

Article

# Modelling of Hydrophilic Interaction Liquid Chromatography Stationary Phases Using Chemometric Approaches

Meritxell Navarro-Reig , Elena Ortiz-Villanueva, Romà Tauler and Joaquim Jaumot \* 

Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, 08034 Barcelona, Spain; meritxell.navarro@idaea.csic.es (M.N.-R.); elena.ortiz@idaea.csic.es (E.O.-V.); rtaqam@idaea.csic.es (R.T.)

\* Correspondence: joaquim.jaumot@idaea.csic.es; Tel.: +34-934-006-100

Received: 15 July 2017; Accepted: 21 October 2017; Published: 24 October 2017

**Abstract:** Metabolomics is a powerful and widely used approach that aims to screen endogenous small molecules (metabolites) of different families present in biological samples. The large variety of compounds to be determined and their wide diversity of physical and chemical properties have promoted the development of different types of hydrophilic interaction liquid chromatography (HILIC) stationary phases. However, the selection of the most suitable HILIC stationary phase is not straightforward. In this work, four different HILIC stationary phases have been compared to evaluate their potential application for the analysis of a complex mixture of metabolites, a situation similar to that found in non-targeted metabolomics studies. The obtained chromatographic data were analyzed by different chemometric methods to explore the behavior of the considered stationary phases. ANOVA-simultaneous component analysis (ASCA), principal component analysis (PCA) and partial least squares regression (PLS) were used to explore the experimental factors affecting the stationary phase performance, the main similarities and differences among chromatographic conditions used (stationary phase and pH) and the molecular descriptors most useful to understand the behavior of each stationary phase.

**Keywords:** hydrophilic interaction liquid chromatography (HILIC); non-targeted metabolomics; stationary phase; chemometrics

## 1. Introduction

Over the past decade, metabolomics has received considerable attention from the scientific community. Metabolomics aims to screen endogenous small molecules (metabolites) present in biological samples, providing a direct measure of the phenotypic state of an organism [1–3]. There are two main metabolomic strategies: targeted and non-targeted. The targeted approach is focused on the investigation of a specific metabolic pathway and, therefore, in the analysis of a reduced and known set of compounds. In contrast, non-targeted metabolomics aims to screen the entire metabolite content of biological samples containing compounds with different physical and chemical properties [4,5].

Due to its high-resolution power, sensitivity and accuracy of  $m/z$  detection, liquid chromatography coupled with high-resolution mass spectrometry (LC-HRMS) has become the analytical platform most used in metabolomics studies. Reversed phased liquid chromatography (RPLC) is useful for the separation of the more hydrophobic compounds, such as lipids. However, RPLC is not recommended for the analysis of some of the most usual metabolite families characterized as being polar and hydrophilic compounds [2,6–9]. In the analysis of polar compounds, and in particular in the metabolomics field, hydrophilic interaction liquid chromatography (HILIC) has become a valuable alternative to RPLC, due to its ability to separate these more hydrophilic compounds [10,11]. Different types of HILIC stationary phases, such as amide, amine, mixed-mode diol and zwitterionic, have

been employed previously in metabolomics studies [11–13]. This variety of HILIC stationary phases allows the separation of metabolites of different properties. Nevertheless, this diversity also makes the selection of the most suitable HILIC stationary phase for a particular study more challenging [10,11,13]. Moreover, the retention mechanism in HILIC has been demonstrated to be more complex than in RPLC mode [14]. In HILIC mode, the separation is primarily achieved due to the partition of analytes between the mobile phase and the hydrophilic layer adsorbed at the surface of the stationary phase. Also, it is proposed that other electrostatic interactions, such as ion exchange interactions, may also contribute to the retention mechanisms. These electrostatic interactions vary among the different types of HILIC stationary phases, making their comparison and selection more difficult [15,16].

Chemometric tools can help in addressing the challenge of finding the most suitable HILIC stationary phase for the analysis of a complex mixture of polar compounds [17]. On the one hand, multivariate data exploratory methods, such as principal component analysis (PCA) [18], can be used to investigate the behavior of HILIC stationary phases and understanding retention mechanisms. Other multivariate statistical methods, like ANOVA-simultaneous component analysis (ASCA) [19], can help evaluating the statistical significance of the experimental factors involved in a non-targeted metabolomics study, such as when different HILIC stationary phases and mobile phase conditions are assessed. On the other hand, building models linking the physicochemical properties of compounds (molecular descriptors, MDs) and their chromatographic behavior (retention factors) could help to get a different insight into the HILIC stationary phase performance [11,20–29]. In addition, these models could also be used to predict the behavior of a chromatographic system and, therefore, can be used to predict the chromatographic retention of unknown compounds and, in some cases, provide additional information to support their identification [11,23,24,30].

In a previous work, different types of HILIC stationary phases have already been evaluated for metabolomics studies [17]. However, in that preliminary work, only 12 metabolites were considered, which simplified the analysis considerably. Moreover, the detection was performed using diode array detector (DAD), which is not the standard in complex studies in the metabolomics field. In contrast, here, a mixture of 54 metabolites is analyzed using LC-HRMS. Four different types of HILIC stationary phases commonly used in metabolomics research have been evaluated under various experimental conditions using different chemometric tools for their application to non-targeted studies. First, PCA and ASCA were applied to explore the general behavior of the considered stationary phases and to evaluate the statistical significance of the three experimental factors discussed in this work (stationary phase, pH and ionic strength). Then, partial least squares regression (PLS) models based on MDs computed from molecular structures of different metabolite families were calculated for evaluating the chromatographic behavior of the four HILIC stationary phases considering the retention factor of the analyzed metabolites.

## 2. Results and Discussion

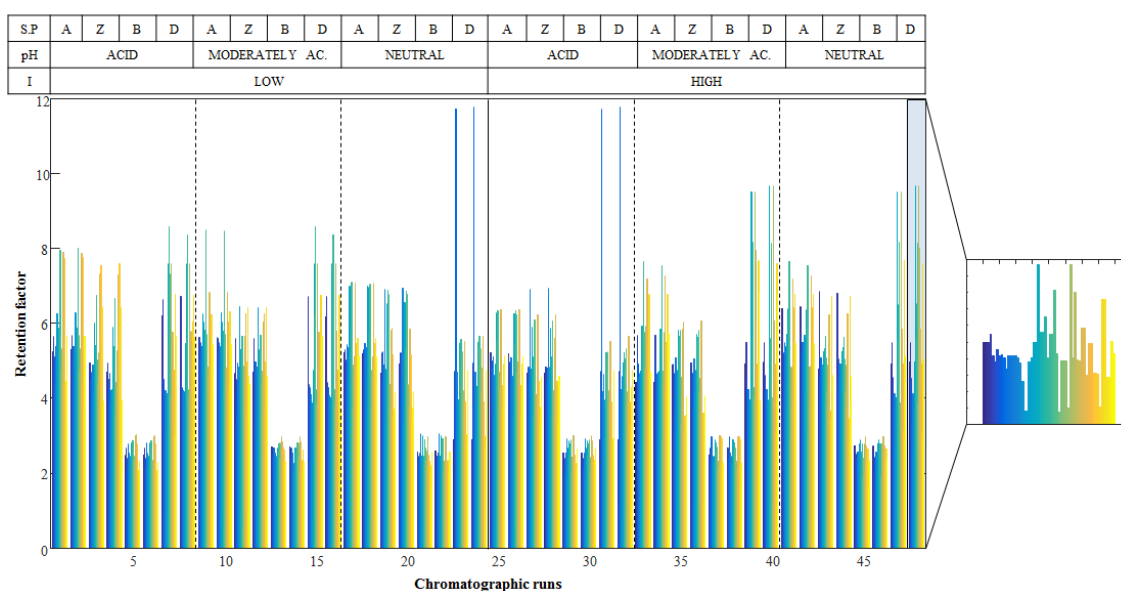
### 2.1. Determination of Retention Factors

A mixture of 54 metabolites from different families (see Table 1) was analyzed using four different stationary phases (Ethylene Bridged Hybrid (BEH) amide, amide, zwitterionic and mixed-mode diol), working with mobile phase at three pH values (acid, moderately acid and neutral) and two ionic strengths (low and high). Each chromatographic condition was injected twice, giving a total of 48 chromatographic runs. The regions of interest (ROI) strategy was used to arrange the 48 chromatographic runs in LC-MS data matrices. Then, each matrix was evaluated to automatically find the  $m/z$  value of each metabolite and provide their retention time in each chromatogram. Lastly, retention factor ( $k$ ) of each metabolite in each chromatographic run was calculated using Equation (1) (see Materials and methods section for more details).

**Table 1.** Metabolites contained in the analyzed mixture.

Metabolite Families					
Nucleosides	Amino Acids	Sugars	Organic Acids	Others	
1-methyladenosine	1-methyl-L-histidine	L-citrulline	D(−)-ribose	citric acid	hypoxanthine
2''-O-methylcytidine	3-methyl-L-histidine	L-glutamic acid	glucose	ketoglutaric acid	L-carnitine
2-thiocytidine	4-hydroxy-L-proline	L-histidine	trehalose	pimelic acid	serotonin
5-methylcytidine	5-hydroxylysine	L-homocystine	mannitol	succinic acid	tryptamine
cytidine	β-alanine	L-isoleucine	-	creatine	-
guanosine	creatinine	L-leucine	-	-	-
inosine	cysteine	L-methionine	-	-	-
pseudouridine	L(−)-proline	L-ornithine	-	-	-
ribothymidine	L(+)-arginine	L-serine	-	-	-
uridine	L(+)-cystathionine	L-threonine	-	-	-
-	L(+)-lysine	L-tryptophan	-	-	-
-	L-2-aminoadipic acid	L-valine	-	-	-
-	L-2-amino-n-butyric acid	taurine	-	-	-
-	L-alanine	sarcosine	-	-	-
-	L-anserine	L-aspartic acid	-	-	-
-	L-carnosine	-	-	-	-

Finally, a matrix, **D**, containing the retention factors of metabolites at each chromatographic condition was built up. This matrix had a number of rows equal to the number of chromatographic runs performed (48 runs) and a number of columns equal to the number of analyzed metabolites (54 compounds). Figure 1 shows the obtained retention factors for the 54 metabolites in the 48 chromatographic runs. The visual inspection of the obtained retention factors already allowed the differentiation of the evaluated chromatographic conditions. For instance, the retention factors of all metabolites in BEH amide are shorter than in the rest of the stationary phases. However, direct evaluation of the HILIC stationary phases behavior is not straightforward. For this reason, matrix **D** was first evaluated using two chemometric exploratory methods, PCA and ASCA, to get a deeper insight into the effects of experimental factors on the retention behavior of metabolites.



**Figure 1.** Retention factors for the 54 metabolites in the 48 chromatographic runs (matrix **D**). The table on the top shows the chromatographic conditions of each sample: (A) indicates the chromatographic runs performed using amide, (Z) zwitterionic, (B) BEH amide and (D) mixed-mode diol stationary phases. As an example, a zoomed view of chromatographic run 48 is depicted.

## 2.2. Evaluation of HILIC Stationary Phases Behavior

PCA was applied to matrix **D** in order to explore the behavior of the chromatographic systems studied in this work (four different stationary phases and mobile phase at three different pH conditions and two different ionic strengths). PCA results indicated that stationary phase was the most critical factor to be considered. In Figure 2a, the PCA scores plot shows that samples analyzed with the four stationary phases were differentiated. Moreover, PC1 distinguished BEH amide stationary phase (XBridge™ Amide) samples (cyan triangles in Figure 2a), which appeared on the left side of PC1, from the rest of the samples. Additionally, PC2 distinguished mixed-mode diol stationary phase (Acclaim™ Mixed-Mode HILIC-1) samples (green squares in Figure 2a), with large positive PC2 scores values, from the rest of the samples. Amide (TSK-Gel Amide-80) and zwitterionic (ZIC-HILIC) stationary phases showed a similar behavior since their samples appeared close to each other in the scores plot. Furthermore, samples analyzed with the three different pH values were also clearly distinguished (see Figure S1 in Supplementary Information), whereas, samples analyzed at different ionic strengths (low and high) could not be differentiated by the PCA scores plot. These results were consistent with the results obtained in the previous authors' work about HILIC stationary phases [17].

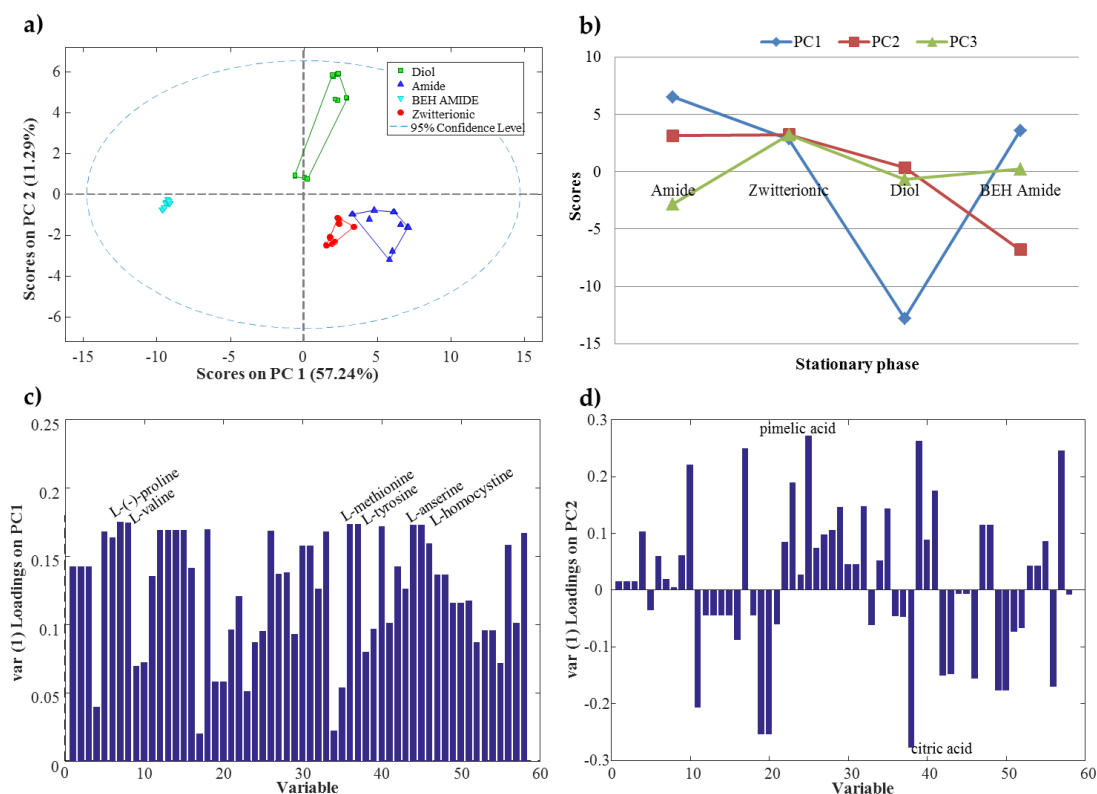
Statistical significance of the three experimental factors considered in this work (stationary phase, pH and ionic strength) and their interactions were assessed by applying ASCA to matrix **D**. Results showed that both stationary phase and pH had statistically significant effects ( $p$ -value of 0.0001). On the contrary, the ionic strength effects were not significant ( $p$ -value of 0.2). Moreover, the interaction of three factors (stationary phase  $\times$  pH, stationary phase  $\times$  ionic strength and pH  $\times$  ionic strength) was also found to be statically significant ( $p$ -value of 0.0001). Hence, from the combination of ASCA and PCA results, the two most relevant factors were defined as the stationary phase and the pH of the aqueous solvent.

Figure 2b shows the ASCA principal component scores (at the mean level) for each HILIC stationary phase. In this scores plot, some trends could be observed. For example, PC1 distinguished the mixed-mode diol stationary phase with a large negative scores value, whereas the two amides and the zwitterionic stationary phases showed a similar positive PC1 scores value. PC2 also differentiated the BEH amide stationary phase with a negative scores value. In contrast, amide, zwitterionic and mixed-mode diol stationary phases had a similar positive PC2 scores value. Finally, it should be mentioned that, as observed in PCA results, amide and zwitterionic stationary phases had similar PC1 and PC2 scores values. Therefore, these two stationary phases showed a similar behavior.

ASCA loadings were useful to know which variables (metabolite retention factors) were the most important to distinguish the stationary phases. Figure 2c,d show the ASCA loadings plot for PC1 and PC2, respectively. For instance, six amino acids (L-(–)-proline, L-valine, L-methionine, L-tyrosine, L-homocysteine and L-anserine) showed a higher loadings value in PC1 (Figure 2c). Consequently, these amino acids were useful to distinguish mixed-mode diol stationary phase from the other three (amide, BEH amide and zwitterionic). In the case of PC2 (Figure 2d), pimelic and citric organic acids showed the highest loadings values. Therefore, these two organic acids appeared as important to differentiate BEH amide from the rest of the stationary phases.

In general, PCA and ASCA results coincided showing that the most important factors were the HILIC stationary phase and the pH of the aqueous solvent. In addition, some facts related to the stationary phase behavior can be highlighted. For instance, the zwitterionic stationary phase showed an intermediate behavior between the two amide stationary phases.

The next step in this work was to find relationships between the observed chromatographic retention observed and the physicochemical properties of metabolites using their molecular descriptors. In addition, since PCA and ASCA results showed that the ionic strength of the mobile phase was not a significant factor in metabolomics studies, these new PLS models were only assessed considering chromatographic runs done at low ionic strength.



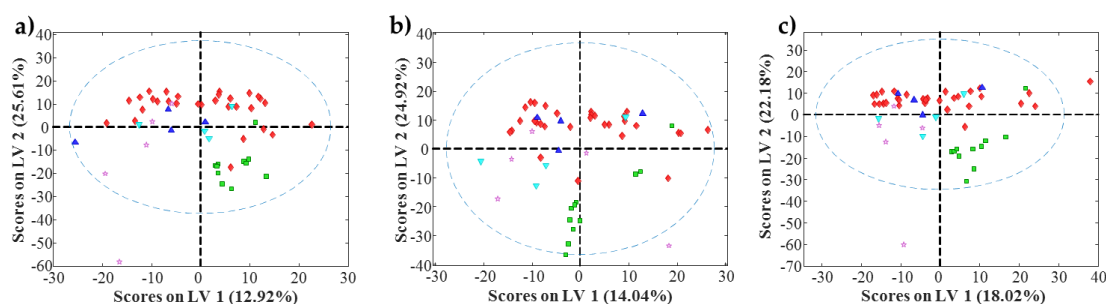
**Figure 2.** (a) Principal component analysis (PCA) scores plot of samples classified according to the stationary phase used in the chromatographic system; (b) ANOVA simultaneous component analysis (ASCA) principal component scores for the stationary phase factor; (c) ASCA PC1 loadings for the stationary phase factor; (d) ASCA PC2 loadings for the stationary phase factor.

### 2.3. Exploratory Relationship between Physicochemical Properties and HILIC Chromatographic Retention

PLS models were independently built for the 12 chromatographic studied systems. They were obtained by the combination of the four stationary phases (BEH amide, amide, zwitterionic and mixed-mode diol) and of the three mobile phases at different pH values (acid, moderately acid and neutral). These models were generated using the experimental retention factors of the 54 metabolites contained in the mixture (a  $y$  vector for each condition) and their corresponding molecular descriptors (MDs) organized in an  $X$  matrix. A preliminary selection step was performed over the whole set of MDs available from PCLIENT to reduce the total number and finally consider 844 of them (see Materials and Methods section for more details).

PLS modelling of the retention factor obtained for each condition using metabolite molecular descriptors required a reduced number of latent variables (between 2 and 3) to explain most of the variance of each  $y$  vector (between 80% and 95%). These models did not show an accuracy enough to be used for the prediction of the retention factors of unknown compounds. However, as these PLS models explain a major part of the retention factor variance, the exploration of scores and loading plots can provide additional insight into the HILIC behavior. First, the analysis of scores plots could allow confirming the differentiation between groups of samples. More interestingly, the analysis of loadings plots could provide information regarding the molecular descriptors related to this differentiation, and could give additional information to know the main physicochemical properties involved in the HILIC retention mechanisms. As an example, Figure 3 shows the scores plots obtained for the amide stationary phase at the three studied pH values of the mobile phases. In these plots, some interesting trends can be observed. First, the differentiation between nucleosides and the rest of the metabolites present in the mixture. In the three cases, a clear group with all nucleosides is visible. Amino acids are

the metabolite family with the most compounds in the mixture. These amino acids are spread along the first latent variable. However, in the three pH conditions, differentiation between two groups can be detected (Figure 3a). An inspection of metabolites forming these two groups allowed observing that the left group was composed of metabolites with a molecular weight lower than 130 Da, whereas the right group was composed of metabolites with a molecular weight larger than 130 Da. Regarding the other families of metabolites present in the mixture (i.e., sugars or organic acids), metabolites were grouped but they were overlapping with amino acids. Similar trends can be observed when considering the other chromatographic conditions.



**Figure 3.** Partial Least Squares (PLS) scores plot for amide stationary phase. (a) Acid pH; (b) Moderately acid pH; (c) Neutral pH. Red diamonds are amino acids ( $\blacklozenge$ ), green squares are nucleosides ( $\blacksquare$ ), blue triangles are organic acids ( $\blacktriangle$ ), purple stars are sugars ( $\blackstar$ ) and cyan triangles are others ( $\blacktriangledown$ ).

The evaluation of which MDs allowed obtaining this retention factor modelling is more appealing. The identification of MDs can be performed using the variable importance on projection (VIP) scores obtained for each PLS model as a feature selection tool allowing to identify which variables (MDs) were the most descriptive of every chromatographic system studied in this work. Table 2 shows the twenty most relevant MDs for each PLS model and their VIP scores values. MDs that appeared to be important in more than one chromatographic system are shown in bold letters. The most repeated MDs values present in Table 2 were the 3D-MoRSE values like the molecular representation of structures based on electronic diffraction, **Mor02p**, **Mor10p**, **Mor15p**, **Mor08u**, **Mor19u**, **Mor24u**, **Mor31u**, **Mor01m**, **Mor06m**, **Mor10m**, **Mor10v**, **Mor15v**, **Mor08e**, **Mor24e** and **Mor31e**. All of them are geometrical descriptors that codify the 3D molecular structure [31–34]. Some other geometrical descriptors (HTcyp, RTe and RTm) also seemed relevant. Geometrical descriptors are calculated from the coordinates of the molecule atoms, interatomic distances and distances from a specific origin. They are geometrical descriptors of the molecular size, shape, symmetry and atom distribution [33,34]. Moreover, these descriptors are weighted by ionization potential, electronegativity, polarizability and molecular mass [11]. Different topological descriptors (J, Lop, MWC03, MWC04, EEig10r, ESpm01d, ESpm07d, ESpm09d, ESpm13d, ESpm14d and ESpm14r) were also highlighted as relevant for the modelling of the studied chromatographic systems. These topological descriptors are numerical quantifiers of molecular topology that are sensitive to one or more structural properties, such as size, shape, symmetry or branching, and can also include chemical information about atom type and bond multiplicity [34–37]. Autocorrelation descriptors (GATS2e, GATS2p, GATS2m, GATS2v, MATS4e, MATS4p, MATS4m, MATS4v, MATS7v, RTm and ATS3m) also appeared to be significant in almost all the studied systems. These descriptors encode both molecular structure and physicochemical properties of a molecule (molecular mass, van der Waals volume, electronegativity or polarizability) [34,38,39]. Constitutional descriptors and molecular properties, like the number of double bonds (nDB) and the number of rotatable bonds (RBN), the unsaturation index (Ui) and the octanol-water partition coefficient (ALOGP and MLOGP) also were relevant in the obtained PLS models. Finally, MDs related to connectivity had significant VIP values for the studied systems. These MDs are called BCUT (Burden eigenvalue descriptors) descriptors (BELe3) [34,40] and Randic connectivity indexes (VRv2 and VRp2) [34,41]. In this case, the selection of different descriptors related to the Sanderson electronegativity can be

mentioned giving a preliminary insight into the interaction mechanism between metabolites and stationary phases.

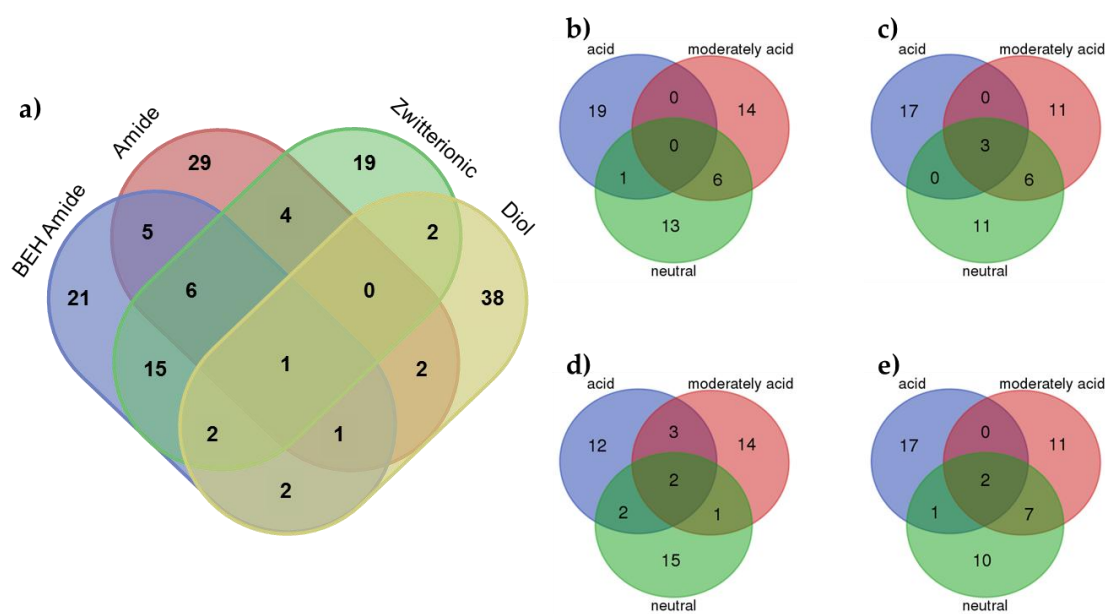
**Table 2.** Variable importance on projection (VIP) scores of the twenty most important molecular descriptors (MD) for each chromatographic system.

BEH Amide Acid		BEH Amide Moderately Acid		BEH Amide Neutral		Amide Acid		Amide Moderately Acid		Amide Neutral	
MD	VIP	MD	VIP	MD	VIP	MD	VIP	MD	VIP	MD	VIP
EEig10r	13.4	<b>R</b> Te	9.62	<b>Mor24u</b>	8.80	G2p	12.3	Mor02e	8.27	<b>Mor31e</b>	13.8
MWC03	11.5	<b>BELe3</b>	7.92	<b>Mor24e</b>	8.62	G2m	11.5	ESpm02x	7.89	<b>Mor31u</b>	13.8
<b>Mor31e</b>	9.76	<b>BLTD48</b>	7.92	RDF080v	7.95	G2u	11.1	<b>Mor31u</b>	6.76	<b>Mor08u</b>	9.85
GATS2v	8.72	<b>BLTA96</b>	7.89	<b>GATS2e</b>	6.10	G2v	11.1	<b>Mor31e</b>	6.70	<b>Mor08e</b>	8.85
GATS2p	8.66	<b>Mor08u</b>	7.84	<b>BELe3</b>	5.64	G2e	11.0	<b>Mor10m</b>	6.61	EPS1	7.04
<b>Mor31u</b>	8.56	<b>Mor24u</b>	7.19	<b>BLTA96</b>	5.61	<b>Mor31e</b>	9.29	<b>ESpm09d</b>	5.87	<b>Mor06m</b>	6.08
<b>Mor02p</b>	8.50	<b>Lop</b>	6.69	<b>BLTD48</b>	5.61	BELe2	9.05	<b>GATS2e</b>	5.81	GATS6v	6.00
<b>GATS2m</b>	8.33	G3v	6.63	ESpm09x	5.00	<b>Mor02p</b>	8.74	<b>Mor06m</b>	5.74	<b>Mor15v</b>	5.79
ESpm05d	8.08	<b>Mor08e</b>	6.59	<b>ESpm14d</b>	4.980	<b>Mor31u</b>	8.45	RDF020p	5.54	Mor08p	5.64
<b>Mor06m</b>	8.04	G3p	6.57	Mor19e	4.87	BEHp1	7.32	CIC0	5.41	<b>Mor15p</b>	5.50
<b>MATS4m</b>	7.45	G3u	6.39	R4e	4.76	G1u	6.99	<b>Mor08u</b>	5.39	<b>Mor10m</b>	5.40
<b>MATS4e</b>	6.74	G3s	6.29	<b>TIC4</b>	4.73	G1e	6.64	WA	5.32	RDF080m	5.31
<b>MATS4v</b>	6.44	G3e	6.27	<b>Mor08u</b>	4.71	G1m	6.60	<b>Mor08e</b>	5.08	<b>GATS2v</b>	5.00
Mor16e	5.39	<b>Mor10p</b>	6.05	<b>TIC3</b>	4.59	G1p	6.48	R5e	4.98	<b>Mor10v</b>	4.64
GATS4v	5.35	<b>Mor24e</b>	6.02	AAc	4.56	G1v	6.45	<b>RTm</b>	4.80	<b>RTm</b>	4.54
<b>HVcpx</b>	5.33	G3m	5.91	IC0	4.56	<b>MWC03</b>	6.21	<b>Mor01m</b>	4.74	Mor08v	4.52
SP03	5.23	<b>Mor10v</b>	5.54	piID	4.49	BEp2	6.10	VDA	4.68	ESpm14x	4.48
Mor17e	5.16	RDF075m	5.39	ESpm02u	4.41	<b>Mor10m</b>	5.25	RDF070u	4.64	<b>RBN</b>	4.47
<b>MATS4p</b>	4.92	<b>VRp2</b>	5.37	<b>Mor19u</b>	4.41	<b>HVcpx</b>	5.22	<b>Mor19u</b>	4.30	<b>ALOGP</b>	4.32
<b>GATS2e</b>	4.71	<b>VRv2</b>	5.37	ESpm12r	4.41	<b>MATS4m</b>	5.15	<b>RBN</b>	4.27	<b>Mor19u</b>	4.32
Zwitterionic Acid		Zwitterionic Moderately Acid		Zwitterionic Neutral		Diol Acid		Diol Moderately Acid		Diol Neutral	
MD	VIP	MD	VIP	MD	VIP	MD	VIP	MD	VIP	MD	VIP
<b>MATS7v</b>	9.20	<b>EEig10r</b>	7.99	<b>MATS4m</b>	1.11	<b>ESpm09d</b>	1.50	G(O..O)	9.65	<b>Ui</b>	6.84
<b>Mor24u</b>	8.03	L3u	7.07	<b>MATS4e</b>	1.00	ESpm02r	1.42	IC3	8.52	<b>ESpm07d</b>	6.54
HTv	7.87	Mor18e	6.69	<b>MATS4v</b>	9.75	ESpm06d	1.12	<b>ATS3m</b>	7.82	<b>J</b>	6.49
<b>Mor24e</b>	7.42	<b>TIC4</b>	6.65	<b>Mor31e</b>	7.65	<b>ESpm13d</b>	7.21	T(O..O)	7.66	<b>ESpm01d</b>	6.41
BELe3	6.72	<b>TIC3</b>	6.40	<b>Mor31u</b>	7.51	<b>RTe</b>	7.08	nO	7.64	<b>nDB</b>	6.13
<b>BLTD4</b>	6.52	TIC5	6.22	GATS1m	7.05	MWC05	7.05	EEig10d	7.46	<b>MWC04</b>	6.05
BELe3	6.51	<b>Mor01m</b>	6.22	<b>EEig10r</b>	6.36	GNar	6.88	<b>J</b>	7.25	Mor23u	5.44
<b>BLTA96</b>	6.48	<b>Lop</b>	5.92	<b>MATS4p</b>	6.24	ATS1p	6.30	<b>ESpm14r</b>	7.12	Mor23e	5.28
<b>Mor08e</b>	5.52	ESpm11x	5.49	ESpm05u	6.11	<b>ATS3m</b>	6.26	<b>MWC04</b>	6.61	<b>ESpm14r</b>	5.21
<b>Mor10p</b>	5.29	<b>GATS2e</b>	5.26	HATS3u	6.09	<b>ESpm14d</b>	6.25	<b>ESpm01d</b>	6.50	<b>ESpm13d</b>	4.77
<b>Mor15p</b>	5.29	<b>Mor31e</b>	5.22	Jhetm	5.75	<b>MLOGP</b>	5.83	<b>nDB</b>	6.41	RDF040m	4.67
<b>MLOGP</b>	5.11	VRv1	5.17	ADDD	5.33	<b>GATS2e</b>	5.78	H0p	6.20	GGI2	4.65
<b>Lop</b>	5.09	<b>Mor31u</b>	4.96	<b>MLOGP</b>	5.31	MATS6p	5.23	<b>ESpm07d</b>	6.16	SPAN	4.65
<b>Mor10v</b>	5.05	ATS2p	4.95	<b>GATS2p</b>	5.23	<b>MATS7v</b>	5.21	AMW	5.93	<b>ATS3m</b>	4.62
<b>Mor31e</b>	4.98	<b>Mor24u</b>	4.95	<b>Mor15p</b>	5.08	ATS1m	5.19	X1sol	5.90	<b>RARS</b>	4.61
<b>Mor15v</b>	4.87	L3e	4.92	Mor07u	4.95	<b>RARS</b>	4.91	AECC	5.79	EEig09d	4.61
<b>VRp2</b>	4.74	ICR	4.82	<b>GATS2v</b>	4.92	<b>Mor24e</b>	4.85	<b>Mor02p</b>	5.73	QXXv	4.58
<b>VRv2</b>	4.74	ESpm08u	4.80	MATS6e	4.78	<b>GATS2m</b>	4.78	<b>Ui</b>	5.68	Mor21u	4.21
<b>Mor31u</b>	4.63	ESpm10x	4.72	<b>GATS2m</b>	4.77	CIC1	4.78	HDcpx	5.08	<b>ALOGP</b>	4.17
Mor03u	4.60	<b>Mor08e</b>	4.64	<b>RBN</b>	4.75	L2e	4.67	<b>ESpm13d</b>	5.03	EEig10x	4.12

Note: Bold format is used to highlight those MDs appearing in more than one chromatographic system.

A deep analysis of the identified molecular descriptors and their relationships with the significant experimental factors (stationary phase and pH value) allowed finding some interesting trends. Figure 4a shows a Venn diagram showing the MDs for each stationary phase considering all pH values. In this plot, the different behavior of the diol stationary phase can be observed. Most MDs (38) were unique, and only some of them appeared as relevant to other stationary phases. This difference in the behavior of the mixed-mode diol stationary phase can be explained by the mixed chemistry of the surface with a hydrophobic alkyl chain with a diol group. These dual properties allow the use of this stationary phase for both RP and HILIC separations but, from our results, modelling using a PLS model approach was more difficult. In addition, despite the ionic strength factor not being significant (using ASCA), the mixed-mode diol stationary phase seemed to be more affected than the other stationary phases, which could also be related to worse modelling. When considering amide,

BEH amide and zwitterionic stationary phases, more similarities in the identified MDs were observed. In addition, from these results, and in accordance with PCA, the zwitterionic stationary phase seemed to have an intermediate behavior between the two amide stationary phases. Finally, evaluation of the MDs identified for the different stationary phases at different pH values showed that most of MDs were unique for a particular condition. However, more similarities can be observed between the moderately acid and neutral pH values (especially in the case of the diol stationary column, confirming its different behavior).



**Figure 4.** Venn diagrams of the VIPs MDs at (a) each stationary phase considering all pH values; (b) BEH Amide at each pH condition; (c) Amide at each pH condition; (d) Zwitterionic at each pH condition and (e) Diol at each pH condition.

### 3. Materials and Methods

#### 3.1. Chemicals and Reagents

Acetic acid ( $\geq 95.0\%$ ), formic acid ( $\geq 95.0\%$ ), ammonia (25%), LC-MS water and Acetonitrile (ACN, LC-MS grade) were obtained from Merck (Darmstadt, Germany). Ammonium acetate ( $\geq 99.0\%$ ) was supplied by Sigma-Aldrich (St. Louis, MO, USA).

A mixture of 54 metabolites was used to evaluate the HILIC stationary phases behavior. Table 1 shows the metabolites contained in the analyzed mixture. All standards were purchased from Sigma-Aldrich (St. Louis, MO, USA). Nucleosides and amino acids were from two mix solutions provided by Sigma-Aldrich (St. Louis, MO, USA). A summary of the polarity of the analyzed metabolites is shown in Table S1 in the supplementary information.

Standard stock solutions ( $1000 \mu\text{g mL}^{-1}$ ) of metabolite mixture were prepared by dissolving an appropriate amount of each metabolite in water and stored at  $-20 \text{ }^\circ\text{C}$  until their use. Working standard solutions ( $20 \mu\text{g mL}^{-1}$ ) were prepared by diluting the stock solution in ACN:H<sub>2</sub>O (1:1).

#### 3.2. Instrumentation

The metabolite mixture was analyzed using an Acquity UHPLC system (Waters, Milford, MA, USA) for the chromatographic separation, equipped with a quaternary pump, an autosampler and a column oven. The mass spectrometer was a triple quadrupole detector (TQD, Waters, Milford, MA, USA) equipped with an electrospray (ESI) ionization source in negative and positive modes. The mass acquisition range was set to 90–1000  $m/z$ .



Four different HILIC stationary phases (BEH amide, amide, zwitterionic and mixed-mode diol) were evaluated (properties summarized in Table 3). The elution gradient was performed using solvent A (acetonitrile) and solvent B (ammonium acetate buffer solution). Chromatographic conditions used for each column are also detailed in Table 3. In order to reproduce the most used chromatographic conditions in metabolomics studies, the experiments were performed using solvent B at three different pH values: acidic (3.0 adjusted with formic acid), moderately acidic (5.5 adjusted with acetic acid) and neutral (7.0 adjusted with ammonia). The pH measurements were performed at 25 °C using an Orion Star A111 pH meter (Thermo Scientific, Waltham, MA, USA), before the additions of the organic solvent. Moreover, two ionic strengths in the aqueous phase were also compared: low (5.0 mM) and high (25 mM).

The metabolite mixture was analyzed with the four stationary phases working with solvent B at the three pH values and the two ionic strengths. Each condition was injected twice giving an experimental design with a total number of 48 chromatographic runs.

### 3.3. Data Analysis

#### 3.3.1. Retention Factor Determination

Figure 5 shows a complete picture of the data analysis strategy from the raw MS data to the chemometric modelling. First, raw chromatographic data files (in .raw format) were converted to the standard CDF format by Databridge function of MassLynx™ v 4.1 software (Waters, Milford, MA, USA). Then, these data files were imported into the MATLAB environment (Release 2015b, The Mathworks Inc., Natick, MA, USA) by using *mzcdfread* and *mzcdf2peak* functions of the MATLAB Bioinformatics Toolbox (4.3.1.version). LC-MS data were then arranged and aligned according to their *m/z* in a data matrix, containing retention times in the rows and selected *m/z* values in the columns. Here, this data matrix was built up using the previously proposed regions of interest (ROI) strategy [42,43]. The ROI approach selects the most relevant mass traces, which are those *m/z* values whose intensity signals are higher than a fixed signal-to-noise ratio threshold and appear a number of times consecutively in the time dimension. These mass traces are searched among all the chromatographic and spectral data. The obtained vectors, containing the intensity of the found ROIs at each time point, are reorganized into a matrix grouping ROIs among all the retention times. The final *m/z* values of each ROI are calculated as the mean of all *m/z* values obtained for that particular ROI. In this work, the parameters for the implementation of this ROI approach are the signal-to-noise ratio threshold (set at 0.1% of the maximum MS signal intensity), the mass accuracy of the mass spectrometer (set at 0.5 Da/e for the TQD MS analyzer used in this work) and the minimum number of consecutive retention times to be considered as a chromatographic peak (set at 25). More details on how this strategy works are given in previous works [1]. Finally, an ROI matrix was obtained for positive and negative ionization modes for each of the 72 chromatographic runs of the present study.

Every ROI matrix corresponding to each chromatographic run was then evaluated to automatically find the *m/z* value of each metabolite and provide their retention time in each chromatogram. Lastly, retention factor (*k*) of each metabolite in each chromatographic run was calculated using their retention times (*t<sub>R</sub>*) and the dead time (*t<sub>0</sub>*, theoretically obtained from the dead volume) as follows:

$$k = \frac{t_R - t_0}{t_0} \quad (1)$$

**Table 3.** HILIC column specifications and chromatographic separation conditions used during the analysis.

Column Specifications				Chromatographic Separation Conditions	
Name	Manufacturer	Stationary Phase	Dimensions	Flow (mL·min <sup>-1</sup> )	Elution Gradient (A: Acetonitrile; B: Water with Ammonium Acetate)
XBridge™ Amide	Waters (Milford, MA, USA)	BEH amide	150 × 4.6 mm <sup>2</sup> i.d., 5 μm	0.15	0–4 min, at 5% B; 4–34 min, from 5% to 70% B; 34–42 min, at 70% B; and 42–44 min, at 5%B
TSK Gel Amide-80	Tosoh Bioscience (Tokyo, Japan)	Amide	250 × 2.0 mm <sup>2</sup> i.d., 5 μm	0.15	0–3 min, at 5% B; 3–27 min, from 5% to 70% B; 27–30 min, at 70% B; and 30–32 min, at 5%B
ZIC-HILIC	SeQuant (Umeå, Sweden)	Zwitterionic	250 × 2.1 mm <sup>2</sup> i.d., 5 μm	0.15	0–3 min, at 5% B; 3–27 min, from 5% to 70% B; 27–30 min, at 70% B; and 30–32 min, at 5%B
Acclaim™ Mixed-Mode HILIC-1	Thermo Scientific (Sunnyvale, CA, USA)	Mixed-mode diol	150 × 2.1 mm <sup>2</sup> i.d., 5 μm	0.15	0–2 min, at 5% B; 2–16 min, from 5% to 70% B; 16–20 min, at 70% B; and 20–22 min, at 5%B

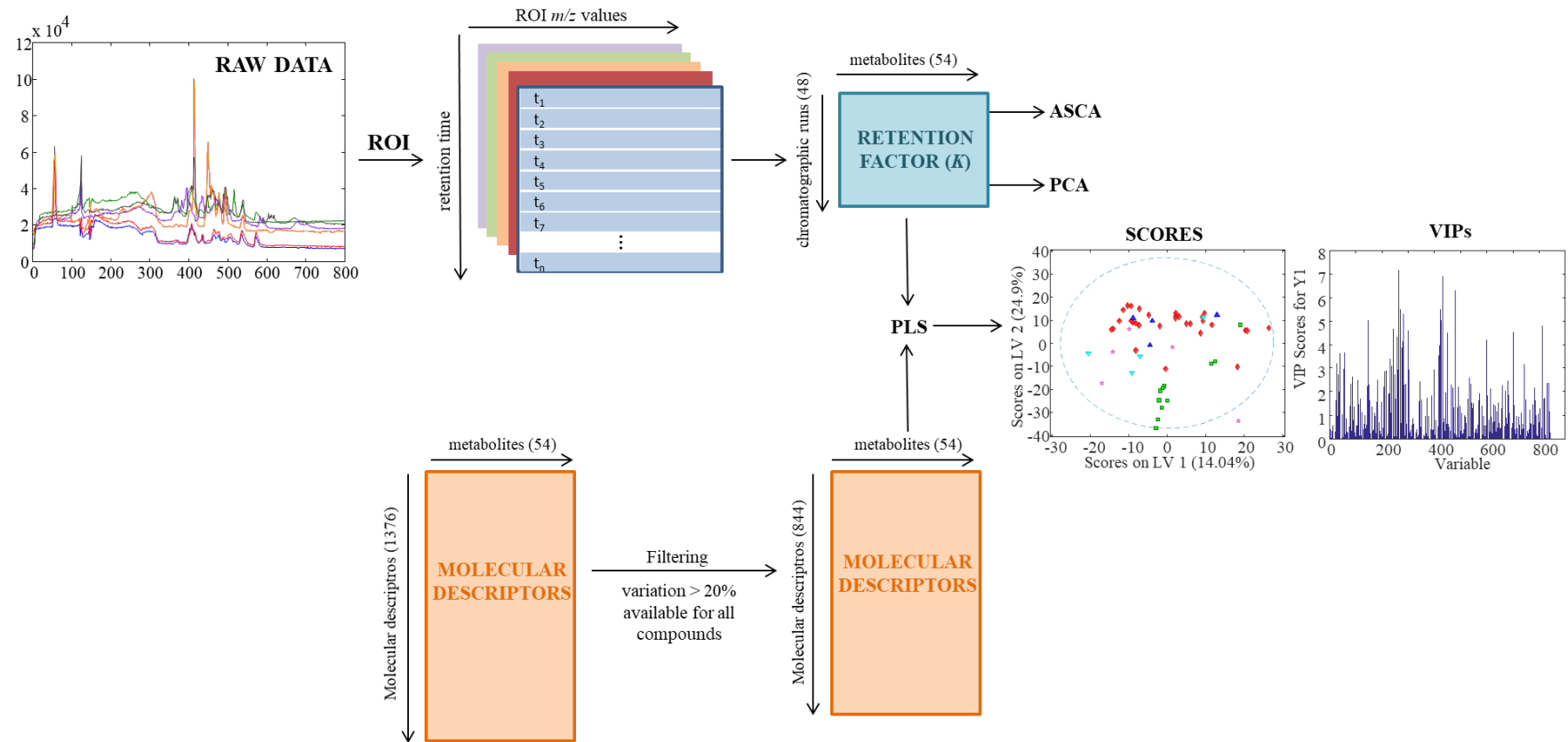


Figure 5. Scheme of data analysis strategy.

### 3.3.2. Molecular Descriptors Determination

Canonical SMILES representations for the standard metabolites were retrieved from PubChem [44] and HMDB [45] databases. These SMILES were input into the PCLIENT software to calculate molecular descriptors (MDs). PCLIENT software can calculate more than 3000 MDs that are divided into 25 logical blocks. Here 1376 MDs were calculated including constitutional, topological, geometrical, electrostatic, physical, shape, and quantum chemical descriptors. Details of MDs calculation can be found in the Handbook of Molecular Descriptors [34]. PCLIENT software is available online at <http://www.vcclab.org>. When 3D atom coordinates were needed for parameter calculation, they were obtained using CORINA software (Molecular Networks GmbH, Nürnberg, Germany).

To reduce the number of MDs descriptors with a percentage of variation (calculated dividing the standard deviation of the MD values by their mean) lower than 20% were excluded. Also, those descriptors not available for all compounds were removed. After this reduction, 844 MDs were obtained for further analysis.

### 3.3.3. Evaluation of HILIC Chromatographic Performance

The retention factors of the 54 metabolites in each HILIC stationary phase at different chromatographic conditions were used to investigate the behavior of the chromatographic systems studied in this work by explorative chemometric methods.

The behaviors of the chromatographic systems studied in this work were evaluated using the retention factors of the 54 metabolites. The retention factor data matrix **D** (containing the retention factor of 54 metabolites at the 48 chromatographic runs) was evaluated using diverse chemometrics exploratory methods: principal component analysis (PCA) [18], ANOVA-simultaneous component analysis (ASCA) [19] and partial least squares regression (PLS).

PCA [18] compresses the information of the original variables into a smaller number of uncorrelated variables known as principal components [18]. In this work, PCA was applied to evaluate the relationships between the experimental conditions studied: stationary phase, pH and ionic strength. Therefore, matrix **D** was analyzed and information about the chromatographic runs and metabolite distribution were obtained in scores and loadings, respectively.

ASCA [19] is a multivariate analysis of variance method that combines the capacity of ANOVA to separate variance sources with the advantages of simultaneous component analysis (SCA, a generalization of PCA for the situation where the same variables have been measured in multiple conditions) [46]. In this work, ASCA was applied to statistically assess the significance of experimental factors in the experimental design: stationary phase, pH and ionic strength. ASCA was performed on a well-balanced experimental design, and 10,000 permutations were used for the permutation test [47]. More details about the ASCA method can be found in the work of Smilde [19] and Jansen [48]. Data were autoscaled prior to applying PCA and centered before applying ASCA.

Finally, PLS regression was used to explore the relationships between obtained retention factors for each chromatographic condition and molecular descriptors (MDs). PLS [49–51] is a multivariate linear regression model used to find correlation models between predictor variables (**X** data matrix) and response values to be predicted (**y** vector). In this work, PLS is used as a regression analysis method to build a model to link the determined retention factor of metabolites (arranged in a vector **y**) using their MDs (arranged in matrix **X**) and to investigate the most influential MDs in the regression. In this work, the optimum number of latent variables for each model was selected using leave-one-out cross-validation.

PLS also provides information about the most relevant variables for achieving the retention factors modelling. For instance, variables importance in projection (VIP) scores can be used for that purpose [51]. According to the common use, variables with a VIP score greater than 1 were important [52]. In this work, these VIP variables corresponded to those MDs that allowed a better description of the retention factor for each considered metabolite.

PCA, ASCA, and PLS were performed using PLS Toolbox 8.0 (Eigenvector Research Inc., Wenatchee, WA, USA) working under MATLAB (The Mathworks, Natick, MA, USA).

#### 4. Conclusions

Results obtained in the assessment of the behavior of HILIC chromatographic stationary phases by means of a variety of chemometric methods showed that the two most important factors to be considered in metabolic studies are the stationary phase and the pH of the aqueous solvent. Moreover, BEH amide and mixed-mode diol stationary phases behaved rather differently compared to amide and zwitterionic phases, which performed similarly. ASCA loadings were useful to know which metabolites were the most important to distinguish the stationary phases. Amino acids appeared to be useful to distinguish the mixed-mode diol stationary phase, while organic acids seemed to distinguish BEH amide from the rest of the stationary phases. In addition, exploratory PLS models allowed linking the retention of metabolites at different chromatographic conditions with molecular descriptors defining their physicochemical properties. Again, a similar behavior was observed for amide and zwitterionic stationary phases whereas the mixed-mode diol stationary phase showed a different performance.

Finally, the obtained PLS models could be considered as a starting point in a more comprehensive work for modelling and prediction of chromatographic retention factors of metabolites in different HILIC stationary phases. However, building up these models requires bigger metabolite datasets with a larger number of compounds for each of the metabolite families. Moreover, efforts should be made in performing a comprehensive external validation of the models. Future work should address these issues.

**Supplementary Materials:** The following are available online at [www.mdpi.com/2218-1989/7/4/54/s1](http://www.mdpi.com/2218-1989/7/4/54/s1), Figure S1: PCA scores plot of samples classified according to the pH of the mobile phase. Table S1: Metabolites' logP values obtained from PCLIENT software.

**Acknowledgments:** The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement Number 320737.

**Author Contributions:** All authors conceived and designed the experiments; M.N.-R. and E.O.-V. performed the chromatographic experiments; M.N.-R. performed the chemometric data analysis; M.N.-R., R.T. and J.J. discussed and interpreted results; M.N.-R. and J.J. drafted the manuscript. All authors revised the article and gave final approval of the submitted version.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### References

1. Gorrochategui, E.; Jaumot, J.; Lacorte, S.; Tauler, R. Data analysis strategies for targeted and untargeted LC-MS metabolomic studies: Overview and workflow. *TrAC Trends Anal. Chem.* **2016**, *82*, 425–442. [[CrossRef](#)]
2. Rochat, B. From targeted quantification to untargeted metabolomics: Why LC-high-resolution-MS will become a key instrument in clinical labs. *TrAC Trends Anal. Chem.* **2016**, *84*, 151–164. [[CrossRef](#)]
3. Jorge, T.F.; Rodrigues, J.A.; Caldana, C.; Schmidt, R.; van Dongen, J.T.; Thomas-Oates, J.; António, C. Mass spectrometry-based plant metabolomics: Metabolite responses to abiotic stress. *Mass Spectrom. Rev.* **2016**, *35*, 620–649. [[CrossRef](#)] [[PubMed](#)]
4. Horvatovich, P.L.; Bischoff, R. Current technological challenges in biomarker discovery and validation. *Eur. J. Mass Spectrom.* **2010**, *16*, 101–121. [[CrossRef](#)] [[PubMed](#)]
5. Monteiro, M.S.; Carvalho, M.; Bastos, M.L.; De Pinho, P.G. Metabolomics analysis for biomarker discovery: Advances and challenges. *Curr. Med. Chem.* **2013**, *20*, 257–271. [[CrossRef](#)] [[PubMed](#)]
6. Madji Hounoum, B.; Blasco, H.; Emond, P.; Mavel, S. Liquid chromatography-high-resolution mass spectrometry-based cell metabolomics: Experimental design, recommendations, and applications. *TrAC Trends Anal. Chem.* **2016**, *75*, 118–128. [[CrossRef](#)]

7. Naz, S.; Moreira Dos Santos, D.C.; García, A.; Barbas, C. Analytical protocols based on LC-MS, GC-MS and CE-MS for nontargeted metabolomics of biological tissues. *Bioanalysis* **2014**, *6*, 1657–1677. [[CrossRef](#)] [[PubMed](#)]
8. Cubbon, S.; Antonio, C.; Wilson, J.; Thomas-Oates, J. Metabolomic applications of HILIC–LC–MS. *Mass Spectrom. Rev.* **2010**, *29*, 671–684. [[CrossRef](#)] [[PubMed](#)]
9. Spagou, K.; Tsoukali, H.; Raikos, N.; Gika, H.; Wilson, I.D.; Theodoridis, G. Hydrophilic interaction chromatography coupled to MS for metabolomic/metabolomic studies. *J. Sep. Sci.* **2010**, *33*, 716–727. [[CrossRef](#)] [[PubMed](#)]
10. Tang, D.Q.; Zou, L.; Yin, X.X.; Ong, C.N. HILIC-MS for metabolomics: An attractive and complementary approach to RPLC-MS. *Mass Spectrom. Rev.* **2016**, *35*, 574–600. [[CrossRef](#)] [[PubMed](#)]
11. Taraji, M.; Haddad, P.R.; Amos, R.I.J.; Talebi, M.; Szucs, R.; Dolan, J.W.; Pohl, C.A. Prediction of retention in hydrophilic interaction liquid chromatography using solute molecular descriptors based on chemical structures. *J. Chromatogr. A* **2017**, *1486*, 59–67. [[CrossRef](#)] [[PubMed](#)]
12. Guo, Y. Recent progress in the fundamental understanding of hydrophilic interaction chromatography (HILIC). *Analyst* **2015**, *140*, 6452–6466. [[CrossRef](#)] [[PubMed](#)]
13. Hendrickx, S.; Adams, E.; Cabooter, D. Recent advances in the application of hydrophilic interaction chromatography for the analysis of biological matrices. *Bioanalysis* **2015**, *7*, 2927–2945. [[CrossRef](#)] [[PubMed](#)]
14. Alpert, A.J. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *J. Chromatogr. A* **1990**, *499*, 177–196. [[CrossRef](#)]
15. Jandera, P. Stationary and mobile phases in hydrophilic interaction chromatography: A review. *Anal. Chim. Acta* **2011**, *692*, 1–25. [[CrossRef](#)] [[PubMed](#)]
16. Buszewski, B.; Noga, S. Hydrophilic interaction liquid chromatography (HILIC)—A powerful separation technique. *Anal. Bioanal. Chem.* **2012**, *402*, 231–247. [[CrossRef](#)] [[PubMed](#)]
17. Ortiz-Villanueva, E.; Navarro-Reig, M.; Jaumot, J.; Tauler, R. Chemometric evaluation of hydrophilic interaction liquid chromatography stationary phases: Resolving complex mixtures of metabolites. *Anal. Method.* **2017**, *9*, 774–785. [[CrossRef](#)]
18. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52. [[CrossRef](#)]
19. Smilde, A.K.; Jansen, J.J.; Hoefsloot, H.C.J.; Lamers, R.J.A.N.; van der Greef, J.; Timmerman, M.E. ANOVA-simultaneous component analysis (ASCA): A new tool for analyzing designed metabolomics data. *Bioinformatics* **2005**, *21*, 3043–3048. [[CrossRef](#)] [[PubMed](#)]
20. Creek, D.J.; Jankevics, A.; Breitling, R.; Watson, D.G.; Barrett, M.P.; Burgess, K.E.V. Toward global metabolomics analysis with hydrophilic interaction liquid chromatography-mass spectrometry: Improved metabolite identification by retention time prediction. *Anal. Chem.* **2011**, *83*, 8703–8710. [[CrossRef](#)] [[PubMed](#)]
21. Falchi, F.; Bertozzi, S.M.; Ottonello, G.; Ruda, G.F.; Colombano, G.; Fiorelli, C.; Martucci, C.; Bertorelli, R.; Scarpelli, R.; Cavalli, A.; et al. Kernel-based, partial least squares quantitative structure-retention relationship model for UPLC retention time prediction: A useful tool for metabolite identification. *Anal. Chem.* **2016**, *88*, 9510–9517. [[CrossRef](#)] [[PubMed](#)]
22. Yousefinejad, S.; Hemmateenejad, B. Chemometrics tools in QSAR/QSPR studies: A historical perspective. *Chemom. Intell. Lab. Syst.* **2015**, *149*, 177–204. [[CrossRef](#)]
23. Goryński, K.; Bojko, B.; Nowaczyk, A.; Bucirński, A.; Pawliszyn, J.; Kaliszan, R. Quantitative structure-retention relationships models for prediction of high performance liquid chromatography retention time of small molecules: Endogenous metabolites and banned compounds. *Anal. Chim. Acta* **2013**, *797*, 13–19. [[CrossRef](#)] [[PubMed](#)]
24. Zisi, C.; Sampsonidis, I.; Fasoula, S.; Papachristos, K.; Witting, M.; Gika, H.G.; Nikitas, P.; Pappa-Louisi, A. QSRR modeling for metabolite standards analyzed by two different chromatographic columns using multiple linear regression. *Metabolites* **2017**, *7*. [[CrossRef](#)] [[PubMed](#)]
25. Randazzo, G.M.; Tonoli, D.; Hambye, S.; Guillarme, D.; Jeanneret, F.; Nurisso, A.; Goracci, L.; Boccard, J.; Rudaz, S. Prediction of retention time in reversed-phase liquid chromatography as a tool for steroid identification. *Anal. Chim. Acta* **2016**, *916*, 8–16. [[CrossRef](#)] [[PubMed](#)]

26. Kritikos, N.; Tsantili-Kakoulidou, A.; Loukas, Y.L.; Dotsikas, Y. Liquid chromatography coupled to quadrupole-time of flight tandem mass spectrometry based quantitative structure–retention relationships of amino acid analogues derivatized via n-propyl chloroformate mediated reaction. *J. Chromatogr. A* **2015**, *1403*, 70–80. [[CrossRef](#)] [[PubMed](#)]
27. Park, S.H.; Haddad, P.R.; Talebi, M.; Tyteca, E.; Amos, R.I.J.; Szucs, R.; Dolan, J.W.; Pohl, C.A. Retention prediction of low molecular weight anions in ion chromatography based on quantitative structure-retention relationships applied to the linear solvent strength model. *J. Chromatogr. A* **2017**, *1486*, 68–75. [[CrossRef](#)] [[PubMed](#)]
28. Aicheler, F.; Li, J.; Hoene, M.; Lehmann, R.; Xu, G.; Kohlbacher, O. Retention time prediction improves identification in nontargeted lipidomics approaches. *Anal. Chem.* **2015**, *87*, 7698–7704.
29. Wolfer, A.M.; Lozano, S.; Umbdenstock, T.; Croixmarie, V.; Arrault, A.; Vayer, P. UPLC–MS retention time prediction: A machine learning approach to metabolite identification in untargeted profiling. *Metabolomics*. **2015**, *12*, 8. [[CrossRef](#)]
30. Cao, M.; Fraser, K.; Huege, J.; Featonby, T.; Rasmussen, S.; Jones, C. Predicting retention time in hydrophilic interaction liquid chromatography mass spectrometry and its use for peak annotation in metabolomics. *Metabolomics* **2015**, *11*, 696–706. [[CrossRef](#)] [[PubMed](#)]
31. Devinyak, O.; Havrylyuk, D.; Lesyk, R. 3D-MoRSE descriptors explained. *J. Mol. Graph. Model.* **2014**, *54*, 194–203. [[CrossRef](#)] [[PubMed](#)]
32. Gasteiger, J.; Sadowski, J.; Schuur, J.; Selzer, P.; Steinhauer, L.; Steinhauer, V. Chemical information in 3D space. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1030–1037. [[CrossRef](#)]
33. Schuur, J.H.; Selzer, P.; Gasteiger, J. The coding of the three-dimensional structure of molecules by molecular transforms and its application to structure-spectra correlations and studies of biological activity. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 334–344. [[CrossRef](#)]
34. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH Verlag GmbH: Weinheim, Germany, 2008.
35. Balaban, A.T. Highly discriminating distance-based topological index. *Chem. Phys. Lett.* **1982**, *89*, 399–404. [[CrossRef](#)]
36. Balaban, A.T. Chemical graphs-XXXIV. Five new topological indices for the branching of tree-like graphs. *Theor. Chim. Acta* **1979**, *53*, 355–375. [[CrossRef](#)]
37. Gozalbes, R.; Doucet, J.P.; Derouin, F. Application of topological descriptors in QSAR and drug design: History and new trends. *Curr. Drug Targets Infect. Disord.* **2002**, *2*, 93–102. [[CrossRef](#)] [[PubMed](#)]
38. Hollas, B. An analysis of the autocorrelation descriptor for molecules. *J. Math. Chem.* **2003**, *33*, 91–101. [[CrossRef](#)]
39. Broto, P.; Moreau, G.; Vandycke, C. Molecular structures: Perception, autocorrelation descriptor and sar studies. *Eur. J. Med. Chem.* **1984**, *19*, 66–70.
40. Burden, F.R. A chemically intuitive molecular index based on the eigenvalues of a modified adjacency matrix. *Quant. Struct.-Act. Relationsh.* **1997**, *16*, 309–314. [[CrossRef](#)]
41. Randić, M. Molecular shape profiles. *J. Chem. Inform. Comp. Sci.* **1995**, *35*, 373–382. [[CrossRef](#)]
42. Tautenhahn, R.; Bottcher, C.; Neumann, S. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinform.* **2008**, *9*. [[CrossRef](#)] [[PubMed](#)]
43. Stolt, R.; Torgrip, R.J.O.; Lindberg, J.; Csenki, L.; Kolmert, J.; Schuppe-Koistinen, I.; Jacobsson, S.P. Second-order peak detection for multicomponent high-resolution LC/MS data. *Anal. Chem.* **2006**, *78*, 975–983. [[CrossRef](#)] [[PubMed](#)]
44. Kim, S.; Thiessen, P.A.; Bolton, E.E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B.A.; et al. PubChem Substance and Compound databases. *Nucleic Acids Res.* **2016**, *44*, D1202–D1213. [[CrossRef](#)] [[PubMed](#)]
45. Wishart, D.S.; Knox, C.; Guo, A.C.; Eisner, R.; Young, N.; Gautam, B.; Hau, D.D.; Psychogios, N.; Dong, E.; Bouatra, S.; et al. HMDB: A knowledgebase for the human metabolome. *Nucleic Acids Res.* **2009**, *37*, D603–D610. [[CrossRef](#)] [[PubMed](#)]
46. Jansen, J.J.; Hoefsloot, H.C.J.; Van Der Greef, J.; Timmerman, M.E.; Smilde, A.K. Multilevel component analysis of time-resolved metabolic fingerprinting data. *Anal. Chim. Acta* **2005**, *530*, 173–183. [[CrossRef](#)]
47. Vis, D.J.; Westerhuis, J.A.; Smilde, A.K.; van der Greef, J. Statistical validation of megavariate effects in ASCA. *BMC Bioinform.* **2007**, *8*. [[CrossRef](#)] [[PubMed](#)]

48. Jansen, J.J.; Hoefsloot, H.C.J.; Van Der Greef, J.; Timmerman, M.E.; Westerhuis, J.A.; Smilde, A.K. ASCA: Analysis of multivariate data obtained from an experimental design. *J. Chem.* **2005**, *19*, 469–481. [[CrossRef](#)]
49. Geladi, P.; Kowalski, B.R. Partial least-squares regression: A tutorial. *Anal. Chim. Acta* **1986**, *185*, 1–17. [[CrossRef](#)]
50. Wold, H. *Multivariate Analysis*; Academic Press: New York, NY, USA, 1966; pp. 391–420.
51. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chem. Intell. Lab. Syst.* **2001**, *58*, 109–130. [[CrossRef](#)]
52. Chong, I.G.; Jun, C.H. Performance of some variable selection methods when multicollinearity is present. *Chem. Intell. Lab. Syst.* **2005**, *78*, 103–112. [[CrossRef](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).