Characterization of the new Celeris™ Arginine column: retentive behaviour through a combination of chemometric tools and potential in drug analysis

Giacomo Russo[1], Maura Vallaro[2], Luca Cappelli[2], Scott Anderson[3], Giuseppe Ermondi[2], Giulia Caron*[2].

1. School of Applied Sciences, Sighthill Campus, Edinburgh Napier University, 9 Sighthill Ct, EH11 4BN Edinburgh, United Kingdom.
2. CASSMedChem Research Group, Molecular Biotechnology and Health Sciences Department, University of Turin, Italy.
3. Regis Technologies Inc., 8210 Austin Ave, Morton Grove IL, 60053, USA.

* Corresponding author

Correspondence:
Prof. Dr. Giulia Caron
Molecular Biotechnology and Health Sciences Department
University of Turin
Via Quarello, 15, 10135
Torino, Italy.
Tel: +39 0116708337
Fax: +39 0112368337
Electronic mail: giulia.caron@unito.it

Abstract

Celeris^{TM} Arginine (ARG) is a mixed-mode stationary phase recently released on the market. To characterize its analytical behavior, the retention factors of a pool (n=100, of which 36 neutrals, 26 acids and 38 bases) of pharmaceutically relevant compounds have been measured on this phase over eight percentages (from 10 to 90% v/v) of acetonitrile (MeCN) as organic modifier. The ARG phase exhibited enhanced affinity for the molecules that are in their anionic form at the experimental pH, whilst basic compounds, albeit over a wide range of lipophilicity and pK_a values, were on average poorly retained. To dissect the separation mechanism of the ARG phase, the overall analytical retention has been deconvoluted into the individual contributions of intermolecular forces by a QSPR/ Partial Least Square (PLS)/Block Relevance (BR) analysis tool recently developed by us.

For the neutrals, the most relevant blocks were found to be Size, describing the interaction due to the dimension of the molecule, and O, representing the solute's hydrogen bond donor properties. The change in sign from positive to negative of the Size block, which occurs between 10% and 20% MeCN, allowed to visually appreciate the switch in the separation mode from reversed phase to normal phase. Some good statistic models for rationalizing the analytical behaviour of neutrals were developed from VS+ descriptors. However, their performance in modelling the analytical retention of acids was substandard, probably due to the intrinsic inefficacy of VS+ descriptors in handling electric charges. This instance was addressed by a complimentary MLR strategy, which led to successfully model the retention of acids on the ARG column and to shed light into their retention mechanism, which seemed to be substantially driven by electrostatics.

Keywords: arginine; block relevance analysis; mixed-mode selectivity; liquid chromatography; drug analysis; chemometrics.

1.0 Introduction

Nowadays liquid chromatography (LC) is one of the most essential and pervasive techniques in the toolbox of analytical chemists, allowing identification, quantification, and purification of the individual components from a mixture[1-3]. For instance, LC is widely applied in almost all stages of the drug discovery/development process[4] to check the identity and the purity of new chemical entities before testing their potency against the desired molecular target(s). Other noteworthy applications relate to therapeutic drug monitoring[5, 6], which is conducted by determining the analytes of interest in biological specimens, such as blood, urine or tissues during the preclinical and clinical phase. LC is also massively exploited to quality control, impurity checks, stability investigations, and many other purposes relevant in drug development[7].

Albeit LC is extremely widespread and of rather common use, analytical method development can be an extremely daunting process, as +1000 stationary phase chemistries are available on the market and their commercial offer keeps widening[8]. Moreover, the complexity of samples to screen is constantly increasing, posing to separation scientists unprecedented challenges.

Whichever purpose the separation scientist pursues, a deep understanding of the intermolecular interactions establishing between the analytes and the stationary phase[9] should be regarded as mandatory not only to select the right analytical column for each sample composition, but also in the analytical method development[10]. However, the overall analytical retention results from the interplay and overlapping of rather complex molecular forces. In method development, the selectivity of two analytical columns is often compared to find out how they perform in separating a set of solutes. The separations achieved by two different columns are considered “orthogonal” if their mechanisms are independent from each other, therefore providing complementary selectivities. The separations achieved by two different columns are instead considered “equivalent” if separation mechanisms coincide, as for instance occurs when both are driven by solute’s hydrophobicity. It might be useful to have equivalent columns to identify an alternative column for running a method or to replace one that is no longer available commercially. Separations’ orthogonality has become increasingly sought after in recent years, also due to the introduction of two-dimensional liquid chromatography (2DLC). 2DLC allows the simultaneous combination of more separation modes, significantly expanding peak capacity of the separation. In that case, the increase in resolving power depends upon the degree of orthogonality exhibited by the separation mode in each dimension, being the greater the orthogonality, the higher the resolving power.

The Celeris™ Arginine column (from now onwards called ARG phase) is a recently marketed chromatographic column which is amenable both in LC (as detailed below in this paper), and in supercritical fluid chromatography (SFC) mode [11]. The arginine (ARG) phase (Figure 1) is a silica surface modified with the amino acid arginine, exhibiting both acidic and basic functionality. *A priori*, the ARG phase is expected to exhibit strong affinity for hydrophilic compounds as this supports a number of polar atoms. Moreover, the ARG phase is supposed to retain preferably anions as it has a positive (+1) overall charge and to support mixed-mode selectivity.

To the best of our knowledge, the analytical retentive behaviour of the ARG phase has never been investigated before. Wu and co-workers developed an arginine functionalized stationary phase for hydrophilic interaction liquid chromatography back in 2015[12], however the chemistry of this phase is rather different and this analytical column is not commercially available.

To study the separation mechanisms of chromatographic columns, it is necessary to deconvolute the individual contributions of intermolecular forces from the overall analytical retention. Solid approaches to realize this include: Abraham's Linear solvation energy relationships[13], the Hydrophobic-Subtraction Model[14], the Tanaka-parameter based approach[15] and the Geometric Approach to Factor Analysis[16]. Indeed, Quantitative Structure-Property Relationship (QSPR) strategies have been successfully applied to the modelling of chromatographic indexes[17] from a variety of separation modes. Recently we introduced the QSPR/ Partial Least Square (PLS)/Block Relevance (BR) analysis (hereafter named BR analysis), a chemoinformatic tool which affords an interpretation of QSPR models based on a selected pool of descriptors and a PLS algorithm[18-22]. The main readout of BR analysis is a couple of plots in which the main components of the intermolecular interactions are quantified and output as blocks. To model physicochemical properties five blocks of intermolecular interactions are essential: the DRY block (hydrophobic interaction), the OH2 block (interaction with water), the O block (the HB interaction between solute HBD and system HBA); the N1 block (between solute HBA and system HBD) and the Others block (additional molecular descriptors that represent the unbalance of hydrophilic and hydrophobic regions on the surface target). Molecular dimensions are also crucial to characterize drug-like candidates, therefore a sixth block of size and shape descriptors (the Size block) was added.

The major drawback of BR analysis is due to the limits of the VolSurf+ descriptors when applied to completely ionised compounds. To overcome this limit, a second QSPR strategy based on a different pool of *ad hoc* physicochemical descriptors has been proposed[23]. In brief, after computing the Gasteiger-Marsili[24] charges of the compounds, an array of physico-chemical and topological

charge-dependent descriptors were calculated (Table S1) and subsequently used to develop statistic models for the various capacity factors.

In this study, a systematic characterization of the main intermolecular forces driving analytical retention on the ARG column has been undertaken. To achieve this aim, we a) measured the capacity factors, in logarithmic scale (log k), of a number of pharmaceutically relevant compounds supporting acidic, basic, and no ionizable (neutral molecules) moieties at eight different concentration of organic modifier; b) studied whether or not the analytical retention on the ARG phase related with that on other stationary phases commercially available; c) applied BR analysis, to visually inspect the molecular interactions driving analytical retention, and d) performed multilinear regression (hereafter named MLR) implementing a pool of charge-based descriptors to model the retention of completely ionised compounds and to evaluate if and to what extent any prediction of chromatographic affinity was feasible.

Our final aim is shedding light and systematically dissect the retention mechanisms of the ARG phase in LC setups to identify potential applications of this new phase in drug analysis/separation science.

2.0 Materials and Methods

2.1 Dataset

The investigated dataset contains 36 neutral, 26 acidic and 38 basic pharmaceutically relevant compounds. The SMILES codes are reported in Table S2.

2.2 Chemicals and sample preparation.

The solutes were obtained from three commercial sources (Aldrich (www.sigmaaldrich.com, Darmstadt, Germany), VWR (www.vwr.com, Milano, Italy), Alfa Aesar (www.alfa.com, Kande, Germany), and their purity was equal to or higher than 98%. Acetonitrile (HPLC grade) was purchased from VWR, and Ammonium Acetate (reagent grade ≥ 98%) was purchased from Alfa Aesar.

All the compounds were solubilised in the mobile phase, at a concentration range of 50-100 µg mL^{-1}.

2.3 Chromatographic hardware

An HPLC Varian ProStar chromatograph (Agilent, 5301 Stevens Creek Blv, Santa Clara, CA, USA) equipped with a 410 autosampler with a built-in thermostatable column compartment, a PDA 335 LC Detector and Galaxie Chromatography Data System Version 1.9.302.952 was used. The column was

a Celeris™ Arginine 100 × 4.6 mm, 5 µm, 100Å from Regis Technologies (Austin Avenue, Morton Grove, IL, USA).

2.4 Chromatographic conditions

All LC analyses were performed at 30°C with a 20 mM ammonium acetate buffer pH 7.0 in mixture with acetonitrile at various percentages (from 10 to 80%, v/v). Flow rate was 1.0 ml min^{-1} and the injection volume was 10 µL. Capacity factors results from the averages of at least three independent measurements.

2.5 Postprocessing of chromatographic signals

Capacity factors on the ARG phase were accounted for by Eq. 1:

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

In which t_r is the retention time (min) of the analyte of interest and t_0 the dead time, determined by monitoring the baseline disturbance. Plotting and data analysis was done by Microsoft Excel for Office 365 v 16.0 at 64 bits.

2.6 Computational analysis

2.6.1 Principal Components Analysis (PCA)

PCA was performed with a MatLab script (ver. R2019a, https://it.mathworks.com/).

2.6.2 Comparison between ARG selectivity and those of other marketed analytical columns

A data matrix including log k of 36 neutrals obtained on the Celeris™ Arginine and those measured on other stationary phases/experimental conditions was submitted to Matlab to calculate the correlation matrix. The columns used for the comparison are the following:

- ABZ[25] (Supelco, Bellefonte, PA, USA 5 µm, 5 cm × 4.6 mm, 120 Å)
- X-Bridge[25] (Waters, Milford. MA, USA 5 µm, 5 cm × 4.6 mm, 130 Å)
- PLRP-S[26] (Agilent, Santa Clara, CA, USA 5 µm, 5 cm × 4.6 mm, 100 Å)
- IAM.PC.DD2[27] (Regis Technologies Inc., Morton Grove, IL, USA 10 µm, 100 × 4.6 mm 300 Å)
- IAM.SPH[18] (synthesized *in house*[28], 5 µm 10 cm × 2.1 mm, 300 Å)

- ZIC®-cHILIC[22] (Merck, Darmstadt, Germany, 3 µm, 10 cm × 4.6 mm, 100 Å)
- ZIC®-HILIC[22] (Merck, Darmstadt, Germany, 5 µm 10 cm × 4.6 mm, 200 Å)
- ZIC®-pHILIC[22] (Merck, Darmstadt, Germany, 5 µm 10 cm × 4.6 mm, 200 Å)

2.6.3 BR analysis

BR analysis was accomplished as detailed elsewhere[19]. The SMILES codes (Table S2) of the 88 compounds were used as an input for VolSurf+ software (www.moldiscovery.com, ver 1.1.2). The electrical state was assigned by pK_a calculations implemented in the software and an average conformation was build and minimised. The 82 descriptors directly obtained from 3D molecular interaction fields (MIFs) were then calculated. The data matrixes, including descriptors and chromatographic data, were submitted to Matlab to perform PLS and VIP analysis. As already discussed elsewhere[19], since here the PLS model is used for interpretative and not predictive purposes, only internal validation was performed. Outliers were identified from the residual plots, when exceeding ±0.5.

Finally, an *in-house* Matlab script grouped the descriptors in blocks and processed the corresponding VIPs to draw the BR plots. Processing was done on a laptop equipped with a 4 cores Intel i7-4700MQ and 12 GB of RAM operating with Windows 10.

BR analysis interpretation is obtained by two graphical outputs: a) the *absolute BR plot* that shows the relevance of any block to the PLS model independently of the sign (the higher, the more relevant) and b) the *BR plot with signs* which splits the contribution of any block into positive BR (+) and negative BR (-) portions. BR (+) indicates how much the considered block favours the considered descriptor (*e.g.,* log k^{ARG}) whereas BR (-) shows how much the block lowers the descriptor. Blocks with small and comparable positive and negative contributions indicate the high noise and inter-correlation of the descriptors of the block itself and thus are poorly relevant in the description of the investigated phenomenon.

2.6.4 MLR

MLR analysis was accomplished by VEGA ZZ x64 software 3.2.0.9[29] implemented on a one 8 core i7 at 3.1 Ghz CPU and 32 GB of RAM Windows desktop machine. Physico-chemical and topological properties (Virtual log P[30], lipole[31], volume, polar surface area, surface accessible to the solvent, gyration radius, ovality, mass, number of atoms, angles, dihedrals, etc) were calculated by VEGA ZZ software (Table S1) and finally, all molecules were inserted into a Microsoft Access database. An additional number of descriptors (HLB, polarizability, log P) were calculated by

MarvinSketch v. 21.3 operated on an 8-core Mac computer. Detailed information is reported in here[32]. In brief, the starting three-dimensional structures of the considered molecules were downloaded from PubChem database [33, 34], and they were considered in both zero atomic charge and ionized form (acids and bases). Furthermore, a weighted average according the the experimental pK_a values was performed. The Gasteiger−Marsili method[24], along with CHARMM force field [35-37], was applied to calculate the atomic charges. After that, structures were minimized by AMMP software[38] (conjugate gradients, 3000 iterations, toler 0.01). The best independent variables were selected by calculating the correspondent equation with a single regressor. Regressions with r^2 value less than 0.10 automatically determine the exclusion of the independent variable. Collinear independent variables were identified by calculating the Variance Inflation Factor (VIF) value for each regressor pair. Variable pairs with VIF > 5.0 were not considered in the model calculation. Statistic models with a number of regressors from one to three were developed by using either the zero-charge or the ionized forms of the compounds. For each model, a cross-validation procedure (leave-one-out) is performed. For the sake of conciseness, only LOO models were discussed.

3.0 Results and discussion

3.1 Relationships between selectivity of ARG phase and other phases on the market.

The relationships between the various log k values measured on the ARG phase of the dataset of 36 the compounds in Table S3 and that of other commercially available stationary phases have been studied. It is noteworthy that none of the tested phases supports mixed-selectivity. An exemplative chromatogram is shown in Figure S1.

Results are shown in Table S3, which lists the r values of the correlation matrix. An *r* of -1 indicates a perfect negative linear relationship between variables, an *r* of 0 indicates no linear relationship between variables, and an *r* of 1 indicates a perfect positive linear relationship between variables. Consequently, $r = 0$ implies max orthogonality, which takes place when the separation mechanisms of each system are fully independent from each other.

Data in Table S3 supports that only ARG affinity values measured at 90/10 buffer 20 mM ammonium acetate/MeCN exhibit some degree of similarity of retention on other octadecylsilyl (ODS)- and IAM.SPH- based chromatographic systems. In fact, *r* values range between 0.65 and 0.70 when we consider ABZ and X-Bridge. This suggests that the analytical selectivities overlap to some extent. However, for most chromatographic systems and for all the other eluent compositions tested on the ARG phase, values very close to zero are observed. This implies that the features of the ARG phase are not depicted by any other chromatographic system among those tested, suggesting strong

orthogonality and, therefore, originality in the separation process afforded by this phase. Consistently, an extensive characterization of the analytical behaviour of the ARG phase was deemed relevant and hereby undertaken.

3.2 PCA analysis

To verify the dataset distribution, we performed PCA using the VS+ descriptors (see 2.6.3).

For the 36 neutral compounds, results show that more than 90% of the variance is explained by the first two principal components (Fig. S2A). The scores plot (Figure S2B) shows that the compounds are distributed in the four quadrants. For acidic compounds, more than 90% of the variance was found to be explained by the first two PCs (Figure S3A). The scores plot (Fig S3B) shows a good although not optimal compounds distribution. Similar results were obtained for basic compounds, which show an optimal distribution in the VS+ descriptors chemical space (Figure S4).

3.3 Dependency of analytical retention on ARG column upon organic modifier concentration.

Log k^{ARG} values of the 36 neutral compounds were isocratically obtained using mobile phases with MeCN content varying from 10 to 80%. ARG capacity factors (Table S4) evidence that retention values could be determined for most but not all the organic modifier concentrations, being a number of compound poorly retained over the 70/30 and 80/20 buffer/MeCN eluent compositions.

When monitoring the dependency of log k *versus* organic modifier concentration, most compounds could be classified in four classes according to the different trends (descending, minimum, snake and bell, Figure S5). However, we could not associate trends with the presence of common substructures and compounds showing similar trends are not always clustered in the same region of the PCA scores plot (Fig. S2B).

Capacity factors of 26 acids at eight concentrations of MeCN were also determined (Table S5). Acidic compounds were retained by the ARG column to a greater extent than neutral molecules, suggesting effectiveness of this phase in the retention and separation of anionic molecules. This is plausible, since the ARG phase bears a +1 total electric charge at the experimental pH (Figure 1). Figure 2, which reports log k^{ARG} of the 26 acidic compounds as a function of the mobile phase composition, suggests that most molecules feature a similar descending trend.

The calculated pK_a values, which are listed in Table S6, are in the 3-5 range for most monoprotic acids. This supports that these compounds interact with the ARG phase prevalently in their negatively charged forms. In fact, according to Henderson-Hasselbalch equation, they are in their undissociated:ionic form ratio to an extent spanning from 1 : 100 and 1: 10.000. The dataset does include some molecules featuring more than one acidic moiety *e.g.*, captopril, citric acid, furosemide, valsartan. However, the lowest pK_a value of these polyprotic acids lies again the in the same 3-5

range, supporting that also these compounds would interact with the ARG stationary phase preferentially in their anionic form.

Log k^{ARG} of the 38 basic compounds (Table S7) at eight concentrations of MeCN were also measured (Figure S6 shows the log k *vs* mobile phase composition plot). Evidently, a number of basic compounds were poorly retained or not retained at all. A reason supporting this analytical behaviour might be the occurrence of repulsive electrostatic interactions between these molecules and the ARG phase. Moreover, the negative charge of the carboxy group is located only onto the outer part of the phase, allowing only a superficial interaction and preventing these basic solutes to establish a deeper engagement with the ARG phase. However, the bases considered span a wider calculated pK_a range *i.e.*, 6-10 (Table S8) than the studied acidic compounds. This implies that not all the compounds would interact with the stationary phase prevalently in their cationic forms, but some in their neutral forms. Since the analytical retention of many bases could not be measured at various organic modifier concentrations, no solid evidence could be drawn with regards to trends. However, data collected indicates that ARG stationary phase may not be the best choice if one aims at separating a mixture of bases.

3.4. BR analysis

The three dataset subclasses (neutrals, acids and bases) were submitted to BR analysis.

3.4.1 Neutrals

An overview of the statistics of the final PLS models is shown in Table 1.

Table 1. PLS final models' overview. Legend: LV (number of latent variables chosen), R2 (goodness-of-fit measure), Q2 (statistical measure of the goodness of prediction of the model), RMSE_CV (how close the observed data points are to the model's predicted values), N (number of compounds in the model) with the compounds eliminated as outliers listed in the brackets

MeCN (v/v) %	***LV***	***R^2***	***Q^2***	***RMSE_CV***	***N (outlier)***
10	2	0.8769	0.5016	0.3536	27 (3,5-dichlorophenol, antipyrine, caffeine, diazepam, griseofulvin, nifuroxime)
20	3	0.8952	0.5726	0.2028	28 (hydrocortisone, prednisone, testosterone, tolnaftate)

30	2	0.7326	0.5053	0.4091	26 (bromazepam, lorazepam, lormetazepam, prednisolone)
40	2	0.6974	0.5698	0.2884	32 (hydrocortisone 21-acetate)
50	2	0.7188	0.5072	0.2532	30 (antipyrine)
60	2	0.7620	0.5103	0.2584	24 (3,5-dichlorophenol, antipyrine, hydrocortisone-21-acetate, paracetamol)

Notably, log k^{ARG} values achieved at 70% and 80% MeCN eluent compositions were not considered since many compounds were poorly retained and the size of their datasets was too limited to draw any solid conclusion. An analysis of the statistics listed in Table 1 suggest that more accurate models were achieved for ARG affinity values measured in prevalently aqueous eluents, and specifically in 20% MeCN. It is reasonable to assume that in prevalently aqueous medium the tendency of the ARG phase to ionise is greater than that in eluents richer in acetonitrile. Since all these solutes are neutrals, this does not affect dipole-dipole interactions, being that all the molecules are zero charge. However, this may well play a role in dipole-dipole induced interactions, which are magnified at lower organic modifier concentrations. Indeed, most neutrals support polar atoms, and consequently, polarized bonds.

Figure 3 shows the BR analysis graphical output for the retention data of the 36 neutrals over increasing concentration of acetonitrile. Overall, as schematized in Fig. 3G the plots show the relevance of any block to the model: the higher, the more important the block. Blocks which either show similar positive and negative contributions or are small (about less than 0.5) do not impact the investigated property. The reverse is true for large blocks.

Figure 3 highlights two major results. First, the Size block changes its sign from negative to positive when switching from 10% to 20% MeCN and then the sign remains positive over all the other concentrations. In reversed phase (RP), analytical retention is positively related with molecular size, being the bulkier the molecules, the more hydrophobic and consequently longer retained in the chromatographic system. In normal phase (NP) instead, the mechanism is specular, given that the stationary phase is hydrophilic and exhibits greater affinity for polar solutes. In this instance, molecular size contributes subtractively to the analytical retention, and the higher the molecular mass, the shorter the retention time in NP. The change in sign of the Size block highlighted by the BR

graphical output allows to visualize that the separation mechanism switches from NP to RP at MeCN concentrations > 10%. This is not a common behaviour among the stationary phases so far studied. The reason of the change in sign of the Size block can be attributed to the mixed-mode selectivity of the ARG phase. Indeed, mixed-mode phases have become increasingly popular in the last decades, [39] and the number of new mixed/multi-mode sorbents is growing fast. Unlike single-mode stationary phases, perfectly suited for the separation of the analytes possessing similar physicochemical properties, for instance reversed-phase chromatography for hydrophobic solutes, mixed-mode sorbents providing multimodal interactions can render better separation selectivity for complex mixtures of solutes differing significantly in their physicochemical characteristics, especially if performed in gradient elution programs. As Figure 1 displays, the ARG phase bears both polar and hydrophobic moieties and its folding might depend on the polarity of the mobile phase.

A not common behavior is also observed for the O (HBD solutes properties) block. However, this trend is unclear. In fact, the O block shows a high negative value for the 10% of CH_3CN. Conversely, in the presence of 20 and 30% MeCN, there is a high positive O block value. From 40% to 60% MeCN, the relevance of the O block is poor. Notably the N1 block, which represents the hydrogen bond acceptor (HBA) of the solute, shows a linear growth, from a negative value in the 10% of CH_3CN to a positive one for the 60%. This agrees with the evidence of higher likelihood of molecules to engage H-bonds in a medium that is prevalently aqueous. The three remaining blocks (OH2, DRY and Others) do not show any particular trend. The other results of BR analysis, including error distribution, absolute BR and experimental vs calculated plots are shown in Figures S7-S12.

3.4.2 Acids

PLS models for acidic compounds were not statistically significant and thus BR analysis could not be performed. Since PCA showed that acidic compounds are sufficiently well distributed in the PC1 vs PC2 chemical space (Fig. S3), we hypothesized that VS+ descriptors do not properly handle electric charges. For this reason, we resorted to a complimentary modelling approach capable of better modelling the retention of compounds that are prevalently ionized at the experimental pH (see 3.5).

3.4.3 Bases

As previously mentioned, the retention of basic compounds was often not experimentally accessible and thus PLS analysis was not performed.

3.5 MLR

3.5.1 Neutrals

The models with the highest predictive strength are listed in Table 2, along with their statistic validation. The plot predicted *vs* experimental log k^{ARG} of the best model is instead shown in Figure S13. The choice of the molecular descriptors (Table S1) operated by the script may incidentally provide valuable information about the nature of the interactions taking place between these neutrals and the stationary phase. Firstly, the change in the separation mode *i.e.*, from RP to NP, that occurs upon increasing the MeCN concentration can be observed also from this approach. Indeed, in Eq.(2), the lipophilicity based- descriptor, which is chemaxon log P, has positive sign, implying that the higher the log P, the longer the retention. However, from 40% to 60% MeCN, the lipophilicity based-descriptor, which is VirtualLogP, has in all cases a negative sign supporting that the degree of lipophilicity is inversely proportional to the analytical retention *i.e.*, log k^{ARG}. Another descriptor which is listed three times in the best models is HLB, which is the hydrophilic–lipophilic balance (HLB) and represents a measure of the partitioning tendency of surfactant between oil and water. This can be calculated according to either Griffin or David methods[40]. The aspect that ARG analytical retention sounds dependent on HLB seems reasonable as ARG is an ampholyte which exist at the experimental pH prevalently in a form supporting 1 negative and 2 positive charges. Therefore, it is plausible to assume that the solute having HLB similar to that the ARG phase are more retained in the chromatographic system. Consistently, the sign of HLB is always positive, except in one case. Interestingly, analytical retention of a wide (*n*=205*)* range of pharmaceutically relevant compounds was found to be similarly driven by HLB on the main IAM.PC phases *i.e.*, IAM.PC.MG and IAM.PC.DD2 in a recent study of ours[41]. These are similarly based on phosphatidylcholine (PC) but differ from each other in the end capping of the free aminopropyl groups, which is performed by reaction with either methyl glycolate (PC.MG) or with C_3 and C_{10} anhydrides (PC.DD2). The PC based analytical columns share with the ARG phase the amphiphilic character, as both these support electric charges of opposite sign. This is consistent also with data presented in Table S3, as a r^2 values equal to 0.64 was obtained when studying the correlation matrix between capacity factors of the neutrals on the ARG phase at 10% (v/v) MeCN and the capacity factors extrapolated to 100% aqueous phase on the IAM.PC.DD2 phase. These may suggest that the intermolecular forces involved in the separation mechanism overlap to some extent when the eluents are prevalently aqueous. The loss of correlation that takes place at higher organic modifier is reasonable as it is well-established[32] that the H-bonding and ionization is perturbed at lower dielectric constant of the medium. The definition of the other molecular descriptors is listed in Table S1.

391 *Table 2. Equations and statistics (SE = standard error and F = Fisher coefficients) of the models*
392 *achieved for the subgroup of neutrals.*
393

MeCN (v/v) % (Model equation)	Best optimized models (*n* – 1)	R^2	SE	F
10 (Eq 2)	-0.9241 + 0.2964 chemaxon log P + 0.0437 Davies HLB – 0.0230 Impropers	0.80	0.268	38.00
20 (Eq 3)	-0.6307 – 0.0179 Angles + 0.0490 Torsions + 0.0976 VirtualLogP	0.61	0.330	13.90
30 (Eq 4)	5.1548 – 5.5762 Ovality + 0.0668 polarizability +0.1201 Davies HLB	0.67	0.342	17.24
40 (Eq 5)	0.2936 – 0.0027 MSA AS+ - 0.0042 Angles – 0.2067 VirtualLogP	0.82	0.199	42.50
50 (Eq 6)	0.0121 – 0.1966 VirtualLogP – 0.0074 Griffin HLB – 0.0176 Atoms	0.83	0.152	41.68
60 (Eq 7)	-0.3652 + 0.0646 chemaxon log P – 0.3131 VirtualLogP – 0.0068 Angles	0.68	0.240	16.05
70 (Eq 8)	-1.1806 + 0.1514 chemaxon HLB – 0.0026 MSA AS	0.96	0.100	103.94

394
395
396 *3.5.2 Acids*
397 Predictive statistic models were achieved for acidic compounds and are listed in Table 3. Acids were
398 considered both in their undissociated and in their anionic forms. Eventually, though, a weighted
399 average of the physico-chemical descriptors according to the experimental pK_a values was performed.
400 This approach was the one that led to the development of models with the highest predictive strength,
401 which are the only ones that are discussed hereby. Some of the developed models feature rather high
402 r^2 (up to 0.94 in the best optimized models) values. Notably, all the models are based on Charge_WA,
403 which is the weighted average of the electric charge according to the pK_a values of the solutes. This
404 suggests that the retention of acidic compounds is heavily driven by electrostatics practically at any
405 eluent composition.
406 Another polarity- related descriptor that is selected by the script is dipole moment, however its role
407 seems to be much more marginal as it appears in only one model.
408 The enhanced selectivity of the ARG phase that was observed for the acids leaves much room for
409 several considerations. First, since retention of the acids seems to be greatly affected by the average
410 electric charge exhibited by the mixture of the species at the experimental pH (7.0), a hypothesis that
411 needs further studies is that a more elegant control of analytical retention could be better achieved by
412 modulating the pH of the eluent, rather than by playing with the organic modifier concentration. Of

course, the ARG phase offers superior stability when operated in a 3.0-7.5 range, so there are doubtlessly operational constrains when it comes to the pH ranges that could be accessed.

Another point that seems interesting is that the selectivity of the ARG phase could be better exploited in LC by using gradient elution programs *e.g.,* from 0 to 90% (v/v) MeCN. This would allow the sequential combination of opposite separation modes (RP and NP) which could in principle noticeably widen the separation window and provide usefulness in the determination of compounds with an ample range of ionization constants. Finally, since some of these models allow a rather accurate prediction, by using these equations it would be possible to assess to which extent a given acid is expected to be retained by the ARG phase, and since all the descriptors are calculated *in silico*, such assessment is feasible for hypothetical molecules or for compounds not yet synthesized as well. An exemplative experimental *vs* predicted log k^{ARG} plot is shown in Figure S14.

Table 3. Equations and statistics (SE = standard error and F = Fisher coefficients) of the models achieved for the subgroup of acids.

MeCN (v/v) % (Model number)	Best optimized models (*n* – 1)	R^2	SE	F
10 (Eq 9)	-0.9563 – 1.2784 Charge_WA + 0.2177 chemaxon log P	0.85	0.269	50.49
20 (Eq 10)	-1.3301 – 1.7797 Charge_WA + 0.0095 Dipole_WA	0.86	0.259	58.33
30 (Eq 11)	-0.5431 – 1.1330 Charge_WA – 0.0169 Impropers_WA	0.94	0.128	170.91
40 (Eq 12)	-0.5603 – 1.0259 Charge_WA – 0.0149 Impropers_WA	0.93	0.136	124.77
50 (Eq 13)	-0.6677 – 1.0481 Charge_WA – 0.0215 Impropers_WA	0.79	0.259	38.14
60 (Eq 14)	-0.6281 – 1.4040 Charge_WA – 0.2723 Rings	0.87	0.245	67.58
70 (Eq 15)	-0.7420 – 1.4714 Charge_WA – 0.2808 Rings	0.82	0.271	45.75

3.5.3 Bases

No regression was developed for bases due to the limited size of the bases featuring appreciable retention on the ARG column under the experimental conditions.

4.0 Conclusions

The Celeris™ Arginine is a mixed-mode stationary phase exhibiting a good degree of selectivity when compared with other stationary phases of common use. To characterize this phase, analytical retention of a pool of neutral, basic and acidic compounds was measured at various concentration of organic modifier. The phase was found to have greater affinity for molecules existing prevalently as anions at the experimental pH. Conversely, the retention of bases, albeit covering a wide range of both pK_a and lipophilicity, seemed to be quite limited. This may support that electrostatic interactions

of repulsive nature realize in solution between the ARG phase and the basic solutes, which both feature a positive overall charge.

BR analysis provided substantial assets in deconvoluting the overall analytical retention into its elementary blocks. Specifically, for the neutrals, the most relevant blocks were found to be Size, which describes the interaction due to the dimension of the compounds, and O, which represent the hydrogen bond donor (HBD) properties of the solute. The change in sign of the Size block allowed to visually appreciate the switch in the separation mode from RP – which occurs at 10% MeCN – to NP – which takes place at MeCN% > 10%. VS+ descriptors allowed the development of some good models for rationalizing the analytical behaviour of neutrals. However, their performance in modelling the analytical retention of acids was poor, probably due to their intrinsic inefficacy in handling electric charges. This was overcome by a complimentary MLR approach, which allowed to successfully model the retention of acids on the ARG column and to shed light into the retention mechanism of these compounds, which seemed to be overwhelmingly driven by electrostatics.

Overall, the ARG phase proved to exhibit a selectivity that is not straightforwardly offered by any other phase of common use and whose separation mode holds potential for applications in drug analysis.

REFERENCES

[1] H. Hoja, P. Marquet, B. Verneuil, H. Lotfi, B. Penicaut, G. Lachatre, Applications of liquid chromatography--mass spectrometry in analytical toxicology: a review, J Anal Toxicol 21(2) (1997) 116-26. https://doi.org/10.1093/jat/21.2.116.

[2] M. Wood, M. Laloup, N. Samyn, M. del Mar Ramirez Fernandez, E.A. de Bruijn, R.A. Maes, G. De Boeck, Recent applications of liquid chromatography-mass spectrometry in forensic science, J Chromatogr A 1130(1) (2006) 3-15. https://doi.org/10.1016/j.chroma.2006.04.084.

[3] A. Roux, D. Lison, C. Junot, J.F. Heilier, Applications of liquid chromatography coupled to mass spectrometry-based metabolomics in clinical chemistry and toxicology: A review, Clin Biochem 44(1) (2011) 119-35. https://doi.org/10.1016/j.clinbiochem.2010.08.016.

[4] A. Ahmad Dar, P.L. Sangwan, A. Kumar, Chromatography: An important tool for drug discovery, J Sep Sci 43(1) (2020) 105-119. https://doi.org/10.1002/jssc.201900656.

[5] F. Aucella, V. Lauriola, G. Vecchione, G.L. Tiscia, E. Grandone, Liquid chromatography-tandem mass spectrometry method as the golden standard for therapeutic drug monitoring in renal transplant, J Pharm Biomed Anal 86 (2013) 123-6. https://doi.org/10.1016/j.jpba.2013.08.001.

[6] Y. Zhang, R. Zhang, Recent advances in analytical methods for the therapeutic drug monitoring of immunosuppressive drugs, Drug Test Anal 10(1) (2018) 81-94. https://doi.org/10.1002/dta.2290.

[7] P. Masson, Quality control techniques for routine analysis with liquid chromatography in laboratories, J Chromatogr A 1158(1-2) (2007) 168-73. https://doi.org/10.1016/j.chroma.2007.03.003.

[8] H. Qiu, X. Liang, M. Sun, S. Jiang, Development of silica-based stationary phases for high-performance liquid chromatography, Anal Bioanal Chem 399(10) (2011) 3307-22. https://doi.org/10.1007/s00216-010-4611-x.

[9] D. Sykora, J. Vozka, E. Tesarova, Chromatographic methods enabling the characterization of stationary phases and retention prediction in high-performance liquid chromatography and supercritical fluid chromatography, J Sep Sci 39(1) (2016) 115-31. https://doi.org/10.1002/jssc.201501023.

[10] C. West, S. Khater, Characterization of retention and separation mechanisms with Pirkle-type enantioselective stationary phases in supercritical fluid chromatography, J Chromatogr A 1626 (2020) 461352. https://doi.org/10.1016/j.chroma.2020.461352.

[11] R. Technologies, Celeris Arginine stationary phase, 2021. https://www.registech.com/celeris-arginine-2/.

[12] S. Wu, X. Li, F. Zhang, G. Jiang, X. Liang, B. Yang, An arginine-functionalized stationary phase for hydrophilic interaction liquid chromatography, Analyst 140(12) (2015) 3921-4. https://doi.org/10.1039/c5an00570a.

[13] M.J. Kamlet, R.M. Doherty, J.L. Abboud, M.H. Abraham, R.W. Taft, Linear solvation energy relationships: 36. Molecular properties governing solubilities of organic nonelectrolytes in water, J Pharm Sci 75(4) (1986) 338-49. https://doi.org/10.1002/jps.2600750405.

[14] L.R. Snyder, J.W. Dolan, P.W. Carr, The hydrophobic-subtraction model of reversed-phase column selectivity, J Chromatogr A 1060(1-2) (2004) 77-116.

[15] K. Kimata, K. Iwaguchi, S. Onishi, K. Jinno, R. Eksteen, K. Hosoya, M. Araki, N. Tanaka, Chromatographic Characterization of Silica C18 Packing Materials. Correlation between a Preparation Method and Retention Behavior of Stationary Phase, Journal of Chromatographic Science 27(12) (1989) 721-728. https://doi.org/10.1093/chromsci/27.12.721.

[16] Z. Liu, D.G. Patterson, M.L. Lee, Geometric Approach to Factor Analysis for the Estimation of Orthogonality and Practical Peak Capacity in Comprehensive Two-Dimensional Separations, Analytical Chemistry 67(21) (1995) 3840-3845. https://doi.org/10.1021/ac00117a004.

[17] M.J. den Uijl, P.J. Schoenmakers, B.W.J. Pirok, M.R. van Bommel, Recent applications of retention modelling in liquid chromatography, J Sep Sci 44(1) (2021) 88-114. https://doi.org/10.1002/jssc.202000905.
[18] G. Russo, G. Ermondi, G. Caron, D. Verzele, F. Lynen, Into the first biomimetic sphingomyelin stationary phase: Suitability in drugs' biopharmaceutic profiling and block relevance analysis of selectivity, Eur J Pharm Sci 156 (2021) 105585. https://doi.org/10.1016/j.ejps.2020.105585.
[19] G. Ermondi, G. Caron, Molecular interaction fields based descriptors to interpret and compare chromatographic indexes, J Chromatogr A 1252 (2012) 84-9. https://doi.org/10.1016/j.chroma.2012.06.069.
[20] G. Ermondi, A. Visconti, R. Esposito, G. Caron, The Block Relevance (BR) analysis supports the dominating effect of solutes hydrogen bond acidity on DeltalogP(oct-tol), Eur J Pharm Sci 53 (2014) 50-4. https://doi.org/10.1016/j.ejps.2013.12.007.
[21] G. Ermondi, G. Caron, Block relevance (BR) analysis and polarity descriptors in property-based drug design, ADMET and DMPK 6(3) (2018). https://doi.org/10.5599/admet.532.
[22] M. Vallaro, G. Ermondi, G. Caron, Chromatographic HILIC indexes to characterize the lipophilicity of zwitterions, Eur J Pharm Sci 145 (2020) 105232. https://doi.org/10.1016/j.ejps.2020.105232.
[23] A. Pedretti, A. Mazzolari, S. Gervasoni, L. Fumagalli, G. Vistoli, The VEGA suite of programs: an versatile platform for cheminformatics and drug design projects, Bioinformatics (2020). https://doi.org/10.1093/bioinformatics/btaa774.
[24] J. Gasteiger, M. Marsili, Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges, Tetrahedron 36(22) (1980) 3219-3228. https://doi.org/http://dx.doi.org/10.1016/0040-4020(80)80168-2.
[25] G. Ermondi, M. Vallaro, G. Goetz, M. Shalaeva, G. Caron, Experimental lipophilicity for beyond Rule of 5 compounds, Future Drug Discovery 1(1) (2019) FDD10. https://doi.org/10.4155/fdd-2019-0002.
[26] G. Ermondi, F. Catalano, M. Vallaro, I. Ermondi, M.P.C. Leal, L. Rinaldi, S. Visentin, G. Caron, Lipophilicity of amyloid beta-peptide 12-28 and 25-35 to unravel their ability to promote hydrophobic and electrostatic interactions, Int J Pharm 495(1) (2015) 179-185. https://doi.org/10.1016/j.ijpharm.2015.08.075.
[27] G. Ermondi, M. Vallaro, G. Caron, Learning how to use IAM chromatography for predicting permeability, Eur J Pharm Sci 114 (2018) 385-390. https://doi.org/10.1016/j.ejps.2018.01.001.
[28] D. Verzele, F. Lynen, M. De Vrieze, A.G. Wright, M. Hanna-Brown, P. Sandra, Development of the first sphingomyelin biomimetic stationary phase for immobilized artificial membrane (IAM) chromatography, Chem Commun (Camb) 48(8) (2012) 1162-4. https://doi.org/10.1039/c2cc16872c.
[29] A. Pedretti, L. Villa, G. Vistoli, VEGA--an open platform to develop chemo-bio-informatics applications, using plug-in architecture and script programming, J Comput Aided Mol Des 18(3) (2004) 167-73.
[30] P. Gaillard, P.A. Carrupt, B. Testa, A. Boudon, Molecular lipophilicity potential, a tool in 3D QSAR: method and applications, J Comput Aided Mol Des 8(2) (1994) 83-96.
[31] A. Pedretti, L. Villa, G. Vistoli, Modeling of binding modes and inhibition mechanism of some natural ligands of farnesyl transferase using molecular docking, J Med Chem 45(7) (2002) 1460-5.
[32] G. Russo, L. Grumetto, R. Szucs, F. Barbato, F. Lynen, Determination of in Vitro and in Silico Indexes for the Modeling of Blood-Brain Barrier Partitioning of Drugs via Micellar and Immobilized Artificial Membrane Liquid Chromatography, J Med Chem 60(9) (2017) 3739-3754. https://doi.org/10.1021/acs.jmedchem.6b01811.
[33] S. Kim, P.A. Thiessen, E.E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B.A. Shoemaker, J. Wang, B. Yu, J. Zhang, S.H. Bryant, PubChem Substance and Compound databases, Nucleic Acids Res 44(D1) (2016) D1202-13. https://doi.org/10.1093/nar/gkv951.

554 [34] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B.A. Shoemaker, P.A. Thiessen,
555 B. Yu, L. Zaslavsky, J. Zhang, E.E. Bolton, PubChem 2019 update: improved access to chemical
556 data, Nucleic Acids Res 47(D1) (2019) D1102-D1109. https://doi.org/10.1093/nar/gky1033.
557 [35] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus,
558 CHARMM: A program for macromolecular energy, minimization, and dynamics calculations,
559 Journal of Computational Chemistry 4(2) (1983) 187-217. https://doi.org/10.1002/jcc.540040211.
560 [36] A.D. MacKerell, B. Brooks, C.L. Brooks, L. Nilsson, B. Roux, Y. Won, M. Karplus, CHARMM:
561 The Energy Function and Its Parameterization, Encyclopedia of Computational Chemistry, John
562 Wiley & Sons, Ltd2002. https://doi.org/10.1002/0470845015.cfa007.
563 [37] B.R. Brooks, C.L. Brooks, 3rd, A.D. Mackerell, Jr., L. Nilsson, R.J. Petrella, B. Roux, Y. Won,
564 G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer,
565 J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor,
566 C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M.
567 York, M. Karplus, CHARMM: the biomolecular simulation program, J Comput Chem 30(10) (2009)
568 1545-614. https://doi.org/10.1002/jcc.21287.
569 [38] R.W. Harrison, Stiffness and energy conservation in molecular dynamics: An improved
570 integrator, Journal of Computational Chemistry 14(9) (1993) 1112-1122.
571 https://doi.org/10.1002/jcc.540140912.
572 [39] D. Sýkora, P. Řezanka, K. Záruba, V. Král, Recent advances in mixed-mode chromatographic
573 stationary phases, Journal of Separation Science 42(1) (2019) 89-129.
574 https://doi.org/https://doi.org/10.1002/jssc.201801048.
575 [40] R.C. Pasquali, M.P. Taurozzi, C. Bregni, Some considerations about the hydrophilic–lipophilic
576 balance system, Int J Pharmaceut 356(1) (2008) 44-51.
577 https://doi.org/https://doi.org/10.1016/j.ijpharm.2007.12.034.
578 [41] G. Russo, L. Grumetto, F. Barbato, G. Vistoli, A. Pedretti, Prediction and mechanism elucidation
579 of analyte retention on phospholipid stationary phases (IAM-HPLC) by in silico calculated physico-
580 chemical descriptors, Eur J Pharm Sci 99 (2017) 173-184. https://doi.org/10.1016/j.ejps.2016.11.026.
581