analytical chemistry

Article

pubs.acs.org/ac

Physicochemical Parameters Affecting the Electrospray Ionization Efficiency of Amino Acids after Acylation

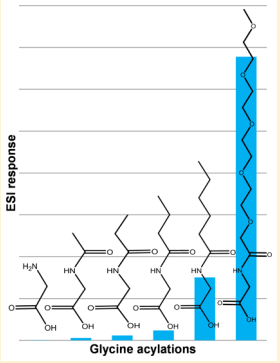
Jos Hermans,[†] Sara Ongay,[†,§] Vadym Markov,[‡] and Rainer Bischoff*[,†]

[†]Analytical Biochemistry, Department of Pharmacy, University of Groningen, Antonius Deusinglaan 1, 9713AV Groningen, The Netherlands

[‡]Department of Chemical Metrology, Kharkov V. N. Karazin National University, Svoboda Square 4, 61022 Kharkov, Ukraine

Supporting Information

ABSTRACT: Electrospray ionization (ESI) is widely used in liquid chromatography coupled to mass spectrometry (LC−MS) for the analysis of biomolecules. However, the ESI process is still not completely understood, and it is often a matter of trial and error to enhance ESI efficiency and, hence, the response of a given set of compounds. In this work we performed a systematic study of the ESI response of 14 amino acids that were acylated with organic acid anhydrides of increasing chain length and with poly(ethylene glycol) (PEG) changing certain physicochemical properties in a predictable manner. By comparing the ESI response of 70 derivatives, we found that there was a strong correlation between the calculated molecular volume and the ESI response, while correlation with hydrophobicity (log *P* values), pK_a, and the inverse calculated surface tension was significantly lower although still present, especially for individual derivatized amino acids with increasing acyl chain lengths. Acylation with PEG containing five ethylene glycol units led to the largest gain in ESI response. This response was maximal independent of the calculated physicochemical properties or the type of amino acid. Since no actual physicochemical data is available for most derivatized compounds, the responses were also used as input for a quantitative structure−property relationship (QSPR) model to find the best physicochemical descriptors relating to the ESI response from molecular structures using the amino acids and their derivatives as a reference set. A topological descriptor related to molecular size (SPAN) was isolated next to a descriptor related to the atomic composition and structural groups (BIC0). The validity of the model was checked with a test set of 43 additional compounds that were unrelated to amino acids. While prediction was generally good (R^2 > 0.9), compounds containing halogen atoms or nitro groups gave a lower predicted ESI response.



Electrospray ionization (ESI) is the most commonly used ionization technique in combination with liquid chromatography−mass spectrometry (LC−MS) for the identification and quantification of a large variety of compounds in a wide range of areas (e.g., biochemical, environmental, and food analysis). Sensitivity of mass spectrometers has improved over the last decades due to better ion transmission using new features like ion funnels, orthogonal spray sources, and/or simply a bigger orifice diameter in combination with more powerful vacuum pumps. Further enhancements were realized by reducing LC flow rates down to nanoflow conditions where electrospray ionization efficiency is greatly improved.[1,2] Nevertheless, ESI-MS responses can differ by many orders of magnitude for a set of compounds while measured under the same conditions.[3] Thus, many researches have tried to understand the underlying reasons, finally leading to the postulation of four theoretical models that assist in understanding ESI response differences: the charge residual model (CRM),[4] the ion evaporation model (IEM),[5] the equilibrium partition model (EPM),[6] and the chain ejection model (CEM).[7] In general, during ESI a liquid flow is led through a needle to which a voltage of 2−3 kV is applied, resulting in charged droplets and producing ionized molecules due to charge-transfer reactions. Next to this, a drying gas leads to solvent evaporation and the subsequent reduction in initial droplet diameter. The CRM, first proposed by Dole et al.,[4] postulates that charge is concentrated on the droplet surface where it can be transferred to neutral or already charged molecules. According to this theory, the charged droplet leaving the needle will form a so-called Taylor cone followed by a fine stream (jet) of tiny droplets depending on the drying gas flow and temperature, the solvent/solution surface tension, and conductivity. As solvent continues to evaporate, charged droplets continue to decrease in size concentrating the charged molecules until a critical point (the Rayleigh limit) is reached. At this point, the repulsive power of all concentrated ions of the same charge results in an explosion of the droplet into many smaller droplets, also known as Coulomb fission and visualized by Duft et al.[8] This process repeats itself until there is no solvent left and all molecules are evaporated into the gas phase. This theory is thought to be most valid for large molecules like intact/unfolded proteins.[9] For small ions the IEM proposed by Iribarne and Thomson[5] is more generally accepted. In this model ions evaporate into the gas phase during the solvent

Received: May 19, 2017
Accepted: July 24, 2017
Published: July 24, 2017

DOI: 10.1021/acs.analchem.7b01899
Anal. Chem. 2017, 89, 9159−9166

evaporation process due to repulsion between ions of the same charge as soon as the droplets are sufficiently small (<10 nm). This phenomenon reduces and even replaces the fission process described by the CRM, since the evaporation rate of ions becomes faster than the evaporation rate of solvent, preventing droplets to reach the Rayleigh limit.[5] It is likely that both processes coexist in many cases. The EPM considers equilibrium between ions in the droplet interior and on the charged surface explaining the often better responses for surface-active compounds. The CEM, recently postulated by Konerman et al.[7] as a fourth mechanism, implies that large multiply charged polymers and denatured (unfolded) proteins are present near the charged droplet surface due to their hydrophobic character. Consequently, they are pulled out of the droplet by repulsive electrostatic forces while acquiring charge from the droplet prior to droplet fission. A recent review about ESI processes by Banerjee and Mazumdar gives an overview of the current theories.[10]

Although the formation of ions with electrospray is still not fully understood, the models described above indicate parameters that might influence the electrospray response. These can be distinguished in instrumental, solution, and compound-related parameters or properties. Differences in ionization efficiency between instruments are mainly due to the design, which cannot be influenced by the user and, once chosen, is fixed. The effect of solution conditions mainly relates to solvents and their additives like formic acid or ammonia as well as to the composition of the sample matrix, affecting parameters like pH, surface tension, and ionic strength. A complex sample matrix may cause suppression of the ESI response due to a competition for charge between the different substances in solution near the limited droplet surface, a well-known disadvantage of this technique.[11,12] Suppression can often be reduced by sample cleanup and/or a liquid chromatographic separation.[13] Another response-affecting phenomenon often seen with ESI and related to the choice of solvent/additive, is adduct formation (e.g., Na^+, K^+, and NH_4^+) which may coexist next to protonated ions. The formation of adducts is highly component-, matrix-, as well as instrument-related but can often be reduced by spraying under acidic conditions using additives like formic or acetic acid or their ammonium salts[14] or masked using metal-chelating additives.[15] Nonetheless, adduct formation is often difficult to avoid. As adducts are also ions originating from the compounds under investigation, they should be accounted for in ESI efficiency calculations. Identical arguments hold with respect to the presence of dimers (2M + H/Na) and in-source fragments, which can coexist with adducts. The former may occur especially for compounds having a good ESI response at concentrations >10 μM due to ion-saturated droplet surfaces,[12,16] while the latter mostly indicate nonoptimal source parameters.

Finally, the ESI response may be affected by solution- and compound-related physicochemical properties like surface tension/activity, conductivity, hydrophobicity (log P), ionizability (pK_a), and molecular volume. Numerous articles have focused on ESI response enhancements due to different solvent compositions and with different additives,[17−19] or by changing the physicochemical properties of compounds through derivatization,[20−23] or using peptides differing in only one amino acid residue,[24] while others tried to find ESI response relations from a diverse set of compounds with orders of magnitude response differences at fixed instrumental conditions.[3] Here we focus on a fixed set of amino acids that overall are not very responsive to ESI and study the response enhancement after derivatization in order to change compound properties in a systematic manner. We used acylation with acid anhydrides of increasing hydrophobicity as well as with poly(ethylene glycol) (PEG) based on previous work[21] and investigated the relation between response and a number of physicochemical parameters. As these experiments were performed by reversed-phase LC−MS (RPLC−MS), the effect of solvent composition on the response was investigated first using flow injection analysis (FIA) of individually labeled amino acids.

Since finding relevant relationships between the electrospray response and physicochemical parameters may be hampered by a lack of reported data, in particular for the derivatives, we developed a quantitative structure−property relationship (QSPR)[25,26] model to elucidate descriptors and finally to predict electrospray ionization efficiencies from molecular structures.

EXPERIMENTAL SECTION

Reagents and Materials. Solutions (100 mM) of each of the 20 natural amino acids histidine (H), lysine (K), asparagine (N), arginine (R), aspartic acid (D), serine (S), glutamine (Q), glycine (G), threonine (T), glutamic acid (E), alanine (A), proline (P), cysteine (C), tyrosine (Y), methionine (M), valine (V), isoleucine (I), leucine (L), phenylalanine (F), and tryptophan (W) (Sigma; Zwijndrecht, The Netherlands) were prepared in 0.1 M sodium hydroxide (Sigma; Zwijndrecht, The Netherlands) and stored at −20 °C. C and N were prepared at 50 mM, D, E, and W at 10, and Y at 1 mM due to limited solubility. These standards were diluted further individually as well as in a mixture to a final concentration of 50 μM in 0.1 M sodium phosphate buffer, pH 8.

The 50 mM solutions of aniline, *p*-toluidine, 4-chloroaniline, 4-nitroaniline, cyclohexylamine, 4-aminobenzoic acid, 2-amino-5-bromobenzoic acid, and L-(+)-α-phenylglycine (all Sigma; Zwijndrecht) were prepared in 0.1 M sodium hydroxide and further diluted to 50 μM in 0.1 M sodium phosphate buffer pH 8. Individual solutions as well as a mixture including A, H, and W were stored at −20 °C until use.

Labeling Procedure. Amounts of 100 μL of the 50 μM mixtures were labeled with 1 μmol of the $^{13}C_1$-pentafluorophenyl-activated ester of a poly(ethylene glycol) derivative (PEG-OPFP, synthesized as described in Abello et al.[21]) containing five ethylene glycol units or with 20, 15, 12, and 9 μmol (2 μL each) of acetic, propionic, butyric, or hexanoic acid anhydride (Sigma; Zwijndrecht, The Netherlands), respectively, for 15 min in a thermomixer (Eppendorf; Hamburg, Germany) at 450 rpm and 25 °C. Esterification of the phenolic OH group of Y was reversed by incubating the sample for 15 min at 99 °C with vortexing (450 rpm). Finally, formic acid (Merck KGaA; Darmstadt, Germany) was added to the samples to a final concentration of 1% (v/v) prior to analysis by LC−MS. The nonderivatized compounds were injected individually to avoid ionization suppression.

Instrumental Parameters. The HPLC part of the analytical system consisted of an Agilent series 1100 capillary LC system (Waldbronn, Germany) comprising a degasser, a binary pump with stream splitter and flow controller (50 μL/min), a thermostated autosampler (4 °C), and a thermostated column compartment (40 °C). Compounds were chromatographically separated with an Atlantis dC18 column (Waters;

DOI: 10.1021/acs.analchem.7b01899
Anal. Chem. 2017, 89, 9159−9166

Etten-Leur, The Netherlands; 1.0 mm × 150 mm, particle size 3 μm). Mobile phase A consisted of 0.1% (v/v) formic acid in ultrapure water. Mobile phase B was 0.1% (v/v) formic acid in acetonitrile (HPLC-S gradient grade; Biosolve; Valkenswaard, The Netherlands). Injection volume was 1 μL. The separation was performed starting at 2% B and a gradient of 0.5% B/min starting right after injection up to 37% B followed by a 2 min gradient to 90% B to regenerate the column for 8 min before going back to 2% B in 2 min and an 8 min final conditioning step.

Analytes were measured in a Bruker HCT ion trap mass spectrometer (Bruker Daltonik GmbH; Bremen, Germany) equipped with an ESI source operated in positive mode. MS data were acquired over a mass-to-charge (m/z) scan range of 100−700 or 52−700 in single-MS mode for the labeled and nonlabeled components, respectively. Skimmer voltage was set to 30. Capillary exit and trap drive settings were adjusted by the smart parameter settings (SPS) using m/z 150, 200, and 350 as average optimum for non-, acyl-, and PEG-labeled compounds, respectively. The nebulizer and drying gas flows were set to 20 and 7 L/min, respectively, using a source temperature of 250 °C. Extracted ion chromatograms (EICs) were retrieved with a tolerance of ±0.5 Da. To compensate for day-to-day variations, all ESI responses to be used in the QSPR model were normalized to the response of butanoylated W measured in the same series using peak areas.

Fractionation and Flow Injection. Fractionation of PEG-labeled amino acids was performed manually employing the LC method described above but injecting 10 μL. Up to 15 50 μL fractions were collected, evaporated until dryness using a Concentrator 5301 at 30 °C (Eppendorf; Hamburg, Germany), and reconstituted in 100 μL of 20% acetonitrile (HPLC-S gradient grade; Biosolve; Valkenswaard, The Netherlands). Due to the noncomplete LC separation some of the fractions contained more than one PEG-labeled amino acid. Flow injection analysis was implemented on the LC−MS system described above without the use of a column. In each analysis, five injections of 1 μL of sample were preceded and followed by a 1 μL injection of a blank (ultrapure water containing 20% acetonitrile and 0.1% formic acid). The mobile phase consisted of 3%, 10%, 20%, 40%, 60%, or 80% acetonitrile and 0.1% (v/v) formic acid in ultrapure water at a flow rate of 50 μL/min.

Physicochemical and QSPR Parameters. As most physicochemical parameters of the derivatives cannot be derived from literature, they were calculated from molecular structures using the simplified molecular input line entry system (SMILES). The freely available software tool Chemicalize (https://chemicalize.com) was used to calculate molecular volumes based on three-dimensional (3D) structures and van der Waals radii. KOWWIN, included in the EPI suite (v411, U.S. EPA; Washington, DC, U.S.A.), was used to calculate log P values using atom/fragment contribution methods. Marvinsketch (v16.10.3, ChemAxon; Budapest, Hungary) was employed to calculate pK_a values based on physicochemical parameters obtained from ionization site-specific regression equations. For commercially produced chemicals many of these parameters were obtained from the Chemspider Web site (http://www.chemspider.com/),[27] which was used to access surface tensions, as predicted by the Physchem module in the Percepta platform (ACD/Laboratories; Toronto, ON, Canada), based on molecular weight, volume, and density. Chemoffice ChemBio3D Ultra 12.0 (PerkinElmer; Waltham, MA, U.S.A.) was used for 3D molecular structure optimization needed for descriptor calculations by the Dragon 5.5 software (Talette SRL; Milan, Italy). After applying a cross-correlation matrix of descriptors as a primary filter, a fuzzy forward search algorithm was used to isolate the most relevant descriptors as explained in more detail in the Supporting Information. This algorithm was developed using Python 3 in conjunction with the Pandas, Numpy, Scipy, and Sklearn libraries for mathematic computations (Python Software Foundation, https://www.python.org/).

RESULTS AND DISCUSSION

Effect of Instrumental Parameters. Instrumental parameters of the mass spectrometer may have a direct effect on the transmission of ions and thus on the measured signal. Some of these, like the capillary exit voltage and the trap drive value, are mass-dependent. The former affects in-source fragmentation as well as desolvation and declustering, while the latter is the radio frequency field strength needed for optimal trapping. Both values are defined by the smart parameter settings (SPS) for an on average optimal transmission around a specific m/z value. The skimmer voltage affects transmission as well as in-source fragmentation, the latter being compound-dependent and not affected by the SPS settings. Varying these parameters over a range from m/z 76 to 922 resulted in rather Gaussian optimization curves with broad tops of at least 10 V for the skimmer voltage as well as the capillary exit voltage, while the trap drive showed optimization curves with a slightly negative slope (see Supporting Information Figure S-1 for more details). This means that the optima, as defined by the SPS algorithm, were also valid for ions with m/z values in a window of about ±50% allowing the use of average SPS settings of m/z 150, 200, and 350 for the non-, acyl-, and PEG-labeled amino acids, respectively.

The skimmer voltage was set to 30, close to the individual optima for all compounds. No in-source fragmentation was observed at this value. In that way transmission efficiency differences due to differences in m/z values were always within ±20%, which is negligible with respect to the orders of magnitude of difference seen after derivatization. This allowed examining ionization efficiency differences independently of small effects related to ion transmission from ion source to detector.

Effect of Acetonitrile Concentration. As the solvent composition is often reported to affect electrospray ionization efficiency and a solvent gradient is needed for proper separation of compounds from a mixture using RPLC, we studied the influence of the acetonitrile concentration on ionization efficiency by FIA of PEG-labeled amino acids, since they provided the highest gain in sensitivity, showed high retention, and consequently, should be strongly affected by this parameter. The mixtures were first fractionated and analyzed as individual PEG-labeled amino acids, minimizing confounding effects due to ionization suppression or adduct formation due to buffer components or residual reagent byproducts as described under the Experimental Section. Figure 1 shows the total ion chromatogram traces of PEG-labeled amino acids. As K contains two reactive amino groups, this resulted in the presence of three products indicated as K′, K*, and K′* for singly and doubly PEG-labeled K, respectively. Because of these coexisting products and the coelution of PEG-labeled K* and PEG-labeled Q having almost the same masses, both amino acids were excluded from further study. A few additional components had to be excluded from the study as well for the

DOI: 10.1021/acs.analchem.7b01899
Anal. Chem. 2017, 89, 9159−9166

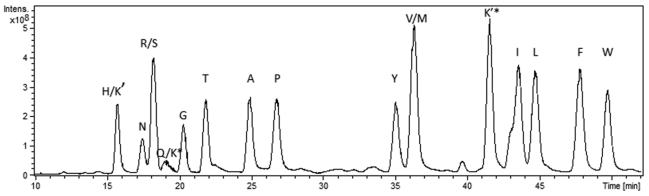


Figure 1. Total ion chromatogram of 50 pmol PEG-labeled amino acids.

following reasons. Y was excluded, because O-acylation was only completely reversed by heating for the PEG-labeled amino acid but not for acylated Y, C was excluded because of its potential for oxidation, and D and E because they appeared to be unstable in the stock solutions at −20 °C.

For this study we collected seven chromatographically well-resolved PEG-labeled amino acid derivatives (A, P, Y, I, L, F, and W) for FIA, while others were collected in one fraction (H and K′; N, R and S; G and D; T and E; V and M; K*, Q, and S). The mass spectra of most flow-injected PEG-labeled amino acids showed abundant sodium adducts of up to 50% of the protonated ion for F against less than 5% when using chromatography. Abundant sodium adducts are commonly observed in ESI mass spectrometry, especially in the absence of chromatography in front of the mass spectrometer.[22] They occur due to the presence of sodium ions in the sample and/or solvent or due to a contaminated ion source. We assume that in this case sodium ions were introduced from the glass vials in the auto sampler prior to FIA. It is striking that the first seven eluting PEG-labeled amino acids did not show any sodium adducts upon FIA, while the rest of the PEG-derivatized amino acids showed a slight increase in sodium adducts with increasing acetonitrile concentration (see Figure S-2 in Supporting Information). In this work we included adduct ions when their response contribution was >20%. In-source fragments were not detected at levels higher than 10% of the base peak and were therefore disregarded.

The effect of the acetonitrile concentration on the response of PEG-labeled amino acids using the combined signals from protonated and sodium-adduct ions ranged from no influence to a maximum increase by a factor of 2.5 [see Figure S-3 in Supporting Information for representative traces of PEG-labeled I (strongest effect), PEG-labeled H (representative effect seen for all PEG-labeled amino acids) and PEG-labeled S (no effect)]. This appeared also to be valid when looking at protonated and sodium adduct ions separately. We conclude that the effect of a changing acetonitrile concentration during gradient elution on the ESI response is no bigger than 3-fold even when comparing results at 3% and 80% acetonitrile, respectively. As all labeled amino acids elute between 7% and 27% acetonitrile, we estimate that the effect of acetonitrile on the ESI response during RPLC is a factor of 2 at the most, which is much smaller than the observed effects due to derivatization.

These results are in agreement with findings by others, who showed that there is only a relatively small dependence of the Rayleigh limit on the type of solvent.[9] Comparing ESI responses after chromatographic separation or FIA at 20% acetonitrile showed that the ESI response after chromatography is on average 2.5 ± 1 times higher (see Figure S-4 in Supporting Information). This is likely due to removal of ionic matrix components by chromatography leading to less ion suppression. These comparisons showed that it is justified to compare ESI responses after separation by gradient RPLC using acetonitrile as organic modifier and to assume that between compound response differences bigger than a factor 2 are primarily related to compound properties rather than solvent properties or m/z-dependent differences in ion transmission.

Effect of Derivatization. Derivatization of amino acids with acyl chains of increasing length increases the hydrophobicity of the resulting derivatives as also indicated by an increased retention time upon RPLC (see Figure S-5 in the Supporting Information). The PEG-labeled amino acids have lower retention times than their corresponding C6-labeled counterparts indicating that they are less hydrophobic.

Figure 2 shows that the response of almost all amino acids steadily increases with the length of the acyl chain. PEG

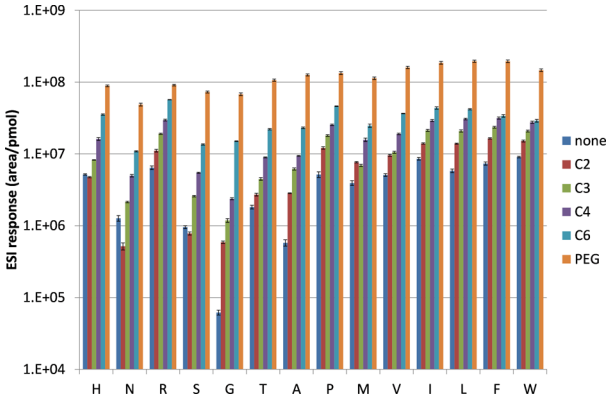


Figure 2. Effect of derivatization of the amino group with carboxylic acids with acyl chain lengths ranging from 2 to 6 CH_2 groups in comparison to PEG-labeling with PEG containing 5 (CH_2CH_2)−O− moieties. Average reproducibility was 15 ± 6% (n = 3).

derivatization gives the largest response for any amino acid demonstrating that hydrophobicity alone cannot explain the gain in ESI response. In fact, PEG derivatization of G increases the ESI response by 3 orders of magnitude. The observed gain in ESI response for PEG-labeled W of a factor 12 corresponds well to the previously reported factor 17 by Abello et al.[21] using a similar LC−MS system. Not all results were in line with the expected response gain. H, N, and S showed a reduced response for the C2 label, which may be attributed to their low retention and consequently poor separation from the buffer matrix as well as from each other, leading to ion suppression.

As shown above, only response enhancement factors up to 2 can be explained by the effect of differences in solvent composition across the applied acetonitrile gradient and retention time of the analytes. This shows that the observed differences are truly related to the molecular properties of the derivatized amino acids. It is thus of interest to gain a better understanding how physicochemical molecular properties affect the ESI response.

Physicochemical Molecular Properties Affecting the ESI Response. The existence of at least 38 different published hydrophobicity scales for amino acids indicates the complex nature of this parameter which tries to combine many physicochemical properties in a single value. Amino acid hydrophobicity scales can roughly be divided into two classes, one based on the physicochemical properties of the individual amino acids and the other one based on their characteristics as part of proteins,[28] the latter being of less relevance to the

DOI: 10.1021/acs.analchem.7b01899
Anal. Chem. 2017, 89, 9159−9166

current work. Hydrophobicity is also often expressed as the distribution coefficient of a compound between water and *n*-octanol also known as the log K_{ow} or log *P* value. Although there is considerable scattering when looking at the entire data set, correlations between log *P* and ESI response become clearer when grouping the data according to the individual labels (Figure 3) and are even more striking when looking at

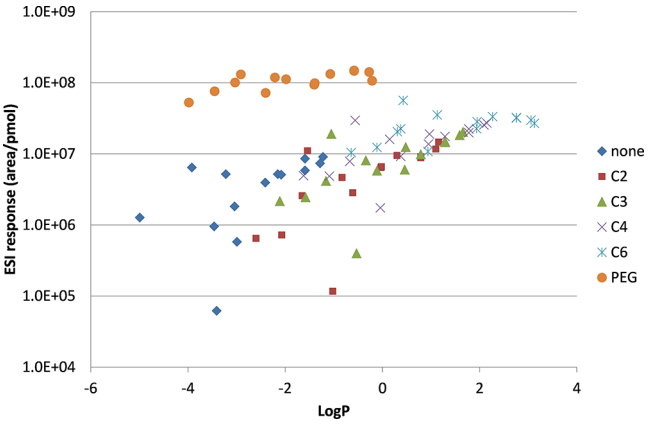


Figure 3. Electrospray response in relation to the log *P* of derivatized amino acids with respect to the length of the acyl chain and PEG moiety.

individual amino acids (see Figure S-6 in Supporting Information). The correlation with retention time instead of log *P* is even stronger, indicating that this parameter reflects the effect on ESI response better for the current set of compounds (see Figure S-7 and Table S-2 in Supporting Information). Correlations are comparable with average correlation coefficients of 0.81 ± 0.11 when the PEG-labeled compounds are excluded (see Table S-2 in Supporting Information). The retention on a reversed-phase column is known to be strongly correlated to ESI response[29] as well as log *P*,[30] although structurally diverse compounds may not always show a clear correlation. Especially R, H, and G show different patterns compared to the other acyl-labeled amino acids, notably when plotted against retention time (Figure S-7 in Supporting Information). G has a steeper slope indicating a major effect of acylation on ESI response. This may be related to the very low ESI response of nonderivatized G and the absence of a side chain in this molecule. The higher responses of R and H prior to acylation may be related to the presence of additional positively charged nitrogen atoms in their side chains resulting in a high proton affinity. It is noteworthy that PEG-labeled amino acids give the highest ESI response independent of their calculated log *P* value or retention time, forming a separate cluster as depicted in Figure 3. This indicates that the introduction of a PEG moiety having hydrophobic as well as hydrophilic properties may be a generic derivatization approach to enhance the ESI response. This was further studied with another set of compounds later on.

As the compounds used in this study are not volatile, they will be concentrated in the droplets during ESI. Compounds with a higher surface activity are expected to accumulate at the droplet/air interface resulting in better ionization efficiency.[11] Since surface activity is defined as a property that influences the surface tension of a liquid, the latter may be good indicator for this parameter, although log *P* is sometimes also used to deduce surface activity.[31] The effect of a compound on the surface tension of a liquid can be measured with the sessile or pendant drop method. This method, which determines the droplet contact angle on a flat surface between 40° and 90°, is limited to droplets of more than 20 μL.[32] Unfortunately this method failed to register any influence of the labeled amino acids on surface tension at concentrations up to 50 μM. We thus resorted to calculations as indicated under the Experimental Section. There is a weak negative correlation between the ESI response and the calculated surface tension for individual amino acids and their derivatives (see Supporting Information Figure S-8, top panel; surface tension values for the PEG-labeled amino acids are not included in this figure, as they were unavailable from the Chemspider database) with considerable scatter. It may also be that effects of the pure compounds on surface tension may be of less relevance due to their low concentrations in solution.

A striking feature of our data is that the smallest amino acid G has the highest and the biggest amino acid W the lowest gain in response upon derivatization. We therefore plotted the calculated molecular volume against the measured ESI response and found a strong correlation (Figure 4). When specified per

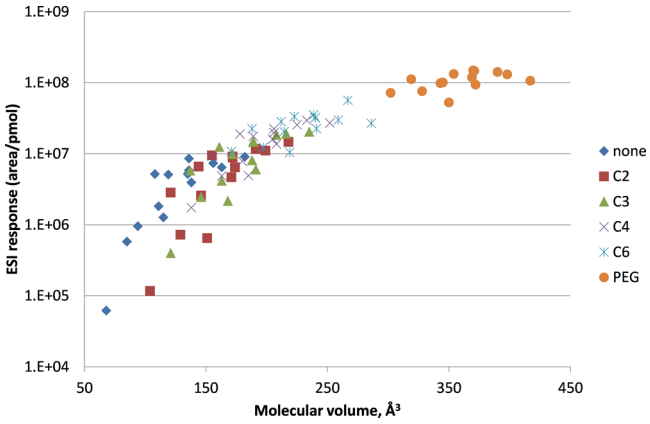


Figure 4. Relation between the ESI response and the calculated molecular volume specified for each acyl chain length and PEG moiety.

amino acid, correlations are even stronger (Figure S-9 and Table S-2, Supporting Information). G has a low ESI response due to its small size and high polarity and, hence, distributes homogeneously in the ESI droplets. With the introduction of acyl groups responses increase according to their length, and hence, their hydrophobicity as the derivatized amino acid enriches at the charged droplet surface, acquiring charge more easily.

When plotted against pK_a, the non-, acyl-, and PEG-labeled amino acids form three separate clusters (Figure S-10, lower panel, Supporting Information). A correlation between pK_a and ESI response, often reported in the literature,[3,31] was observed for the acyl-labeled amino acids notwithstanding the small pK_a window of one unit (Figure S-10 and Table S-2 Supporting Information). This indicates a relation between the length of the acyl chain the pK_a and the ESI response, especially for individual amino acids. However, since pK_a, hydrophobicity, surface tension, and molecular volume are all related to each other, we conclude that molecular volume is most strongly correlated with the ESI response and that pK_a plays only a minor role. A correlation between pK_a value and the ESI response was not observed for PEG-labeled amino acids and only marginally for the underivatized amino acids.

DOI: 10.1021/acs.analchem.7b01899
Anal. Chem. 2017, 89, 9159−9166

Since pH can affect the charge state, it is of interest to follow it during the electrospray process as it may relate to the ionization efficiency. Zhou et al. used laser-induced fluorescence spectroscopy with and without buffering around pH 7 and concluded that the pH in ESI droplets decreases by 1 unit at the most during evaporation.[33] From an analytical and stochastic modeling point of view a pH reduction of several units (from pH 4 to pH 1) was predicted in the presence of a buffering macromolecule like a protein, as deduced from its maximum charge state.[34] Since all compounds in this study are low molecular weight compounds, a pH reduction of maximally one unit may be expected during the ESI process. Solutions containing 0.1% formic acid (pK_a = 3.75) have a pH of about 2.7, which is in the pK_a range of the carboxylic acid group of most underivatized amino acids and about 1 unit below the pK_a of their derivatives. This means that we are essentially dealing with neutral molecules (<10% deprotonation of the carboxylic acid group), in the case of derivatized amino acids, at the acidic pH of the initial ESI droplets, that need to be charged by protonation/adduct formation. The nonderivatized amino acids, however, are already protonated at their amino group under these low-pH conditions, although the low responses demonstrate that this does not lead to more efficient ESI, indicating that pH is not a critical parameter in this context.

Since ESI transfers ions to the gas phase, a parameter like proton affinity, which is related to gas-phase basicity, may give a better relation to ionization efficiency and response. Proton affinities are difficult to measure and have only been determined for a limited number of compounds, among which amino acids but not for their derivatives. Although some models to predict proton affinity from molecular structure have been published,[35] they are to our knowledge not publicly available and/or still under development. While the reported proton affinity difference for nonderivatized G and W (887 vs 949 kJ/mol according to the NIST Chemistry WebBook[36]) is in line with their ESI response difference, this data is too sparse to draw conclusions. Investigations of proton affinities of amino acid derivatives are needed to link them to the measured ESI response.

QSPR Results. It would be of great interest to be able to predict how structural properties of molecules can be translated into molecular descriptors that can then be used to model the ESI response differences. This would allow us to "predict" how a certain type of derivatization strategy would affect the ESI response, taking some of the trial and error out of this approach

The measured electrospray responses of all compounds and their derivatives were used to find the best compound-related descriptors for building a QSPR model to predict the ESI response from molecular structures. To correct for day-to-day variations the responses were normalized to the response of butanoylated W measured in the same series, making the model transferable to other instrumental settings. Most of the compounds used to validate the model were not related to amino acids but contained an amine group that was derivatized using the same protocol as used for the amino acids. When ESI responses were clearly affected by limited solubility of the reaction products (e.g., for some of the hexanoylated derivatives showing reduced responses compared to butanoylated derivatives), results were rejected.

After rejection of all constant descriptors generated by the DRAGON application, the remaining descriptors (1552 from 3224) were subjected to cross-correlation analysis and grouped if their R^2 > 0.99. Only one descriptor from every group was used for further model calculations, which reduced the total number of descriptors further from 1552 to 788. Such clustering of descriptors prevents singular matrix formation during calculation. As preliminary research showed that direct prediction of the ESI response is unreliable, $\log_{10}$ values were used instead. Thereafter, all possible log ESI response dependencies with two descriptors were generated using the fuzzy forward searching algorithm and all combinations with the addition of a third descriptor were checked from the top 1000 pairs (see Supporting Information p S-11 for more details). The three-descriptor equations showed no significant advantage over the two-descriptor model (R^2 = 0.937 vs R^2 = 0.911) having similar uncertainties, although the isolated descriptors did not belong to a cluster. In this way QSPR modeling finally resulted in two non-cross-correlated descriptors, BIC0 and SPAN, to predict the ESI response. BIC0 is the bonding information content index proposed by Basak et al.[37] and generally describes the diversity of atomic composition and structural groups. The SPAN geometrical index is a simple size descriptor. It is the radius of a sphere centered in the molecule center of a mass enclosing the entire molecule.[38]

The data set used for modeling contained 84 compounds in the training set and 43 compounds in the validation set (see Table S-1 in Supporting Information). A second data set, excluding the PEGylated substances, was used to evaluate whether the somewhat different ESI response of PEGylated compounds affected the model. However, the same descriptors were isolated with a slightly worse correlation. This model was not used further. Despite the good correlation coefficient for the training set (R^2 > 0.9), leading to function 1, prediction of the ESI response across the entire validation set was worse (R^2 = 0.7).

$$\mathrm{Log(ESI)} = 0.075(\pm 0.013)\mathrm{SPAN} - 9.92(\pm 0.58)\mathrm{BIC0} + 2.43(\pm 0.24) \quad (1)$$

While there is generally a good correlation between predicted and measured ESI response for the acylated derivatives as well as for the PEGylated compounds, correlation is rather poor for halogen and nitro group-containing compounds (Figure 5).

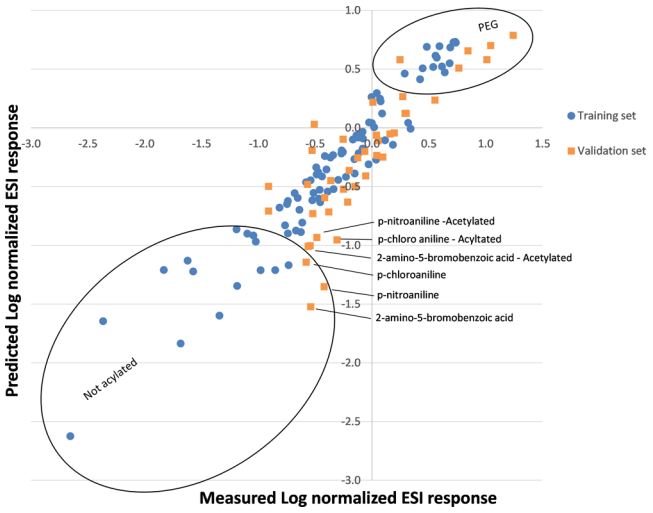


Figure 5. Comparison of predicted ESI responses for molecules from the training and the validation set to the measured responses using the isolated QSPR equation. Outliers corresponding to halogen or nitro group-containing compounds are indicated as well as the non- and PEG-labeled compound clusters.

DOI: 10.1021/acs.analchem.7b01899
Anal. Chem. 2017, 89, 9159−9166

The model tends to underestimate the ESI response for these molecules. Molecules with halogen atoms or nitro groups were not part of the training set of compounds, indicating that modeling the ESI response is restricted to structurally related compounds. Notably the superior ESI response of PEGylated compounds is correctly modeled.

CONCLUSIONS

In a systematic study of amino acids and their acyl and PEG derivatives, we found correlations between the ESI response and physicochemical compound parameters. Most notable was the correlation between the calculated molecular volume and the ESI response. Weaker, albeit still clear, correlations were observed with hydrophobicity (log P), dissociation constant (pK_a), and surface activity (negative correlation), especially when PEG labels were not taken into account. Notwithstanding the fact that a nonderivatized amino acid like G is already positively charged in solution, it does not give a very strong ESI response due to its high polarity caused by the absence of a side chain. Introduction of an acyl group increases the ESI response which correlates with chain length and, hence, hydrophobicity as well as RPLC retention time and surface activity. It is of interest to note that the PEG derivatives, which gave the strongest ESI response enhancement, did not fit the correlation plots of the acyl derivatives of increasing chain length except with respect to the molecular volume. We therefore consider that the calculated molecular volume is the best overall predictor of ESI response. This may be due to the fact that voluminous (charged) molecules occupy more space on the limited charged surface of the ESI droplet and consequently evaporate more easily following the ion evaporation model. Theoretically, they will also reach a solvent-free stage earlier following the charge residual model. Molecular volume is related to other physicochemical properties like molecular weight, molecular surface area, surface tension, boiling point, and solubility as well as to critical constants like critical pressure and critical temperature, which all may play a role, especially at the extreme conditions during Coulombic explosions of droplets.[39] It is interesting to note that ESI responses after PEG labeling are in the same order of magnitude for all amino acids as well as for the 41 compounds in the validation compound set. Considering the high solubility and favorable chromatographic properties (slight increase in hydrophobicity) of PEG derivatives, our work shows that labeling compounds with PEG is a viable and possibly generic strategy to enhance the ESI response of small molecules. More experimental work is needed to gain a better understanding of the physicochemical mechanisms behind this effect, but it is clearly of practical value. In situ studies on droplets during the ESI process are particularly promising to shed more light on this complex process that is of such an importance to modern analytical chemistry.[8]

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.7b01899.

MS parameter optima, FIA sodium adduct percentages, chromatographic/FIA response ratios, TIC chromatograms, ESI response plots against acetonitrile percentage, log P, retention time, surface tension, pK_a, and molecular volume specified per amino acid and/or label, QSPR modeling substances overview, correlation coefficients, QSPR modeling procedure and schema (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: r.p.h.bischoff@rug.nl. Phone: +31 0 503633338. Fax: 031 0 503637582.

ORCID

Rainer Bischoff: 0000-0001-9849-0121

Present Address

§S.O.: Hexal AG, Industriestrasse 25, 83607 Holzkirchen, Germany.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Wilm, M. S.; Mann, M. *Int. J. Mass Spectrom. Ion Processes* **1994**, *136*, 167−180.
(2) Wilm, M. *Mol. Cell. Proteomics* **2011**, *10*, M111.009407.
(3) Oss, M.; Kruve, A.; Herodes, K.; Leito, I. *Anal. Chem.* **2010**, *82*, 2865−2872.
(4) Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B. *J. Chem. Phys.* **1968**, *49*, 2240−2249.
(5) Iribarne, J. V.; Thomson, B. A. *J. Chem. Phys.* **1976**, *64*, 2287−2294.
(6) Enke, C. G. *Anal. Chem.* **1997**, *69*, 4885−4893.
(7) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. *Anal. Chem.* **2013**, *85*, 2−9.
(8) Duft, D.; Achtzehn, T.; Muller, R.; Huber, B. A.; Leisner, T. *Nature* **2003**, *421*, 128−128.
(9) Kebarle, P.; Verkerk, U. H. *Mass Spectrom. Rev.* **2009**, *28*, 898−917.
(10) Banerjee, S.; Mazumdar, S. *Int. J. Anal. Chem.* **2012**, *2012*, 282574.
(11) Cech, N. B.; Enke, C. G. *Mass Spectrom. Rev.* **2001**, *20*, 362−387.
(12) Zook, D. R.; Bruins, A. P. *Int. J. Mass Spectrom. Ion Processes* **1997**, *162*, 129−147.
(13) Van Eeckhaut, A.; Lanckmans, K.; Sarre, S.; Smolders, I.; Michotte, Y. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2009**, *877*, 2198−2207.
(14) Mortier, K. A.; Zhang, G.; Van Peteghem, C. H.; Lambert, W. E. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 585−592.
(15) Siegel, D.; Permentier, H.; Bischoff, R. *J. Chromatogr. A* **2013**, *1294*, 87−97.
(16) Bruins, A. P. In *Electrospray Ionization Mass Spectrometry Fundamentals, Instrumentation & Applications*; Cole, R. B., Ed.; Wiley: New York, 1997; pp 107−136.
(17) Leitner, A.; Emmert, J.; Boerner, K.; Lindner, W. *Chromatographia* **2007**, *65*, 649−653.
(18) Hahne, H.; Pachl, F.; Ruprecht, B.; Maier, S. K.; Klaeger, S.; Helm, D.; Medard, G.; Wilm, M.; Lemeer, S.; Kuster, B. *Nat. Methods* **2013**, *10*, 989−991.
(19) Yu, P.; Hahne, H.; Wilhelm, M.; Kuster, B. *Anal. Bioanal. Chem.* **2017**, *409*, 1049−1057.
(20) Kulevich, S. E.; Frey, B. L.; Kreitinger, G.; Smith, L. M. *Anal. Chem.* **2010**, *82*, 10135−10142.
(21) Abello, N.; Geurink, P. P.; Toorn, M. v. d.; Oosterhout, A. J. M. v.; Lugtenburg, J.; Marel, G. A. v. d.; Kerstjens, H. A. M.; Postma, D. S.; Overkleeft, H. S.; Bischoff, R. *Anal. Chem.* **2008**, *80*, 9171−9180.
(22) Rebane, R.; Rodima, T.; Kütt, A.; Herodes, K. *J. Chromatogr. A* **2015**, *1390*, 62−70.
(23) Null, A. P.; Nepomuceno, A. I.; Muddiman, D. C. *Anal. Chem.* **2003**, *75*, 1331−1339.
(24) Cech, N. B.; Enke, C. G. *Anal. Chem.* **2000**, *72*, 2717−2723.

DOI: 10.1021/acs.analchem.7b01899
Anal. Chem. 2017, 89, 9159−9166

(25) Golubović, J.; Birkemeyer, C.; Protić, A.; Otašević, B.; Zečević, M. *J. Chromatogr. A* **2016**, *1438*, 123−132.
(26) Chalcraft, K. R.; Lee, R.; Mills, C.; Britz-McKibbin, P. *Anal. Chem.* **2009**, *81*, 2506−2515.
(27) Pence, H. E.; Williams, A. *J. Chem. Educ.* **2010**, *87*, 1123−1124.
(28) Trinquier, G.; Sanejouand, Y. H. *Protein Eng., Des. Sel.* **1998**, *11*, 153−169.
(29) Cech, N. B.; Krone, J. R.; Enke, C. G. *Anal. Chem.* **2001**, *73*, 208−213.
(30) Poole, S. K.; Poole, C. F. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2003**, *797*, 3−19.
(31) Ehrmann, B. M.; Henriksen, T.; Cech, N. B. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 719−728.
(32) Noordmans, J.; Busscher, H. J. *Colloids Surf.* **1991**, *58*, 239−249.
(33) Zhou, S.; Prebyl, B. S.; Cook, K. D. *Anal. Chem.* **2002**, *74*, 4885−4888.
(34) Malevanets, A.; Consta, S. *J. Chem. Phys.* **2013**, *138*, 184312.
(35) Pedraza-Gonzalez, L.; Romero, J.; Ali-Torres, J.; Reyes, A. *Phys. Chem. Chem. Phys.* **2016**, *18*, 27185−27189.
(36) Hunter, E. P.; Lias, S. G. In *NIST Chemistry WebBook, NIST Standard Reference Database*; Linstrom, P. J., Mallard, W. G., Eds.; National Institute of Standards and Technology: Gaithersburg, MD, 2016.
(37) Basak, S. C.; Harriss, D. K.; Magnuson, V. R. *J. Pharm. Sci.* **1984**, *73*, 429−437.
(38) Volkenstein, M. V. *Configurational Statistics of Polymeric Chains*; Interscience: New York, 1963; Chapter 4.
(39) Mebane, R. C.; Schanley, S. A.; Rybolt, T. R.; Bruce, C. D. *J. Chem. Educ.* **1999**, *76*, 688.

DOI: 10.1021/acs.analchem.7b01899
Anal. Chem. 2017, 89, 9159−9166