Negative Electrospray Ionization via Deprotonation: Predicting the Ionization Efficiency

ARTICLE *in* ANALYTICAL CHEMISTRY · APRIL 2014

Impact Factor: 5.64 · DOI: 10.1021/ac404066v · Source: PubMed

4 AUTHORS, INCLUDING:

Anneli Kruve
University of Tartu
22 PUBLICATIONS **406** CITATIONS
SEE PROFILE

Jaanus Liigand
University of Tartu
6 PUBLICATIONS **28** CITATIONS
SEE PROFILE

analytical chemistry

Article

pubs.acs.org/ac

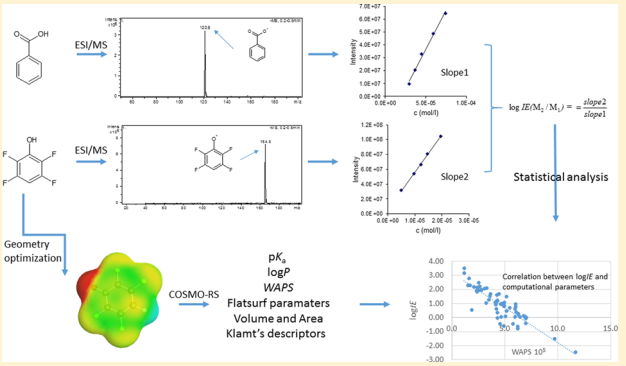
Negative Electrospray Ionization via Deprotonation: Predicting the Ionization Efficiency

Anneli Kruve,* Karl Kaupmees, Jaanus Liigand, and Ivo Leito

University of Tartu, Institute of Chemistry, Ravila 14a, Tartu 50411, Estonia

Supporting Information

ABSTRACT: Electrospray ionization (ESI) in the negative ion mode has received less attention in fundamental studies than the positive ion electrospray ionization. In this paper, we study the efficiency of negative ion formation in the ESI source via deprotonation of substituted phenols and benzoic acids and explore correlations of the obtained ionization efficiency values (log*IE*) with different molecular properties. It is observed that stronger acids (i.e., fully deprotonated in the droplets) yielding anions with highly delocalized charge [quantified by the weighted average positive sigma (*WAPS*) parameter rooted in the COSMO theory] have higher ionization efficiency and give higher signals in the negative-ion ESI/MS. A linear model was obtained, which equally well describes the log*IE* of both phenols and benzoic acids (R^2 = 0.83, S = 0.40 log units) and contains only an ionization degree in solution (α) and *WAPS* as molecular parameters. Both parameters can easily be calculated with the COSMO-RS method. The model was successfully validated using a test set of acids belonging neither to phenols nor to benzoic acids, thereby demonstrating its broad applicability and the universality of the above-described relationships between IE and molecular properties.



ESI/MS is among the most widely used analytical techniques and enables a wide range of compounds to be ionized and analyzed. The mechanism of electrospray ionization is still not fully known, in spite of its wide usage. Although both positive and negative ionization is possible in ESI, positive ionization is significantly more often used and more extensively studied. In ESI source ion formation from small molecules is described via the ion evaporation model[1] and in the case of large analytes (e.g., biomolecules), the charge residue model[2] is used. It has been shown that in the case of the large molecules, the residual charge of the molecule is controlled by the ion evaporation mechanism.[3] For singly and multiply charged cluster ions, Gamero-Castaño et al.[4] observed that both ion evaporation and charge residue mechanism act together, each doing part of the ionization.

In the negative ionization mode, ions are mostly generated via deprotonation, adduct formation with anions,[5] or via simultaneous deprotonation and adduct formation with cations (e.g., ions $[2M - 2H + Na]^-$).[6]

It is known that in both positive and negative ionization mode different compounds have different ionization efficiency.[7] This means that a different number of gas-phase ions are generated from the same number of analyte molecules/ions in solution. The variations in ionization efficiency have been studied for both positive[7] and negative[8] ionization.

Henriksen et al.[9] have observed higher ESI negative response for compounds with higher log*P* values, although the relationship was fuzzy. Huffman et al. [8] have studied the negative ionization for different compound classes (carboxylic acids, phenols, steroids, etc.) in different solvents (methanol, acetone, acetonitrile, and water). The highest ionization efficiencies were observed in methanol, resulting in both higher sensitivity and lower LoD, in both flow injection and chromatographic experiments. No clear dependency of the ESI response on compound class was observed, but the response depended on the molecular structure (e.g., on the length of the alkyl chain of a carboxylic acid).

For positive ionization (via protonation), it has been observed that basicity of the analyte (i.e., conjugate acid pK_a) and the size of the molecule (molecular volume) can be used to predict the response in the ESI source.[7] Similarly, Nguyen et al.[10] have observed positive correlation between ESI response and adjusted mass of the ion [adjusted mass = $n(H)/n(C) \times$ molecular mass]. On the other hand, Chalcraft et al.[11] showed that the ESI signal can be predicted using 4 molecular parameters: molecular volume, octanol–water distribution coefficient, absolute mobility, and effective charge. Ehramm et al.[12] observed that the parameter that most effectively predicted the response in ESI/MS was the analyte's pK_b, a measure of its basicity in solution.

Theoretical studies based on the molecular dynamics simulations by Konermann et al.[13] have revealed that for a molecule to be ejected from the ESI droplet as an ion, it is

Received: December 16, 2013
Accepted: April 14, 2014

   dx.doi.org/10.1021/ac404066v | *Anal. Chem.* XXXX, XXX, XXX–XXX

necessary to move to the surface of the droplet. From the droplets surface, the ionized analyte is ejected with a few solvent molecules and the analyte is fully desolvated later.

It has been shown that negative ionization is less influenced by the ionization suppression caused by the compounds coeluting with the analyte,[14,15] therefore making the negative ionization mode attractive for quantitative analyses.

Attempts have been made to describe the activation energy of ion evaporation (or also called solvation energy of the ion) based on the charge and curvature of the droplet[16] as well as dielectric constant and surface tension.[17] Interestingly, it was shown by Hogan et al.[3] that charge carriers influence the ΔG of ion evaporation. More precisely, for more compact ions, the ΔG was found to be higher, therefore delaying the ion evaporation. [3]

In few studies, trends between molecular structure and ionization efficiency have been revealed.[11,10] However, none of them attempted to establish a quantitative model for predicting the negative ions MS response from molecular properties.

Hereby, we report an ionization efficiency study of 29 phenols, 18 benzoic acids, and 3 bifunctional compounds (salicylic acids) of diverse substitution patterns via deprotonation in negative ESI−MS. The measured ionization efficiency values are correlated with the molecular properties and a model for predicting ionization efficiency from molecular parameters is obtained.

EXPERIMENTAL SECTION

Chemicals. As analytes in training set derivatives of benzoic acid and phenol were used. The structures with origin for both training set as well as validation set are presented in Table S1 of the Supporting Information. Purity of all substances was confirmed by mass spectra. The solvent was prepared from ultrapure water (18.2 MΩ·cm, TOC 1−2 ppb, prepared by a Millipore Milli-Q Advantage A10 water-purification system), acetonitrile (MeCN) (HPLC grade, J. T. Baker, Deventer, Netherlands), and concentrated aqueous ammonia (25% solution, Lachner, Czech Republic).

The aqueous phase pH was measured with a pH meter (Evikon pH Meter E6115), using a glass electrode (Evikon pH631).

MS Studies of Ionization Efficiency. The MS responses were recorded in a flow injection mode. The concentration of the analyte stock solution varied between 110 and 1.2 μM, depending on the analyte properties. Six dilutions of the analyte stock solutions were made (1, 1.25, 1.67, 2, 2.5, and 5 fold) with the solvent (0.1% aqueous ammonia solution and acetonitrile in 20:80 volume ratio) by the autosampler and injected to MS. The injection volume was 10 μL, and the eluent flow rate was 0.2 mL/min. Analyte concentrations in the injected solutions ranged from 2.2×10^{-5} to 2.4×10^{-7} M. In order to estimate the instrument stability during measurements, the reference compound, benzoic acid, was analyzed in the beginning and in the end of each sequence.

The measurements were carried out using an Agilent XCT ion trap mass spectrometer. The MS and ESI parameters were optimized only by setting the Target Mass parameter (see below). Factory defaults of the remaining parameters were used: nebulizer gas pressure 15 psi, drying gas flow rate 7 L/min, drying gas temperature 300 °C. The capillary voltage between the MS and nebulizer was −3500 V. All remaining ion transport parameters were determined by the Target Mass parameter, set by the operator. The ion trap parameters were as follows: Smart Target (parameter characterizing the number of ions accumulated in the trap) was 100000 and the maximum accumulation time 300 ms. Each spectrum was scanned from m/z 30 to 1000.

Measurements were carried out using different target mass (TM) values: 150, 200, 250, 300, and so forth. For each analyte, the TM closest to the m/z of its anion was used. The validation of this approach is shown in Figure S-1 of the Supporting Information.

All compounds except two yielded only the $[M - H]^-$ ions in the mass spectra. The remaining two compounds yielded fragment ions: 2,3,4,5,6-pentafluorobenzoic acid and picric acid. For the 2,3,4,5,6-pentafluororbenzoic acid, the fragment $[M - H - CO_2]^-$ was observed and the anion $[M - H]^-$ was not observed. For picric acid, only the ion corresponding to $[M - NO_2]^-$ was observed and the anion $[M - H]^-$ was not observed. Therefore, picric acid could not be used in the study of ionization efficiency via deprotonation. 2,3,4,5,6-Pentafluorobenzoic acid, however, was used on the assumption that as the first step, the 2,3,4,5,6-pentafluorobenzoate anion is formed and ejected from the droplet and the decarboxylation process occurs only thereafter, in the gas phase. Cluster ions were not observed for any of the compounds.

We term the negative ionization efficiency via deprotonation as IE (or logIE on the logarithmic scale), similar to the ionization efficiency in the positive ESI mode.[18] As it is complicated to measure absolute ionization efficiency,[18] we focus on measuring the relative ionization efficiency (IE) of a compound M_1 relative to benzoic acid (M_2), according to the following equation:

$$\mathrm{RIE}(M_1/M_2) = \frac{R([M_1 - H]^-)}{R([C_6H_5COO]^-)} = \frac{\mathrm{slope}([M_1 - H]^-)}{\mathrm{slope}([C_6H_5COO]^-)} \quad (1)$$

where the slope of the analyte signal versus concentration is estimated via linear regression in the linear range of the signal-concentration plot. In order to make the data easier to present and analyze, the logarithmic scale (logIE) was used. In order to find absolute logIE values, the scale was anchored to the logIE of benzoic acid, taken arbitrarily as 0.

In this work, every compound was measured individually (alone in solution), not in pairs (two compounds simultaneously in solution) as in earlier works.[18,7] In order to minimize the influence of possible differences in experimental conditions when measuring M_1 and benzoic acid, two steps were taken: (1) each acid was measured on at least 3 different runs (on 3 different days) and the results were averaged and (2) benzoic acid was measured in the beginning and end of each run on each day.

For validating this measurement method, the results of the absolute and relative measurements (as described in ref 18) were compared in the case of four compounds−benzoic acid, 3,5-dinitrobenzoic acid, 4-*tert*-butylbenzoic acid, and pentakis-(trifluoromethyl)phenol. The differences in ionization efficiency values were not statistically significant, according to the t-test. It was concluded that ionization efficiency values are independent of the measurement method (if sufficiently diluted solutions are used).

COSMO-RS/Turbomole Computations. COSMO-RS method[19] was used for calculating the aqueous pK_a and log$P$ (solvent/vacuum) values as well as for generating sigma profiles

dx.doi.org/10.1021/ac404066v | Anal. Chem. XXXX, XXX, XXX−XXX

needed for calculating the *WAPS* parameter.[20] The ionization degrees α of the acids were calculated from the pK_a values calculated with COSMOtherm and the directly measured pH of the aqueous phase.

COSMO-RS was chosen because it enables calculations in solvent mixtures[19] and is able to handle preferential solvation of solutes by one of the solvent mixture components.[20]

As the first step, full geometry optimization and energy calculation at the DFT BP TZVP level with the RI approximation and applying the COSMO continuum solvation model was carried out for all species and solvent molecules using Turbomole, version 6.3.[21] Different starting geometries were used, based on common chemical knowledge of the species involved. For most species, several conformers corresponding to different local energy minima were found. These were all taken into account by statistical weighting inherent in the COSMO-RS procedure (see below). The default convergence criteria of Turbomole were used (wave function convergence: max difference 10^{-6} Hartree; geometry convergence: max gradient $|dE/dxyz|$ 10^{-3} Hartree/Bohr). This first computation step yields for every conformer the following data: (1) the geometry of the conformer, (2) detailed data on the shape of the molecular cavity, (3) the polarization charge densities mapped onto the cavity surface, (4) the total electronic energy of the species submerged into a virtual conductor ($\varepsilon = \infty$), and (5) molecular surface area and volume.

As the second step, the COSMO-RS calculation was carried out on all species using the above listed data as input data with the COSMOtherm, version C3.0, release 13.01.[22] COSMO-RS calculation takes into account the interactions between the species and the solvent/medium molecules, as well as between the solvent molecules themselves (the implicit solvation model is used). The solvent composition is input as it was in the experiment (i.e., including the concentration of water in the solvent). Zero concentrations of the studied species are used. This way the interactions between the species and the solvent components are taken into account but not the interactions between the species themselves. This situation corresponds well to the reality of very low concentrations used in the experiments and is the common practice of carrying out such calculations.[23] Both van der Waals interactions (electrostatic interactions, such as dipole−dipole, ion−dipole, etc. forces, as well as dispersion forces) and hydrogen bonds (in the implicit way) are taken into account. The interactions are quantified via statistical counting and averaging of energies of pairwise interactions of molecular surface segments using the polarization charge density maps of the species created in the first step, taking into account the concentrations of the respective species in the solution.[19] Terms accounting for vibrational contributions to the G_{tot} are also added in this step. This is done implicitly, as these are represented through the experimental data used for parametrization of the method. The energetics of these interactions are calculated at 298 K, using a statistical thermodynamics procedure whereby also the conformers of all the interacting species are taken into account and statistically weighted based on their relative stabilities.[19] This way, the entropy effect of the same species present in multiple conformers is also accounted for. As a result, a G_{tot} value is found for every species.

WAPS is a parameter for quantitatively evaluating the charge delocalization in anions and is calculated as follows:[20]

$$WAPS = \frac{\int_{\sigma=0}^{\infty} \sigma p(\sigma)\, d\sigma}{A \int_{\sigma=0}^{\infty} p(\sigma)\, d\sigma} \quad (2)$$

Where σ is the polarization charge density on ion surface, $p(\sigma)$ is the probability function of σ and A is the surface area of the anion. The smaller the *WAPS* value, the more delocalized the charge in the anion. It has been proposed that *WAPS* values above 4.5 indicate anions with localized charge.[20]

The Flatsurf module in COSMOtherm is used to calculate the orientation and energy of molecules or ions at the surface of two solvents or (as in our case) a solvent and the gas phase (vacuum). Vacuum represents the gas phase in COSMOtherm as its properties are insignificantly different from the air (similarly negligible interactions between molecules). Using the same approach as described above for the general COSMOtherm calculations, the gain in energy is calculated when moving the molecule or ion from solvent to the surface so that certain segments of molecular surface are embedded in the liquid phase, whereas the rest is in the gas phase. All relevant positions and orientations are sampled in order to find the energy minimum at the flat surface. Flatness of the surface is, in this case, an approximation because during ESI ionization, the surface of the droplets is not flat. As a result, the total energy gain when moving from solvent to the surface is calculated taking into account all conformers and their possible orientations on the surface. Also, the most surface active conformer is found for which the same energy gain can be calculated as well as the surface area of molecular segments in both phases.

The COSMO-RS theory implies that the solvent space is 5-dimensional and that any log-partition coefficient is a linear combination of 5 σ-moments:[24] polarity/polarizability, hydrogen bond acidity and basicity, and the surface area (CSA) descriptor. These were all calculated using COSMOtherm. If ionization efficiency were just a partition coefficient then these descriptors could be used to construct a quantitative structure−property relationship (QSPR).The molecular properties calculated using COSMOtherm are presented in Table S-2 of the Supporting Information.

Statistical Model Evaluation. Multilinear regression analysis was used to obtain the model for describing the relationship of log*IE* with the other parameters. For building, the model training set, containing phenols, benzoic acids and salicylic acids, was used. In order to obtain a model easily usable by other laboratories, the input parameters as well as the log*IE* values were not normalized or scaled. All molecular parameters that proved to be statistically insignificant in the multilinear regression were left out from the model. In order to prevent overfitting, every time when a parameter, although with statistically significant slope, was added to the model, it was also compared to the less complex model, according to the remaining unexplained variance. All parameter additions that failed to yield statistically significant reduction of unexplained variance of log*IE* were rejected.

The model was validated with the validation set of acids, for which the log*IE* values were measured after the model with training set (containing phenols and benzoic acids) was built. The validation set contained compounds from different compound classes: sulfonic acids, sulphonamides, a sulphonimide, a perfluorinated alcohol, nonaromatic carboxylic acids, and carboxylic acid imide derivatives. The model's lack of fit

dx.doi.org/10.1021/ac404066v | *Anal. Chem.* XXXX, XXX, XXX−XXX

Table 1. Measured ($\log IE_m$) and Predicted ($\log IE_p$) Ionization Efficiency Values for Training and Validation Set Compounds, Ionization Degree in Solution (α), Calculated pK_a Values in Water, *WAPS*, Number of Halogen Atoms *n*(Hal), and Molecular Volume (MV)

molecule	group	$\log IE_m$	$\log IE_p$	α	pK_a (H_2O)	$WAPS \cdot 10^5$	*n*(Hal)	MV (Å^3)
4-aminobenzoic acid	benzoic acid	−0.06	−0.28	1.00	5.72	7.0	0	159.3
benzoic acid	benzoic acid	0	−0.33	1.00	4.63	7.1	0	144.4
4-isopropylbenzoic acid	benzoic acid	0.05	0.89	1.00	4.90	4.7	0	209.2
2,4-dinitrobensoic acid	benzoic acid	0.07	0.68	1.00	1.82	5.1	0	206.7
3-aminobenzoic acid	benzoic acid	0.09	−0.23	1.00	5.11	6.9	0	159.3
4-hydroxybenzoic acid	benzoic acid	0.13	−0.18	1.00	5.33	6.8	0	154.8
3-chlorobenzoic acid	benzoic acid	0.48	0.63	1.00	4.42	5.2	1	167.3
pentafluorobenzoic acid	benzoic acid	0.61	1.14	1.00	2.50	4.2	5	187
3-(dimethylamino)benzoic acid	benzoic acid	0.71	0.48	1.00	5.19	5.5	0	200.7
3,5-dinitrobenzoic acid	benzoic acid	0.79	0.79	1.00	2.94	4.9	0	205.2
3-(trifluoromethyl)benzoic acid	benzoic acid	1.16	1.14	1.00	4.28	4.2	3	187.3
4-*tert*-butylbenzoic acid	benzoic acid	1.2	1.09	1.00	4.91	4.3	0	229.8
3-[(trifluoromethyl)sulphanyl]benzoic acid	benzoic acid	1.28	1.49	1.00	4.36	3.5	3	216.8
3-[(trifluoromethyl)sulphonyl]benzoic acid	benzoic acid	1.69	1.44	1.00	3.77	3.6	3	240.1
3-[(heptafluoropropyl)sulphanyl]benzoic acid	benzoic acid	1.84	2.05	1.00	4.29	2.4	6	294.3
3-[tris(trifluoromethyl)methyl]benzoic acid	benzoic acid	2.14	2.05	1.00	4.29	2.4	9	291.6
3-[tris(trifluoromethyl)methylsulfonyl]-benzoic acid	benzoic acid	2.15	2.05	1.00	3.72	2.4	9	340.8
3-[(heptafluoropropyl)sulphonyl]benzoic acid	benzoic acid	2.2	1.95	1.00	3.72	2.6	6	313.2
2-bromophenol	phenol	−0.8	0.11	0.99	8.48	6.2	1	147.6
2-nitrophenol	phenol	−0.62	0.13	1.00	6.51	6.2	0	149.6
2-isopropylphenol	phenol	−0.61	−0.28	0.52	10.71	4.9	0	182.9
2-*tert*-butylphenol	phenol	−0.46	−0.42	0.37	10.98	4.5	0	204.1
2-phenylphenol	phenol	−0.01	0.62	0.79	10.17	4.3	0	211.2
4-chlorophenol	phenol	0.03	−0.06	0.96	9.34	6.4	1	141.4
3-chlorophenol	phenol	0.22	0.03	0.98	9.09	6.3	1	142
2-cyanophenol	phenol	0.28	0.08	1.00	6.73	6.3	0	145
4-phenylphenol	phenol	0.35	0.96	0.94	9.53	4.3	0	212.6
2,4,6-tritert-butylphenol	phenol	0.55	0.19	0.05	12.00	1.9	0	375.4
3-bromophenol	phenol	0.63	0.18	0.98	9.01	6.0	1	148.3
3-nitrophenol	phenol	0.78	0.23	1.00	7.83	6.0	0	150.8
2,4-dinitrophenol	phenol	0.84	0.79	1.00	3.44	4.9	0	181.6
pentafluorophenol	phenol	0.91	1.19	1.00	6.08	4.1	5	160.7
2,3,5,6-tetrafluorophenol	phenol	0.97	0.94	1.00	6.26	4.6	4	152.2
4-nitrophenol	phenol	0.97	0.23	1.00	5.29	6.0	0	149.9
2,5-dinitrophenol	phenol	0.98	0.89	1.00	4.71	4.7	0	180.5
4-phenylazophenol	phenol	1.02	1.65	1.00	7.41	3.2	0	238.1
2,6-dinitrophenol	phenol	1.06	0.74	1.00	3.12	5.0	0	180.9
2,4,6-trinitro-1,3-benzenediol	phenol	1.07	1.19	1.00	0.64	4.1	0	219.4
4-(4-nitrophenylaso)-1-napthol	phenol	1.77	2.20	1.00	4.61	2.1	0	323.8
heptafluornapht-2-ol	phenol	1.94	1.85	1.00	5.71	2.8	6	231.3
pentabromophenol	phenol	1.96	1.95	1.00	5.21	2.6	5	258.6
pentachlorophenol	phenol	1.97	1.85	1.00	5.24	2.8	5	227.7
4-(pentafluorophenyl)-2,3,5,6-tetrafluorophenol	phenol	2.17	2.20	1.00	5.41	2.1	9	286.7
pentakis(trifluoromethyl)phenol	phenol	2.3	2.51	1.00	1.32	1.5	15	330
2-hydroxy-1,3,5-tris(2,2,2-trifluoroethyl)-1,3,5-benzenetrisulfonic acid ester	phenol	3.14	2.66	1.00	−0.25	1.2	9	505.4
2-hydroxy-1,3,5-tris(2,2,3,3-tetrafluoro-propyl) 1,3,5-benzenetrisulfonic acid ester	phenol	3.49	2.66	1.00	0.51	1.2	12	563.3
salicylic acid	salicylic acid	0.39	0.18	1.00	3.33	6.1	0	151.6
3,5-dinitrosalicylic acid	salicylic acid	1.22	1.09	1.00	0.32	4.3	1	213.4
3,5-diiodosalicylic acid	salicylic acid	2.05	1.70	1.00	2.65	3.1	2	230.1
beta-alanine	validation set	−2.46	−2.98	0.87	9.91	11.7	0	108.1
4-aminobutanoic acid	validation set	−1.52	−2.05	0.84	10.03	9.7	0	131.8
N-hydroxy-phthalimide	validation set	−0.43	0.32	1.00	5.29	5.8	0	174.9
sorbic acid	validation set	−0.40	−0.30	1.00	5.35	7.0	0	147.5
phthalimide	validation set	0.05	0.22	1.00	7.68	6.0	0	162.8
3-nitrobenzenesulphonamide	validation set	1.50	0.63	1.00	7.58	5.2	0	203.9
$(CF_3)_3COH$	validation set	1.54	1.44	1.00	5.82	3.6	9	180.6

dx.doi.org/10.1021/ac404066v | *Anal. Chem.* XXXX, XXX, XXX−XXX

Table 1. continued

molecule	group	logIE_m	logIE_p	α	pK_a (H_2O)	$WAPS \cdot 10^5$	n(Hal)	MV (Å^3)
naphthalenesulfonic acid	validation set	1.60	1.10	1.00	−1.57	4.3	0	221.1
2,4-dinitro-benzenesulfonic acid	validation set	1.68	1.01	1.00	−3.25	4.4	0	230.3
tetradecanoicacid	validation set	1.91	2.08	1.00	5.00	2.3	0	335.7
2-anthraquinonesulfonic acid	validation set	2.10	1.59	1.00	−2.44	3.3	0	295.1
2,3,4,5,6-pentafluoro-*N*-[(2,3,4,5,6-pentafluorophenyl)sulphonyl]-benzenesulphonamide	validation set	2.29	2.38	1.00	−0.94	1.7	10	395.5
4-methyl-*N*-[(4-nitrophenyl)sulphonyl]-benzenesulphonamide	validation set	2.46	1.98	1.00	0.96	2.5	0	367.5
1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8,8-heptadecafluorooctane-1-sulfonic acid	validation set	2.77	2.44	1.00	−4.56	1.6	17	378.6

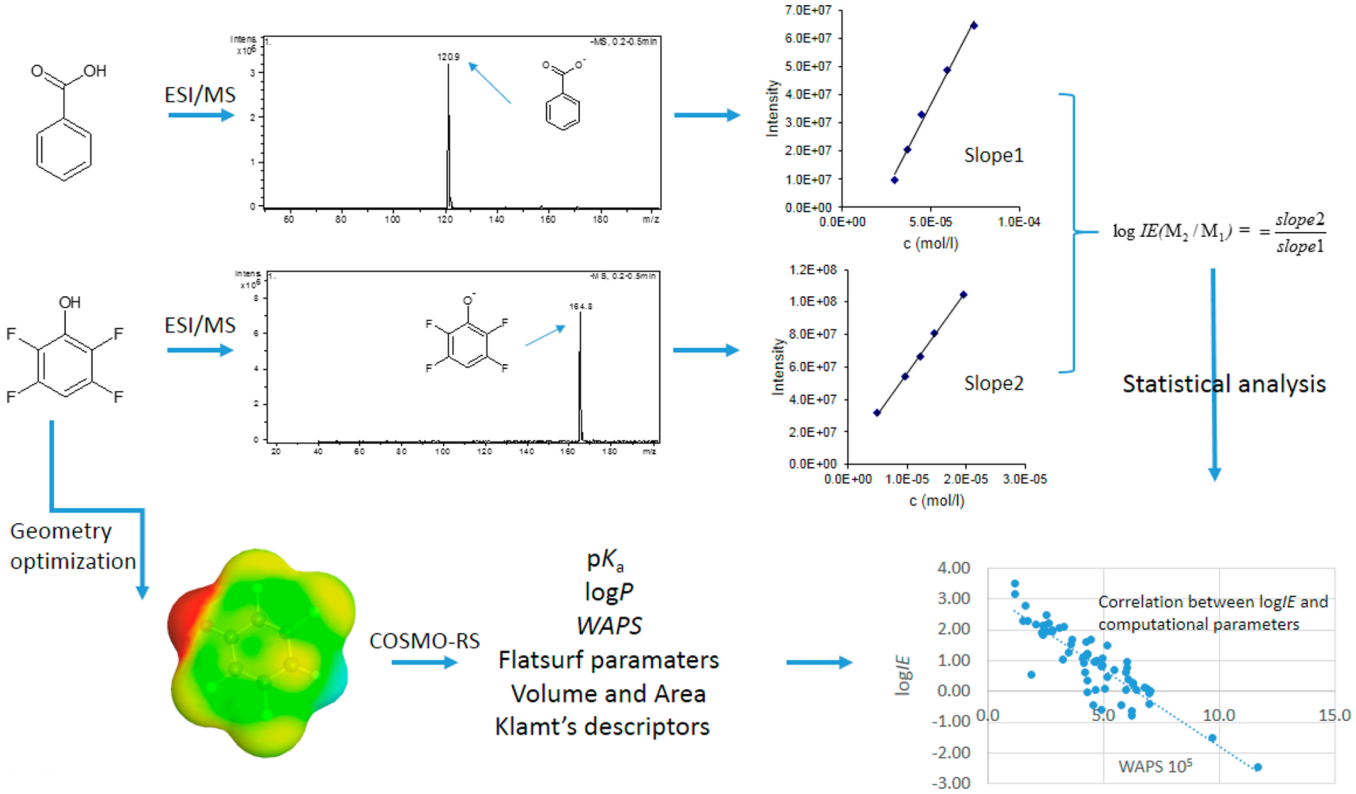


Figure 1. Schematic representation of the work.

RMS deviation (RMSD) for the validation set was evaluated accordingly:

$$\mathrm{RMSD} = \sqrt{\frac{\sum (\log IE_m - \log IE_p)^2}{n}} \quad (3)$$

where logIE_m is the measured ionization efficiency of a compound logIE_p is the predicted ionization efficiency of the compound, according to the obtained model (eq 6), and n is the number of compounds in the validation set.

All statistical tests were performed at the 95% significance level.

RESULTS AND DISCUSSION

Electrospray ionization efficiency of 49 substituted phenols and benzoic acids has been studied (ionization via deprotonation). The obtained logIE values (Table 1) relative to benzoic acid range from −0.8 up to 3.5, indicating that acidic compounds have ionization efficiencies ranging for more than 4 orders of magnitude. The day-to-day reproducibility standard deviation of the logIE values did not exceed 0.26 logIE units (for 2,5-dinitrophenol), and the pooled standard deviation over all assigned logIE values was 0.14 logIE units. The schematic representation of the work flow is given in Figure 1.

Correlation between Molecular Properties and IE for the Training Set. In order to rationalize the relationship between logIE values and molecular properties, correlations of several parameters (calculated with COSMO-RS) with logIE were studied. Namely pK_a values in water and MeCN as well as log$P_{solvent/vacuum}$ of the anions observed in mass spectra, Flatsurf properties estimated by COSMS-RS, molecular surface area (Å^2), molecular volume (Å^3), *WAPS*, and descriptors for log-partition coefficients.[24] Also experimentally available pK_a values in MeCN and gas-phase acidity (GA) values and the number of halogen atoms in the molecule were correlated with measured logIE values.

As a first step, single parameter correlations were studied. Strong negative correlations of logIE with both experimental pK_a (MeCN) and GA values were observed (R^2 = 0.61 in both cases). These two parameters were also strongly correlated to each other (R^2 = 0.90). At the same time, correlation with pK_a

dx.doi.org/10.1021/ac404066v | Anal. Chem. XXXX, XXX, XXX−XXX

(MeCN) calculated with COSMOS-RS was somewhat poorer (R^2 = 0.51).

It may be possible that the correlation accidentally happens to be better because of the limited number of compounds used for correlation with experimental pK_a values. It is necessary to point out that for the compounds for which the experimental pK_a values were available the correlation between log*IE* and calculated pK_a values was statistically insignificantly different from the one observed with experimental pK_a values, although it has been observed that COSMO-RS-predicted pK_a values are not very accurate [e.g., the root-mean-square error of pK_a values (in MeCN) predicted with COSMO-RS has been found to be 1.38 pK_a units].[25]

Correlation between ionization efficiency and computational pK_a values in water (R^2 = 0.35) was statistically similar to pK_a in MeCN.

The correlation between log*IE* and pK_a clearly shows that stronger acids (lower pK_a values) tend to have higher ionization efficiency. The R^2 of the correlation is significantly lower than 1, which implies that the strength of the acid is not the only parameter influencing ionization efficiency. The same has also been observed in the positive ESI.[7] This finding indicates that ESI ionization is more complex than a pure solution equilibrium.

It has been previously observed for the positive ions that ionization efficiency is dependent on the molecule size.[7] Therefore, correlations with molecular surface area and molecular volume of the anion were studied. These parameters showed a good correlation (R^2 = 0.58 and 0.59, respectively) with the measured ionization efficiency. The correlation with the logarithm of molecular surface area or molecular volume was also good (R^2 = 0.61 in both cases). Obviously, molecular surface area and molecular volume are strongly correlated to each other and, therefore, only molecular volume was used as an input parameter for the multilinear regression described below. Therefore, it can be concluded that in this data set bigger molecules tend to have higher ionization efficiency. Molecular volume can also be interpreted as a cavity term describing the energy released when the ion leaves the droplet and the solute cavity collapses.

On the other hand, correlation of log*IE* with hydrophobicity described by the log$P_{solvent/vacuum}$ was poor (R^2 = 0.24). The same has also been observed in positive ESI.[7] This may be caused by the following. First, COSMO-RS calculations can only be made for the neutral solvent, while droplets in ESI are negatively charged (in negative ESI). Therefore, strong ion−ion repulsion occurs on the surface of ESI droplets that cannot be accounted for in this type of calculation. Second, log $P_{solvent/vacuum}$ describes the partitioning between solvent and gas phase (a vacuum in this case), while for the ionization, the partitioning between droplet surface and interior is more important.[26] Droplet surface, however, is a phase with poorly understood composition. Third, calculations were done for the initial solvent composition, while it is known that in the ESI spray, the droplets are gradually (and significantly) enriched with the less volatile solvent component (water in this case). Therefore, the composition of the droplets from which the ions are eventually ejected is different from the initial one.

From these data, it seems that changes in molecular size better describe the partitioning between droplet interior and surface than the log$P_{solvent/vacuum}$. Probably ions formed from larger molecules are more favored on the surface of the droplet. The reason is that for large ions, the charge−charge repulsion is smaller as the charges are spaced farther away from each other because of ion size. Still, in our experiments, the increasing volume is accompanied by the increasing hydrophobicity as large hydrophilic molecules are not included in the data set, due to their unavailability. Also the log$P_{solvent/vacuum}$ depends on the total interaction energy of the ion with the solvent, which is lower in the case of larger ions. This total interaction energy may be inadequate for describing interactions between ions on the droplet surface and solvent because the ion is only partially surrounded by solvent molecules.

The correlation between measured log*IE* and , the anion charge delocalization describing parameter,[20] was good (R^2 = 0.67). Higher ionization efficiency was observed for ions with more efficient charge delocalization (lower *WAPS*). This good correlation implies that the charge density of the anion is important for ionization, which can be explained by the charge-to-charge repulsion, occurring on the surface of the ESI droplet and responsible for "ion evaporation" from the droplets. On the other hand, lower *WAPS* also indicates lower solvation energy and lower tendency for ion pairing. As ion needs to "escape" from the droplet to be detected in MS, the solvation energy describes how much energy is needed to overcome the attractive forces between ion and the solvent molecules. At this point, we have not computationally estimated the solvation energies as such but use the *WAPS* term instead.

Strong correlation (R^2 = 0.59) was observed between log*IE* and the number of halogen atoms in the molecule *n*(Hal). However, also other parameters correlate with the number of halogen atoms (molecular volume was found to be in a positive and *WAPS* in negative correlation).

The correlation between Flatsurf parameters calculated with COSMO-RS (namely, the Gibbs free-energy gain of solute moving from solvent to the surface, molecular surface that is in contact with the gas phase, and percent of molecular surface that is in contact with the gas phase for the most surface active conformer) and measured ionization efficiency values was nearly nonexistent (R^2 below 0.2).

Also the Klamt descriptors of general solvation theory calculated with COSMO-RS were correlated with ionization efficiency. Two of the parameters, polarity (sig2) and hydrogen bond basicity (Hb_don3), were statistically insignificant in the model.

Overall, the model was statistically insignificantly different from the model compiled by physically logical parameters (described below). Therefore, the model was rejected and it can be concluded that partitioning (described by Klamt parameters) is not the (only) process responsible for ionization efficiency changes from compound to compound.

Building the Ionization Efficiency Predictive Model. As no single parameter is capable of fully describing the changes of ionization efficiency, multilinear regression analysis was carried out for phenols and benzoic acids, both separately and together. In the case of separate study salicylic acid, 3,5-diiodosalicylic-acid and 3,5-dinitrosalicylic acid were not classified into either of the groups.

Three models were obtained. For benzoic acids, the best model was

$$\log IE = (3.02 \pm 0.25) + (-0.46 \pm 0.05)\cdot WAPS\cdot 10^5 \quad (4)$$

with R^2 = 0.83 and standard error of the log*IE* estimate *S* = 0.35 was obtained (standard deviations of the regression coefficients are given as ±). *WAPS* turned out to be the only statistically

dx.doi.org/10.1021/ac404066v | *Anal. Chem.* XXXX, XXX, XXX−XXX

significant parameter. Absence of acidity in the model is obviously due to the relatively narrow, 10 orders of magnitude, variability of acidities in the group of benzoic acids (pK_a range from 17.44 to 27.97 in MeCN). Also the pK_a values are sufficiently low, leading to full dissociation of all studied benzoic acids in droplets.

On the other hand, for phenols the best model contained both acidity and *WAPS*:

$$\log IE = (3.96 \pm 0.29) + (-0.087 \pm 0.016)pK_a + (-0.27 \pm 0.07)\cdot WAPS\cdot 10^5 \quad (5)$$

with $R^2 = 0.83$ and standard error of the log*IE* estimate $S = 0.46$. In this case, the acidity variability is 25 orders of magnitude (pK_a from 7.18 to 32.55 in MeCN) and the pK_a is a statistically significant parameter in the model. For phenols, the model was of the same goodness independent of whether pK_a in MeCN or water was used, as they are also strongly correlated to each other.

These differences in the models for benzoic acids and phenols indicate that pK_a values, as such, are not important in the ionization process but rather the degree of ionization of analyte in the droplet. From the literature,[27] it is known that during solvent evaporation from the droplets, water concentration significantly increases in the droplets. This means that even though the solvent initially contains 80% of acetonitrile, the droplets from which ions are formed mainly contain water. Therefore, we calculated the degree of ionization (α) for each analyte from the COSMO-RS pK_a values and measured pH of the aqueous phase (0.1% ammonia in water). The pK_a values calculated via COSMO-RS were used as experimental pK_a's are not available for all of the compounds in the training set. Another reason arises from the fact that we would like to make the model usable also for compounds for which measured data are not available, therefore only a computational approach is used.

This degree of ionization was used in the following model. Out of 48 compounds of the training set, the best model was obtained as follows:

$$\log IE = (1.04 \pm 0.34) + (2.23 \pm 0.34)\cdot\alpha + (-0.51 \pm 0.04)\cdot WAPS\cdot 10^5 \quad (6)$$

with $R^2 = 0.83$ and standard error of the log*IE* estimate $S = 0.40$.

Log*IE* values predicted using eq 6 are plotted in Figure 2 against the experimental values and tabulated in Table 1. Figure 2 demonstrates uniform scatter of data points around the line with unity slope and zero intercept, absent of deviating data points and uniform distribution of data points of the three compound families.

It is observed that even though molecular volume showed good correlation with log*IE*, it did not become statistically significant in the multilinear regression model. This most probably arises from the fact that *WAPS* is strongly correlated with the molecular area: the higher the area, the lower the *WAPS*. Molecular area in turn is correlated with molecular volume.

These obtained models are in good agreement with the general understanding of the ESI ionization process. The compounds having lower pK_a values are more extensively dissociated in the droplets and therefore yield a higher concentration of anions in droplets. On the other hand, anions

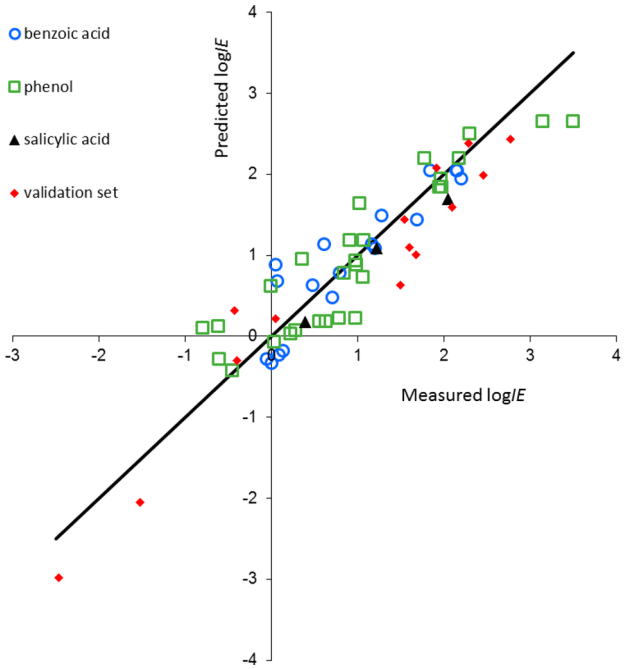


Figure 2. The predicted ionization efficiency values (according to eq 6) against measured values for both training set and validation set). The trendline has zero intercept and unity slope.

that have more delocalized charge (expressed by lower *WAPS*) interact weaker with other molecules and ions and thus have lower solvation energy and lower tendency for ion pairing. Thus, they experience less charge-to-charge repulsion on the surface of the droplet (where the majority of the ions are anions) and at the same time less interaction with the polar water molecules as well as less charge−charge attraction with ions of opposite charge in the droplet interior. The ions with localized charge experience more charge-to-charge repulsion on the surface and more interactions in the droplet interior and are forced into the droplet interior where opposite charges "neutralize" the repulsion effects. The charge delocalization in anions is well-described by *WAPS*: the more delocalized is the charge, the lower is the *WAPS* value.

The obtained model displays quite good consistency with other studies about ion evaporation, determining the charged state of the proteins.[28−30] The critical field strength required for the ion evaporation to happen was found to be dependent on the solvation energy of the ion. In our model (eq 6), the *WAPS* term is strongly correlated to the solvation energy of the ions. Therefore, the ionization efficiency and the charge state (i.e., extent of ionization) of large molecules is likely to be controlled by the same mechanism and therefore also the mathematical models are somewhat similar.

For different analytes, different aspects of the ionization process may become limiting. If the analyte is a very strong acid and therefore mostly dissociated in the droplet, the partitioning to the droplet surface becomes the limiting step. On the other hand, if the analyte has very low *WAPS* value, resulting in a very low charge-to-charge repulsion and low solvation energy, a majority of the anions formed during dissociation are also partitioned to the surface of the droplet and the formation of the anions (dissociation) becomes the limiting process. On the example of benzoic acids, all analytes are almost fully ionized in the droplets ($\alpha \approx 1$); therefore, degree of ionization does not become significant in the model, while partitioning to the

dx.doi.org/10.1021/ac404066v | *Anal. Chem.* XXXX, XXX, XXX−XXX

droplet surface yields the actual log*IE* differences. Therefore, only *WAPS* appears in the predictive model (eq 3). On the other hand, for phenolic compounds the degree of ionization, as well as *WAPS* values, vary on a large scale; therefore, for some compounds dissociation limits ionization while for others moving to the droplet surface limits the ionization. As a result, both α, as well as *WAPS*, appear in the model.

Validating the Ionization Efficiency Predictive Model. In order to validate the model (eq 6) and determine its applicability for a wider range of compounds a validation set of different chemical nature was chosen and measured. The validation set contained 14 compounds from different compound classes, imides, sulphonamides, amino acids, etc. (see Table 1). The log*IE* values measured for the validation set varied from −2.46 to +2.46, thus covering an even wider ionization efficiency range than the training set. The measured and predicted values for the validation set are presented in Table 1 and Figure 2. It can be seen that the agreement between measured values and values predicted by the model is good. The RMSD value, indicating the average disagreement between measured and predicted values, was found (using eq 3) to be 0.48, somewhat higher than observed for the training set. Keeping in mind the variability of the measured log*IE* values, the different compound classes in the validation set and the fit of the model achieved in the training set, the agreement for the validation set is good.

CONCLUSIONS

In this paper, we have extended the previous ionization efficiency studies to negative electrospray ionization (via deprotonation) for 63 compounds altogether (training and validation set). Correlation analysis of log*IE* with different molecular parameters shows that charge delocalization in the anion (*WAPS*) and degree of ionization (α) of the molecule in solution are the most important parameters when describing the efficiency of negative electrospray ionization. Out of all parameters investigated, *WAPS* displays the best single-parameter correlation with log*IE*. A linear model was obtained, which equally well describes the log*IE* of both phenols and benzoic acids (R^2 = 0.83, S = 0.40 log units) and contains only α and *WAPS* as molecular parameters. Both parameters can be easily calculated with the help of the COSMO-RS method.

The model was successfully validated using a validation set of 14 acids belonging neither to phenols nor to benzoic acids, thereby demonstrating the broad applicability and universality of the above-described relationships between IE and molecular properties. Thus, the first promising steps have been taken toward predicting the ionization efficiency of acidic compounds previously not measured in the negative ion mode.

ASSOCIATED CONTENT

Supporting Information

Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: anneli.kruve@ut.ee.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work has been supported by PUT 34 from Estonian Research Council as well as by the Institutional Funding IUT20-14 (TLOKT14014I) from the Estonian Ministry of Education and Science and was in part carried out in the High Performance Computing Centre of the University of Tartu. We are grateful to Dr. Koit Herodes for insightful discussions.

REFERENCES

(1) Iribarne, J. V.; Thomson, B. A. *J. Chem. Phys.* **1976**, *64*, 2284−2294.
(2) Kebarle, P. *J. Mass Spectrom.* **2000**, *35*, 804−817.
(3) Hogan, C. J.; Fernandez de la Mora, J. *Phys. Chem. Chem. Phys.* **2009**, *11*, 8079−8090.
(4) Gamero-Castaño, M.; Fernandez de la Mora, J. *Anal. Chim. Acta* **2000**, *406*, 67−91.
(5) Cech, N. B.; Enke, C. G. *Mass Spectrom. Rev.* **2001**, *20*, 362−387.
(6) Schug, K.; McNair, H. M. *J. Sep. Sci.* **2002**, *25*, 760−766.
(7) Oss, M.; Kruve, A.; Herodes, K.; Leito, I. *Anal. Chem.* **2010**, *82*, 2865−2872.
(8) Huffman, B. A.; Poltash, M. L.; Hughey, C. A. *Anal. Chem.* **2012**, *84*, 9942−9950.
(9) Henriksen, T.; Juhler, R. K.; Svensmark, B.; Cech, N. B. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 446−455.
(10) Nguyen, T. B.; Nizkorodov, S. A.; Laskin, A.; Laskin, J. *Anal. Methods* **2013**, *5*, 72−80.
(11) Chalcraft, K. R.; Lee, R.; Mills, C.; Britz-McKibbin, P. *Anal. Chem.* **2009**, *81*, 2506−2515.
(12) Ehramm, B. M.; Henriksen, T.; Cech, N. B. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 719−728.
(13) Konermann, L.; Ahadi, E.; Rpdriguez, A. D.; Vahidi, S. *Anal. Chem.* **2013**, *85*, 2−9.
(14) Thurman, E. M.; Ferrer, I.; Barcelo, D. *Anal. Chem.* **2001**, *73*, 5441−5449.
(15) Kloepfer, A.; Quintana, J. B.; Reemtsma, T. *J. Chromatogr., A* **2005**, *1067*, 153−160.
(16) Gamero-Castaño, M.; Fernandez de la Mora, J. *J. Mass Spectrom.* **2000**, *35*, 790−803.
(17) Lobowsky, M.; Fenn, J. B.; Fernandez de la Mora, J. *Anal. Chim. Acta* **2000**, *406*, 105−118.
(18) Leito, I.; Herodes, K.; Huopolainen, M.; Virro, K.; Künnapas, A.; Kruve, A.; Tanner, R. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 379−384.
(19) Klamt, A. COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design; Elsevier Science Ltd: Amsterdam, The Netherlands, 2005.
(20) Kaupmees, K.; Kaljurand, I.; Leito, I. *J. Phys. Chem. A* **2010**, *114*, 11788−11793.
(21) Ahlrichs, R.; Bär, M.; Baron, H.-P.; Bauernschmitt, R.; Böcker, S.; Ehrig, M.; Eichkorn, K.; Elliott, S.; Furche, F.; Haase, F.; Häser, M.; Horn, H.; Hattig, C.; Huber, C.; Huniar, U.; Kattannek, M.; Köhn, M.; Kölmel, C.; Kollwitz, M.; May, K.; Ochsenfeld, C.; Öhm, H.; Schäfer, A.; Schneider, U.; Treutler, O.; von Arnim, M.; Weigend, F.; Weis, P.; Weiss, H. *Turbomole*, version 6.3, 2011.
(22) Eckert, F.; Klamt, A. *COSMOtherm*, version C3.0, release 13.01; COSMOlogic GmbH&CoKG: Leverkusen, Germany, 2013; available at http://www.cosmologic.de/.
(23) Tshepelevitsh, S.; Oss, M.; Pung, A.; Leito, I. *ChemPhysChem* **2013**, *14*, 1909−1919.
(24) Zissimos, A. M.; Abraham, M. H.; Klamt, A.; Eckert, F. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1320−1331.
(25) Eckert, F.; Leito, I.; Kaljurand, I.; Kütt, A.; Klamt, A.; Diedenhofen, M. *J. Comput. Chem.* **2009**, *30*, 799−810.
(26) Enke, C. G. *Anal. Chem.* **1997**, *69*, 4885−4948.
(27) Girod, M.; Dagany, X.; Boutou, V.; Broyer, M.; Antoine, R.; Dugourd, P.; Mordehai, A.; Love, C.; Werlich, M.; Fjeldsten, J.; Stafford, G. *Phys. Chem. Chem. Phys.* **2012**, *14*, 9389−9396.

(28) Hogan, C. J.; Carroll, J. A.; Rohrs, H. W.; Biswas, P.; Gross, M. L. *J. Am. Chem. Soc.* **2008**, *130*, 6926−6927.
(29) Hogan, C. J.; Fernandez de la Mora, J. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 158−172.
(30) Allen, S. J.; Schwartz, A. M.; Bush, M. F. *Anal. Chem.* **2013**, *85*, 12055−12061.

dx.doi.org/10.1021/ac404066v | *Anal. Chem.* XXXX, XXX, XXX−XXX