

# Modeling from Theory and Modeling from Data: Complementary or Alternative Approaches? The Case of Ionic Liquids

Alessio Paternò,<sup>[a]</sup> Laura Goracci,<sup>[b]</sup> Salvatore Scire,<sup>[a]</sup> and Giuseppe Musumarra<sup>\*[a]</sup>

In the field of ionic liquids (ILs), theory-driven modeling approaches aimed at the best fit for all available data by using a unique, and often nonlinear, model have been widely adopted to develop quantitative structure–property relationship (QSPR) models. In this context, we propose chemoinformatic and chemometric data-driven procedures that lead to QSPR soft models with local validity that are able to predict relevant

physicochemical properties of ILs, such as viscosity, density, decomposition temperature, and conductivity. These models, which use readily available and easily interpretable VolSurf+ descriptors, represent an unexploited opportunity for experimentalists to model and predict the physicochemical properties of ILs in industrial R&D design.

## 1. Introduction

Inductive and deductive approaches have attracted a long-standing philosophic interest, starting from Plato (everything may be extracted from inborn ideas) and Aristotle, and continued with Galileo Galilei, David Hume, Immanuel Kant, Karl Popper, and Thomas Kuhn, to cite just a few names.

To develop a given field in scientific research, both deductive and inductive approaches may be used. Deductive reasoning is a “top-down” approach aimed at testing an initially postulated hypothesis and then trying to find experimental evidence to support or disprove it. In contrast, inductive reasoning is a “bottom-up” approach based on learning from observations; explanatory hypotheses are eventually formulated towards the end of the process. An inductive approach usually starts with a set of observations, looks for patterns in the data, and then moves from data to theory, from the specific to the general.

The computer age has had an enormous impact on chemical research and given rise to a new field, which was initially named computer chemistry. Now hundreds of molecular-mod-

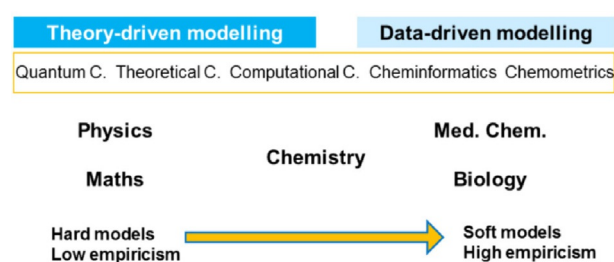


Figure 1. Hard and soft modeling application fields.

eling programs that adopt different investigational approaches are available in different areas of chemistry. For example, from left to right in Figure 1, quantum chemistry, which mainly considers problems related to quantum phenomena; theoretical chemistry, traditionally associated with the formulation of new theories and/or approximations; computational chemistry, a branch of theoretical chemistry in which the objective is to build a mathematical model to calculate molecular properties (e.g. energy, dipole moments, vibrational frequencies); and chemoinformatics, which uses computational and chemometric software to investigate different chemical and biological problems. Although chemometric competence is sought in industry and applied R&D fields, chemometrics, probably due to its highly empirical character, is still not very popular among organic chemists. The term chemometrics was proposed in 1974 by Bruce Kowalski (Seattle, USA) and Svante Wold (Umeå, Sweden) and since then several successful chemometric applications have been reported in pharmaceutical, food, and analytical chemistry. In European scientific societies, such as the Società Chimica Italiana and the Royal Society of Chemistry, the interdisciplinary chemometrics group refers to the Analytical Chemistry Division and is mainly focused on analytical problems.

[a] Dr. A. Paternò, Prof. S. Scire, Prof. G. Musumarra  
Dipartimento di Scienze Chimiche, Università di Catania  
Viale A. Doria 6, 95125 Catania (Italy)  
E-mail: gmusumarra@unicat.it

[b] Dr. L. Goracci  
Laboratorio di Chemiometria e Chemioinformatica  
Dipartimento di Chimica, Università di Perugia  
Via Elce di Sotto 10, 06123 Perugia (Italy)

Supporting Information and the ORCID identification number(s) for the author(s) of this article can be found under <http://dx.doi.org/10.1002/open.201600119>.

© 2016 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

In an elegant paper that included considerations of the psychology and personality types of the scientists, Martens<sup>[1]</sup> noted the gap between the mathematics–statistics culture, which focuses on formal accuracy, and other sciences, which produce an enormous amount of good raw data that are often treated by using limited and uninformed mathematics and statistics. He states that “chemometrics has a lot to learn from other disciplines, mathematics and statistics ... but on the other hand chemometrics has a lot to give to other disciplines” and hopes for a culture that favors warm-hearted cooperation rather than competition. In the same paper, he noted that in the past 40 years science has witnessed a big data explosion paralleled by increased computer capacity with respect to storage space, memory, and CPU power, but unfortunately we are often overwhelmed by this. In this context, Martens stated that chemometrics, in contrast with “black box” approaches, developed a pragmatic scientific culture that attempts to approach the real world by letting the data talk to us but at the same time trying to interpret the results in the light of prior chemical knowledge and the laws of physics.

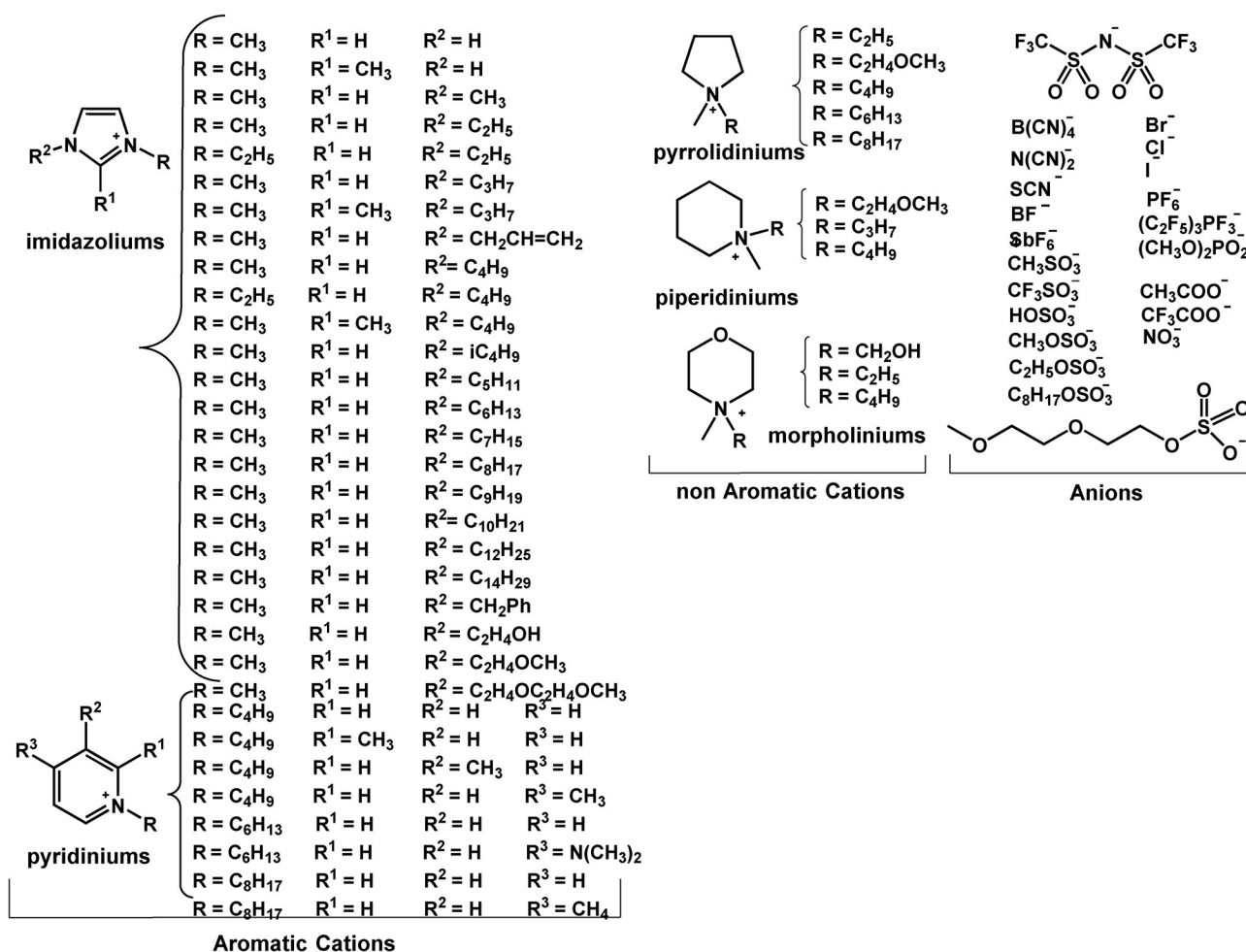
The reason for chemometrics not being applied in organic chemistry has, in our opinion, somewhat paralleled the lack of focus by many Universities on education in physical organic chemistry in the past two decades. This has led organic chem-

ists to delegate these studies to theoreticians (whose aim is to provide a unique model of high complexity able to fit all the data) and to statisticians (whose cultural background emphasizes the importance of high correlation and predictivity, expressed as  $R^2$  and  $Q^2$  values, respectively, which in cases such as biological and physical measurements on ionic liquids (ILs) are difficult to achieve), both believed to be more suited to the job. The field of ILs, low-melting-point salts formed from an organic cation and an inorganic or organic anion, covers a huge experimental space that is difficult to explore; multivariate approaches that lead to soft models with local validity might be useful for their application potential. For this reason, herein we propose the adoption of chemoinformatic and chemometric approaches to model and predict the physicochemical properties for some ILs (Scheme 1), with the aim of providing new opportunities to complement available theory-driven models in the field.

## 2. Results and Discussion

### 2.1. General Survey of IL Modeling Approaches

The increasing number of industrial applications of ILs requires knowledge of their toxicological and environmental properties



Scheme 1. Structures of IL cations and anions.

to comply with the European Union rules on the Registration, Evaluation, Authorization, and Restriction of Chemicals (REACH).<sup>[2]</sup> This has prompted the primary interest in collecting information on the properties and hazards of ILs by using reliable and representative toxicity tests. Available studies addressing specific toxicity “sensors” in different biological systems have been reported in the UFT-Merck Ionic Liquids Biological Effects Database,<sup>[3]</sup> which unfortunately is no longer open access and, even when it was, only provided single toxicity tests for different and in some cases numerically limited numbers of ILs. Indeed, papers that report experimental determinations of both the toxicity and the physicochemical properties of ILs can cover only a very limited portion of the huge potential experimental space, which is estimated to include over one million ILs with different cation and anion combinations.

In this context, the adoption of multivariate approaches has helped to simplify the overall toxicity picture. A multivariate insight into the IL toxicity database recently dealt with four main groups of toxicity: aquatic toxicity, toxicity towards fungi and bacteria, cytotoxicity towards the IPC-81 rat cell line, and acetylcholinesterase enzyme (AChE) inhibition.<sup>[4]</sup> Although several hundreds of ILs were reported in the above database, only 104 aquatic toxicity scores and 87 bacterial and fungi toxicity scores could be derived by performing principal components analysis (PCA) on a matrix that reported experimental data.<sup>[4]</sup>

IL structural features can be related to toxicity and physical properties by using quantitative structure–property relationships (QSPR). QSPR models need 1) good descriptors, 2) good statistical correlation tools, and 3) good experimental data. Available literature models apply different descriptors and correlation tools to different data sets to fit the data after measurements have been made. This situation is of limited use for experimentalists.

In our previous work,<sup>[5]</sup> specific *in silico* structural descriptors for both cationic and anionic counterparts of ILs were recently developed by using the VolSurf+ approach and related to IL properties by using a unique correlation tool, partial least squares (PLS), which provides multiparameter equations with no possibility of collinearity because the descriptors are orthogonal to each other.

The VolSurf+ approach<sup>[6,7]</sup> computes the interaction energies between molecules and four chemical probes and calculates the molecular descriptors derived from the information coded into the 3D GRID molecular interaction fields (MIFs).<sup>[8–11]</sup> VolSurf+ descriptors quantify relevant physicochemical properties, such as molecular size and shape, hydrophilic and hydrophobic regions, hydrogen-bonding ability, interaction energy moments and capacity factors, molecular amphiphilic moments, hydrophilic–lipophilic balance, molecular diffusivity in water as a solvent, partition coefficients in different solvents, pH-dependent water solubility, and molecular flexibility in different solvents. The VolSurf+ descriptors were successfully applied to develop quantitative models for structure–aquatic toxicity<sup>[5]</sup> and structure–polarity relationships.<sup>[12]</sup>

The above IL VolSurf+ descriptors were compacted into nine new *in silico* descriptors (five for cations and four for anions), called IL PPs (principal properties), for 218 cations and

38 anions (hereafter denoted as PP+ and PP–, respectively).<sup>[13]</sup> These descriptors, which have a lower information content than the original VolSurf+ descriptors, have the advantage of providing simpler QSPR models suitable for design purposes. A multivariate approach, such as PLS, was able to correlate quantitatively and simultaneously both cation and anion ILs PPs to IPC-81 cytotoxicity and AChE inhibition for over 200 ILs.<sup>[13]</sup> The practical utility of this approach was demonstrated by the development of a QSPR model that correlated IL structures to *Vibrio fischeri* toxicity by providing a simple three-parameter equation that allowed prediction of IL toxicity toward *Vibrio fischeri* without using chemometric and/or chemoinformatic software.<sup>[14]</sup> The resulting correlation plot is comparable, if not better than, that reported in a subsequent paper,<sup>[15]</sup> in which six QSPR models that used multiparameter equations with between 9 and 12 descriptors were reported. The paper does not provide numerical values for the descriptors, which prevents a check of the results. The authors reported predictions by using “the best model” but at the same time they observe that “classical external validation metrics were unable to portray poor model performances” and this led to the development of new judgment criteria. In our work, all the data are reported in the Supporting Information to allow reproducibility of the results for the examined data set and in a literature reference<sup>[13]</sup> to allow experimentalists to predict for themselves the toxicity of new compounds. The interest in predicting *Vibrio fischeri* toxicity was confirmed in a recent paper that used IL structural descriptors, such as the Gutman topological index, the lopping center information index, and the number of oxygen atoms, in a QSPR model.<sup>[16]</sup>

Research into the conversion from thermal into electric energy has recently focused on the use of ILs in thermoelectrochemical devices. QSPR modeling aimed at the identification of structural features of ILs (mixed with a LiI/I<sub>2</sub> redox couple) that influence the Seebeck effect has been reported.<sup>[17]</sup> In this field, the design capability of IL PPs was demonstrated by using a QSPR model that provided affordable predictions for IL heat capacities, validated by experimental measurements. *In silico* predictions allowed the design of a limited number of structurally different ILs with similar C<sub>p</sub> values, which has provided the possibility to select an optimal IL according not only to its efficiency, but also to its environmental and economic sustainability.<sup>[18]</sup>

The present general procedure, which uses readily available descriptors and adopts an accessible statistical procedure, such as PLS, can be extended to other QSPR models that involve relevant IL PPs. A few applications will be provided below. The choice between adoption of the entire set of VolSurf+ descriptors, with a higher degree of information, or compacted ILs PPs, with a lower information content but easier to handle, will be data-driven by following the ancient motto “*Frustra fit per plura quod fieri potest per pauciora*” (It is vain to do with many what can be done with few).

Viscosity is a very relevant property required for process design in industrial applications, such as heat exchangers, pipelines, or distillation columns. A low viscosity is generally desired for applications of ILs as solvents, to minimize pump-

ing costs and increase mass-transfer rates, whereas higher viscosities may be favorable for other applications, such as lubrication or use in membranes.<sup>[19]</sup> Therefore, it is not surprising that several viscosity QSPR-modeling studies by using different approaches have been reported.

QSPR models for the ionic conductivity and viscosity of ILs with group descriptors, by using the polynomial expansion with a genetic algorithm model based on the type of cation, length of side chain, and type of anion, exhibited relatively good correlation and provided the reverse design of ILs.<sup>[20]</sup> The group contribution method was also applied in QSPR modeling to estimate the viscosity of imidazolium-, pyridinium-, and pyrrolidinium-based ILs that contained several anions and covered wide ranges in temperature.<sup>[19]</sup> A similar approach was adopted to derive a relationship between the viscosity of imidazolium-based ILs and the descriptive parameters of anions and cations by considering temperature, molecular weight, and the number of the branched-chain carbon atoms on the imidazole ring.<sup>[21]</sup>

A group contribution model based on a feed-forward artificial neural network was applied to over 13 000 data points for the temperature- and pressure-dependent viscosity of 1484 ILs published in the open literature in the last three decades. The data were critically revised and divided into training, validation, and testing sets, to develop a new model that allowed *in silico* predictions of the viscosities of ILs on the basis of the chemical structures of their cations and anions as described by 242 building blocks.<sup>[22]</sup>

Many theory-driven QSPR models based on *ab initio* calculations based on CODESSA (comprehensive descriptors for structural and statistical analysis) or COSMO-RS (conductor-like screening model for realistic solvents) methods have been reported. CODESSA derives descriptors by using quantum mechanical methods to develop QSAR (quantitative structure–activity relationship/QSPR) models. A critical analysis of error sources in quantum-chemical computations has been recently reported.<sup>[23]</sup> The CODESSA approach was adopted to establish QSPR correlations for conductivities and viscosities of low-temperature-melting ILs with the bis(trifluoromethylsulfonyl)imide anion. The authors concluded that the models were highly temperature dependent and stressed that the experimental properties of ILs depend heavily on the degree of purity, which cannot always be easily controlled.<sup>[24]</sup> A more recent study at eight different temperatures on ILs that contained bis(trifluoromethylsulfonyl)imide<sup>[25]</sup> concluded that interionic electrostatic interactions are the most important factor that affects viscosity, and this effect changes with temperature. The same research group,<sup>[26]</sup> addressing an extensive database, developed QSPR models and concluded that alongside temperature, pressure, and impurity, the ionic structural characteristics of the IL cation or anion also have significant effects on the viscosity. A QSPR study addressing the viscosity of imidazolium-based ILs<sup>[27]</sup> noted the predominant effects of cation–anion electrostatic interactions whereas other interactions (e.g. interionic hydrogen bonds, van der Waals interactions) or microcharacteristics (e.g. molecular orbitals, electronic population, dipole moments, volume, shape, branching degree, symmetry) provide a minor

contribution. However, it is worth mentioning that this work considered only one heterocyclic scaffold and divided the original dataset into four different sets, each modeled by using multiparameter equations that involved up to 25 “independent” variables with a clear danger of colinearity that would provide overoptimistic correlations. The COSMO-RS approach was adopted in a systematic study of the dynamic viscosity of 27 ILs<sup>[28]</sup> to study anion and cation effects. The COSMO-RS method established relationships between molecular-level features and viscosity data for the investigated families of ions. A QSPR model that considered six molecular parameters by using a genetic function approximation from selected molecular descriptors was developed, and led to a suitable correlative and predictive ability.

The most abundant viscosity data set analyzed by using COSMO-RS molecular descriptors included 1502 experimental data points for 89 ILs under a wide range of temperatures and pressures.<sup>[29]</sup> QSPR linear and nonlinear models were developed and the latter provided better viscosity predictions. Unfortunately, Ref. [29] does not give the opportunity to reproduce the results and to exclude the occurrence of data overfitting due to the lack of numerical values for the descriptors.

An advantage of the COSMO-RS approach is that a unique model can take into account temperature and pressure variations. However, a careful insight into the data shows that in analyzing all available literature data, very similar or even identical values reported in different papers are all included in the analysis. Just to quote an example, seven values are provided for 1-octyl-3-methyl imidazolium hexafluorophosphate at 101.325 kPa and 343.15 K, some in the training set, others in the test set. The inclusion of four to seven values under the same pressure and temperature conditions either in the training or in the test set is common in this theoretical approach. Such a situation, which obviously improves both the model performance and the correlation, would be immediately noticed and not tolerated in a data-based analysis.

To compare the advantages and drawbacks of our modeling approach, we selected the same abundant literature database<sup>[29]</sup> (see Data in the Supporting Information). The inductive data-driven model for analysis of the viscosity data was the PLS approach,<sup>[30,31]</sup> a known multivariate procedure aimed at finding relationships between an  $x$  descriptor matrix (in the present case, the cation and anion IL PPs) and a  $y$ -dependent variable (in the present case,  $\log(\text{viscosity})$ ) for which numerical values are reported (see Data in the Supporting Information). IL PPs<sup>[13]</sup> are compacted descriptors for both IL anionic and cationic counterparts that are orthogonal to each other and, therefore, can be used for multivariate experimental design. In PLS the data are pre-processed by autoscaling all variables to unitary variance, that is, by multiplying the variables by appropriate weights (the reciprocal of the variable standard deviation) to give them unit variance (i.e. the same importance).

A simple chemometric tool, the IL physicochemical (ILPC) predictor, which provides a preliminary qualitative prediction of properties, such as viscosity, has recently been reported.<sup>[32]</sup> The authors state that viscosity depends mainly on hydrogen-bond formation, but “unfortunately this type of intermolecular



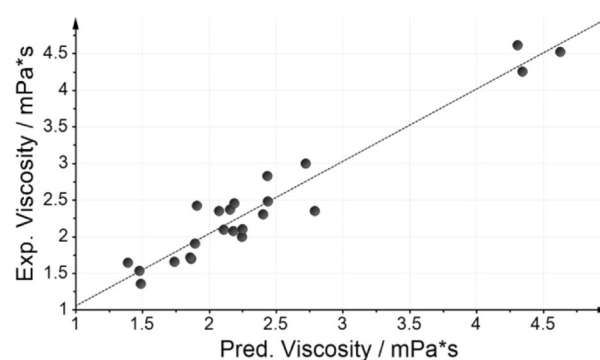
relation is not covered by our set of descriptors". Our approach, which uses hydrogen-bonding molecular descriptors to develop a quantitative relationship, can be considered an extension and a further development of this work.

The literature database<sup>[29]</sup> reports viscosity data at different temperatures and pressures. In our data analysis, the PLS procedure should be applied to data obtained under the same experimental conditions. In the case of multiple literature values, only one was considered, either the most reproducible one or an average value. Initially, to obtain information on the data structure,  $\log(\text{viscosity})$  values at fixed temperature and pressure (283.15 K and 101 kPa, respectively) for 23 ILs were used as the dependent variable and the corresponding nine PPs (five for cations and four for anions) as the descriptor variables in a PLS model. In such a model, three significant PLS components describe 94.9% of the total  $y$  variance with a predictive ability of 0.872 (Table S1). In the VIP (variable importance for projection) plot, which shows the importance of each  $x$  variable in explaining  $x$  variation and correlation to  $y$  (Figure S1),  $PP_{-2}$  and  $PP_{-3}$  are the most important descriptors, followed by  $PP_{+1}$  and  $PP_{+5}$ , whereas all other descriptors appear to be less relevant. To limit the number of descriptors and simplify the model, a new PLS correlation model was built by retaining only four relevant  $x$  descriptors:  $PP_{+1}$  and  $PP_{+5}$  for the cations and  $PP_{-2}$  and  $PP_{-3}$  for the anions. The new simplified  $23 \times 4$  matrix provided a two-PLS-components model that explained 92.6% of the total variance, with a cumulative  $Q^2$  value of 0.903 (see Table S2 for model details), which showed that the exclusion of five low-relevance descriptors improved the "goodness" of this model by considering simultaneous variations in both the cation (heterocyclic core, side-chain length, presence of oxygen atoms in the side chain) and the anion structural features described by only four descriptors (PPs). The correlation between the predicted and experimental values is reported in Figure S2.

There is no simple or unique criterion to decide how many PCs should be considered significant, and  $Q^2$  is only one possibility. It has been noted<sup>[33]</sup> that a high  $Q^2$  value is a necessary but not sufficient condition for the model to have high predictive power and that an external validation with at least five compounds with different structural features that covers the range of measured properties should be used. Since then, the principles of internal and external QSAR model validation have been clearly defined<sup>[34]</sup> and more recently the statistical criteria for evaluation of the external predictivity have been widely discussed.<sup>[35–38]</sup> To check the predictive power of the model by external validation as well, we randomly selected two sets of five structurally different ILs across the experimental viscosity range as external test sets. The statistical parameters of the resulting models, with 18 ILs in the learning sets and 5 ILs in the test sets, are reported in Table S3, and the predicted versus experimental viscosity values for the test-set ILs are given in Table S4. The  $Q^2$  values for the 18 IL models (0.833 and 0.872) are comparable to that of the previous model (0.903). The agreement between experimental viscosity values and the predicted ones for both test sets, reported in Table S4, provides

experimental validation support for the high  $Q^2$  value in the model with 23 ILs.

Theory-driven approaches aim at the best fit for all available data by using a unique, often nonlinear, model, whereas the SIMCA (soft independent modeling of class analogy) approach<sup>[39]</sup> aims at raw data reduction by compacting them into data of higher relevance and eventually adopting different soft models of local validity, which provides more easily interpretable results. In this context we carried out different PLS models at nine different temperatures for which the statistical parameters are reported in Table S2. The correlation plot for 293.15 K is reported in Figure 2 and those for other temperatures are given in Figure S2.



**Figure 2.** Predicted vs. experimental  $\log(\text{viscosity})$  values at 293.15 K ( $R^2 = 0.93$ ).

One advantage of this approach is that  $\log(\text{viscosity})$  values at different temperatures can be easily calculated by using four parameters [Eqs. (1)–(9)].

$$\log \eta_{283.15} = 2.43378 - 0.0467197 PP_{+1} + 0.246482 PP_{+5} - 0.321282 PP_{-2} - 0.299658 PP_{-3} \quad (1)$$

$$\log \eta_{293.15} = 2.12028 - 0.05129 PP_{+1} + 0.200276 PP_{+5} - 0.29018 PP_{-2} - 0.2723 PP_{-3} \quad (2)$$

$$\log \eta_{298.15} = 1.94932 - 0.0719465 PP_{+1} + 0.11487 PP_{+5} - 0.259657 PP_{-2} - 0.197003 PP_{-3} \quad (3)$$

$$\log \eta_{303.15} = 1.962 - 0.0450731 PP_{+1} + 0.165658 PP_{+5} - 0.207737 PP_{-2} - 0.214209 PP_{-3} \quad (4)$$

$$\log \eta_{313.15} = 1.74491 - 0.0563375 PP_{+1} + 0.148687 PP_{+5} - 0.173187 PP_{-2} - 0.176367 PP_{-3} \quad (5)$$

$$\log \eta_{323.15} = 1.55596 - 0.047622 PP_{+1} + 0.0926056 PP_{+5} - 0.145431 PP_{-2} - 0.154512 PP_{-3} \quad (6)$$

$$\log \eta_{333.15} = 1.48278 - 0.0501674 PP_{+1} + 0.124556 PP_{+5} - 0.126096 PP_{-2} - 0.136255 PP_{-3} \quad (7)$$

$$\log\eta_{343.15} = 1.32238 - 0.0423298 \text{PP}_{+1} + 0.0930893 \text{PP}_{+5} - 0.116262 \text{PP}_{-2} - 0.134434 \text{PP}_{-3} \quad (8)$$

$$\log\eta_{353.15} = 1.22789 - 0.041984 \text{PP}_{+1} + 0.106297 \text{PP}_{+5} - 0.0979848 \text{PP}_{-2} - 0.11833 \text{PP}_{-3} \quad (9)$$

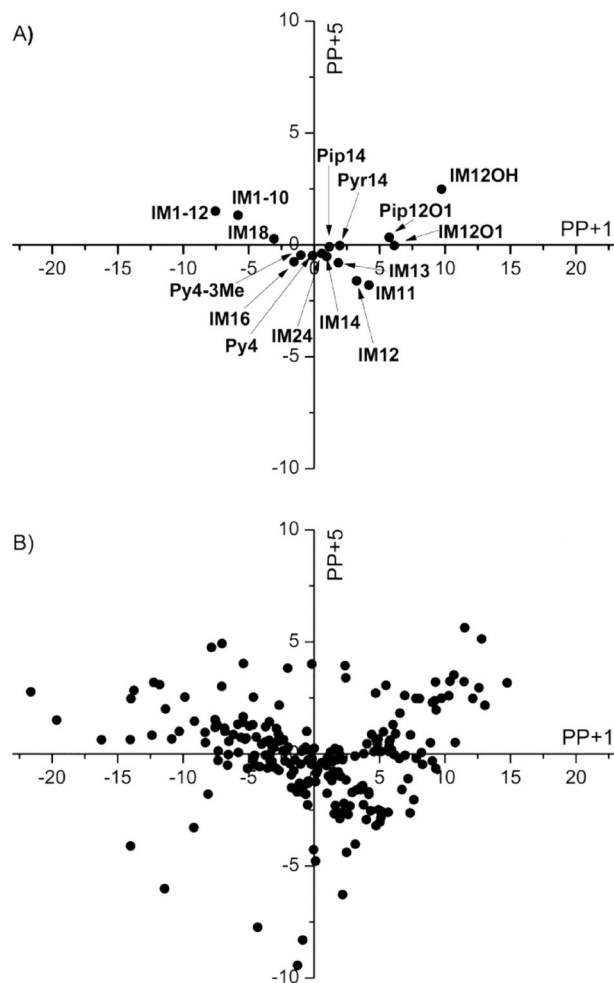
The coefficients of Equations (1)–(9) indicate that the importance of both anionic descriptors decreases as the temperature increases, whereas only the  $\text{PP}_{+1}$  cationic descriptor exhibits the same trend (Figure S3). This can be interpreted by considering that the directional polar interactions, such as the hydrogen-bond interactions (expressed by  $\text{PP}_{+5}$ ), are less efficient as the temperature increases. At 353.15 K,  $\text{PP}_{+5}$  appears to have the same importance as  $\text{PP}_{-2}$  and  $\text{PP}_{-3}$ .  $\text{PP}_{+1}$ , which is a contribution of molecular descriptors that are positively and negatively influenced by temperature, is not affected by temperature variation and anyway provides a lower contribution.

The physicochemical interpretation of cation and anion PPs has been commented on previously.<sup>[13]</sup> In particular,  $\text{PP}_{-2}$  is related to the hydrophilic/hydrophobic character, whereas  $\text{PP}_{-3}$  is related to anionic size/shape and to the ability to form hydrogen bonds as a donor or acceptor. The latter intermolecular interaction, not covered by previous descriptors<sup>[32]</sup> and described by  $\text{PP}_{-3}$ , accounts for the good predictability of the present model.  $\text{PP}_{+1}$  embodies information related to cation solubility, size, flexibility, and molecular weight. High values for  $\text{PP}_{+1}$  indicate high solubility in water, whereas low values are related to molecular size and shape and to solubility in organic solvents, and  $\text{PP}_{+5}$  discriminates the hydrogen-bond donor/acceptor ability.

Another advantage of the adopted approach is that the results can be summarized into plots that allow interpretation and design in addition to data prediction. In Figures 3 and 4, respectively, we report the cation and anion experimental space explored by using our data analysis as compared with the potential experimental space, which could be covered by the cation and anion PPs reported in Ref. [13]. ILs with high viscosity have cations with negative  $\text{PP}_{+1}$  values and anions located in the lower-left quadrant of Figure 4b.

The same considerations can be drawn from inspection of Figure 5, in which the VIPs for cation and anion descriptors are reported. The latter has a higher effect on viscosity; in particular  $\text{PP}_{-2}$  is the most important descriptor in determining viscosity, that is, the lower the  $\text{PP}_{-2}$  value (chlorides and iodides), the higher the viscosity. Accordingly, anions with high  $\text{PP}_{-2}$  values are expected to exhibit low viscosity. Tetracyanoborate and tricyanomethanide, which exhibit the highest  $\text{PP}_{-2}$  values in Figure 4b (7.57 and 6.89 respectively),<sup>[13]</sup> are hydrophobic anions used to generate hydrophobic ILs<sup>[40]</sup> with interesting applications in dye-sensitized solar cells. It has been reported<sup>[41,42]</sup> that a low viscosity seems to be associated with the use of these anions.

This is in excellent agreement with the results of our analysis despite the fact that tetracyanoborate and tricyanomethanide anions (Figure 4b) are both outside the experimental space explored by using anion PPs in our model (Figure 4a).

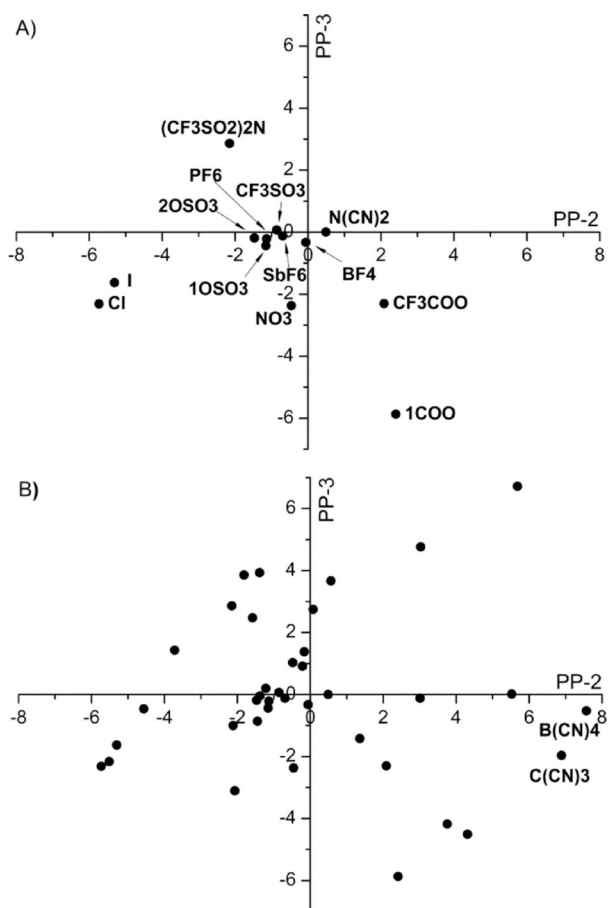


**Figure 3.** A) The cation  $\text{PP}_{+1}/\text{PP}_{+5}$  descriptor space explored by using the PLS model as compared with B) the  $\text{PP}_{+1}/\text{PP}_{+5}$  available descriptor space.<sup>[13]</sup>

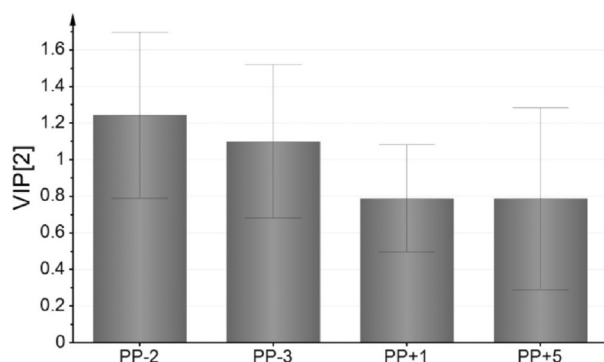
The above consideration indicates that Figures 3 and 4 provide *in silico* design opportunities that can be handled directly by an experimentalist who will be able to evaluate the synthetic affordability and confirm the IL behavior of potential IL candidates for specific applications. The simplicity and the practical utility of this approach in R&D studies of ILs are evident.

## 2.2. Density

Liquid density is a crucial physical property required in the industrial process design of equipment, such as condensers, reboilers, liquid–liquid two-phase mixer–settler units, in liquid metering calculations, and in material and energy balances that involve liquids, vapor–liquid, and liquid–liquid separation processes.<sup>[21,43]</sup> Thus there is longstanding interest in the prediction of IL densities by using several approaches, from Seddon's early studies that used a surface-tension-weighted molar volume, the parachor,<sup>[44]</sup> to QSPR modeling based on semiempirical calculations with 11 molecular descriptors,<sup>[43]</sup> to COSMOS-RS based on quantum-chemistry calculations,<sup>[45]</sup> to



**Figure 4.** A) The anion  $PP_{-2}/PP_{-3}$  descriptor space explored by using the PLS model as compared with B) the  $PP_{-2}/PP_{-3}$  available descriptor space.<sup>[13]</sup>



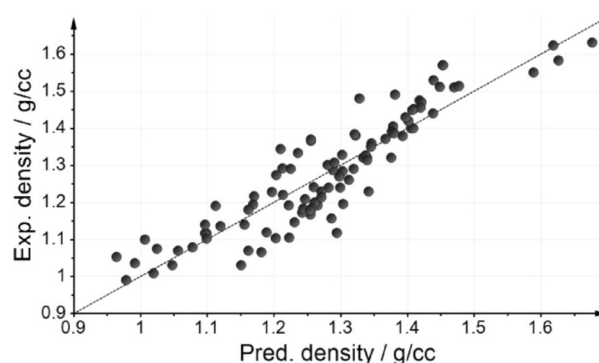
**Figure 5.** VIP bar plot for the viscosity PLS model ( $T = 283.15$  K) displaying the importance of each PP.

a group contribution method that uses the Patel-Teja equation.<sup>[46]</sup>

Recent studies<sup>[47]</sup> suggest that the use of semiempirical methods (faster and less expensive than *ab initio* ones) for geometry optimization provide comparable QSPR models to predict the density of 66 ILs.

Our approach, based on the literature,<sup>[48–53]</sup> considered as many as 109 density values (see Data in the Supporting Infor-

mation). A preliminary PLS analysis (model D1) by using nine PPs as descriptors, the statistical parameters of which are reported in Table S5, revealed that anions such as long-chain sulfates,  $SbF_6$ , bromides, iodides, and nitrates deviate from the linear correlation (Figure S4 a). Deviation from linear behavior may be ascribed to size differences between ions and packing effects.<sup>[50]</sup> Exclusion of the above anions led to soft model D2 for 98 ILs by using nine PPs (five for cations and four for anions) as descriptors. The statistical parameters are reported in Table S5 and the correlation plot and VIP plots are given in Figures S4 b and S5, respectively. In analogy to the procedure adopted for viscosity, a new simplified PLS correlation model included four relevant X descriptors:  $PP_{+1}$  and  $PP_{+2}$  for the cations and  $PP_{-1}$  and  $PP_{-3}$  for the anions (model D3). This model explained 80.2% of the total variance and provided the correlation plot reported in Figure 6.



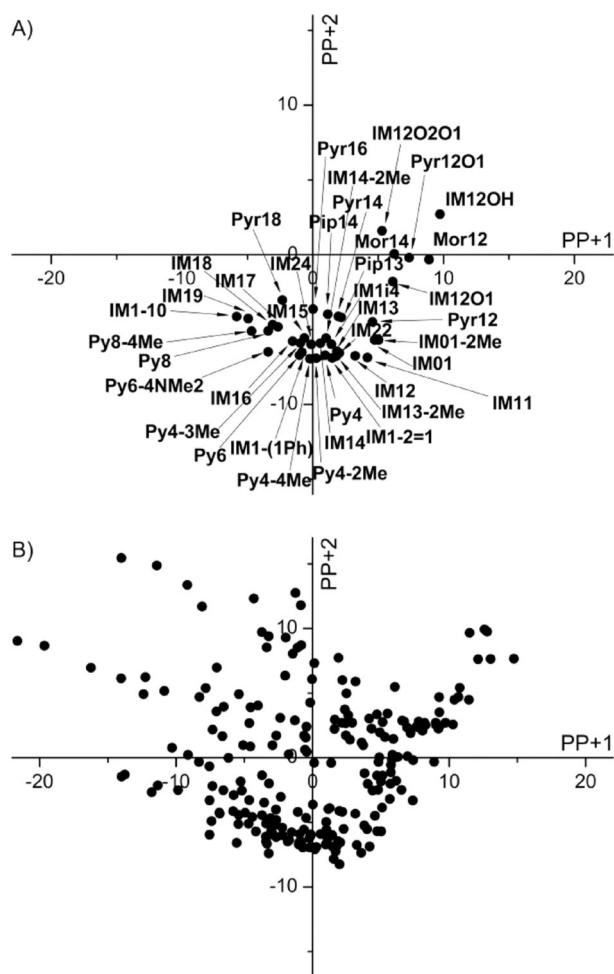
**Figure 6.** Predicted vs. experimental densities from model D3 ( $R^2 = 0.82$ ).

Two sets of ten structurally different IL external sets that covered the experimental density range were selected randomly. The statistical parameters of the resulting models, with 88 ILs in the learning sets and 10 ILs in the test sets, are reported in Table S6 and the predicted versus experimental density values are given in Table S7. The  $Q^2$  values for both 88 IL models (0.797 and 0.776) are comparable to that of the previous model (0.802), which provides external validation for the model.

Figures 7 and 8, respectively, show the cation and anion PPs experimental spaces explored by the present data analysis, compared with those spanned by all PPs in Ref. [13]. It is worth mentioning that the cation experimental space includes many heterocyclic scaffolds. No cation is present in the upper-left quadrant of Figure 7 a because those in the same quadrant of Figure 7 b, characterized by long alkyl chains, are not liquid at 298.15 K.

Predicted density values at 298.15 K can be easily calculated by using the following four-parameter equation [Eq. (10)] with the descriptor values reported in Ref. [13]:

$$D_{298.15K} = 1.1968 + 0.0138095 P_{+1} - 0.0061904 P_{+2} + 0.0244469 P_{-1} + 0.0646624 P_{-3} \quad (10)$$

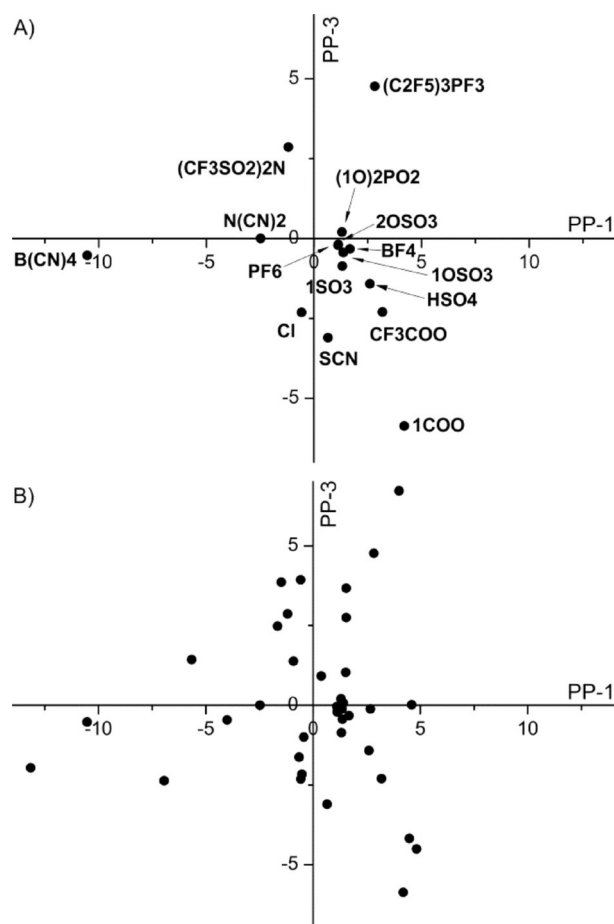


**Figure 7.** A) The cation  $PP_{+1}/PP_{+2}$  descriptor space explored by using the PLS model as compared with B) the  $PP_{+1}/PP_{+2}$  available descriptor space.<sup>[13]</sup>

From Equation (10), it is evident that anions have a higher effect on density, in particular as quantified by the  $PP_{-3}$  descriptor. ILs with high density exhibit high  $PP_{-3}$  values (e.g. tris(pentafluoroethyl)trifluorophosphate and 1,1,1-trifluoro-*N*-(trifluoromethylsulfonyl) methanesulfonamide) whereas low-density ILs have low  $PP_{-3}$  values (e.g. acetates, thiocyanates, and chlorides). High  $PP_{-3}$  values are related to high anionic size, surface, and polarizability,<sup>[13]</sup> which result in higher densities, whereas low  $PP_{-3}$  values indicate a high anion ability to form hydrogen bonds that result in lower densities. As previously illustrated for viscosity, the above plots provide insights for experimental design, although great caution should be adopted because the developed soft model has local validity and cannot be applied to anions not included in the model derivation.

### 2.3. Decomposition Temperature

The decomposition of ILs represents an essential physicochemical property to evaluate their thermal stability and, therefore, the industrially applicable temperature range. The decomposition temperatures for 35 ILs are available in the literature<sup>[48,50,54]</sup>

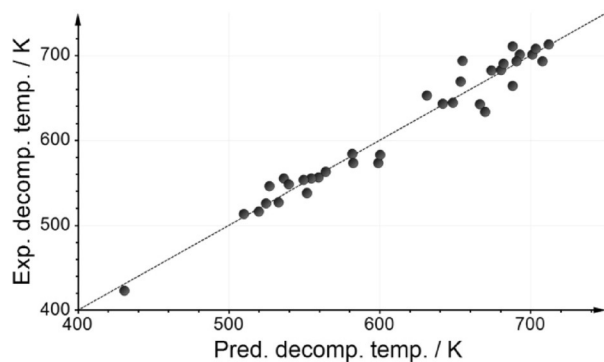


**Figure 8.** A) The anions  $PP_{-1}/PP_{-3}$  descriptor space explored by using the PLS model as compared with B) the  $PP_{-1}/PP_{-3}$  available descriptor space.<sup>[13]</sup>

(see Data in the Supporting Information). A preliminary PLS analysis by using PPs as descriptors provided a statistically unreliable model (M1), as shown by the parameters reported in Table S8. Therefore, to improve the modeling ability and achieve more accurate predictions, a new PLS analysis that used all the original VolSurf+ descriptors and their product terms, and also took into account cation–anion interactions, was performed. The resulting PLS model (M2) included 35 objects (ILs) and 244 variables (128 cation descriptors, 48 anion descriptors, and 48 cross terms) and provided a nonsignificant first PLS component, a second significant PLS component, and a third nonsignificant PLS component (negative  $Q^2$  value, Table S8).

This situation is encountered for particular data structures for which the orthogonal information in the  $x$  block is strong. This is the case for the present data structure. A chemometric analysis able to handle such matrices is the orthogonal PLS (OPLS) approach,<sup>[55–57]</sup> which is able to discriminate between the predictive and orthogonal  $x$  variation that results in a higher  $Q^2$  value. The OPLS model provided, in addition to the predictive OPLS component, seven statistically significant orthogonal PLS components (Table S9). The correlation plot in Figure 9 can be considered as very satisfactory given that it results from an OPLS model with an optimum  $Q^2$  value (see discussion in the Supporting Information).





**Figure 9.** Predicted vs. experimental decomposition temperatures from model M3 ( $R^2 = 0.96$ ).

The predictive power of the model underwent external validation by randomly selecting two sets of five structurally different ILs that covered the experimental decomposition temperature range. The statistical parameters of the resulting models are reported in Table S10 and the predicted versus experimental temperature values for both test sets of ILs are given in Table S11. The  $Q^2$  values for both 30 IL models (0.707 and 0.680) are only slightly lower than that for the 35 IL model (0.795), which provides external validation support for the 35 IL model.

In addition to more accurate predictions, the OPLS analysis with all the original VolSurf+ descriptors and their product terms, which also take into account cation–anion interactions, provides a more detailed interpretation of the descriptors. Table 1 reports the OPLS coefficients higher than 0.03 and lower than  $-0.03$  for cation and anion VolSurf+ descriptors and for their interactions.

Interestingly, Table 1 shows that the decomposition temperature decreases as cation–anion van der Waals interactions increase (D3\_CatxAn, D4\_CatxAn, D5\_CatxAn, D7\_CatxAn), whereas it increases with the increase in hydrogen-bond-derived polar interactions (W5\_CatxAn, W6\_CatxAn, CW5\_CatxAn, CW6\_CatxAn). Therefore, not surprisingly, the thermal stability data for ILs are influenced by both anionic and cationic components, and this finding may allow us to modulate the degradation processes. Such an easy computational approach, which leads to the prediction of the thermal degradation of ILs, may allow selective and application-driven design.

## 2.4. Conductivity

Of the common properties of ILs, conductivity is of crucial importance for their potential applications as electrolytes in electrochemical devices. For example, when ILs are applied as electrolysis solutions for batteries, larger ionic conductivities are required.<sup>[20, 58, 59]</sup>

Conductivity data for 43 ILs are available in the literature<sup>[48, 49]</sup> (see Data in the Supporting Information). In this dataset, the conductivity values for 1-hexyl- and 1-octyl-3-methylimidazolium bromides (12.06 e and  $10.55 \text{ S m}^{-1}$ , respectively) are significantly different from all others (in the range of 0.007–

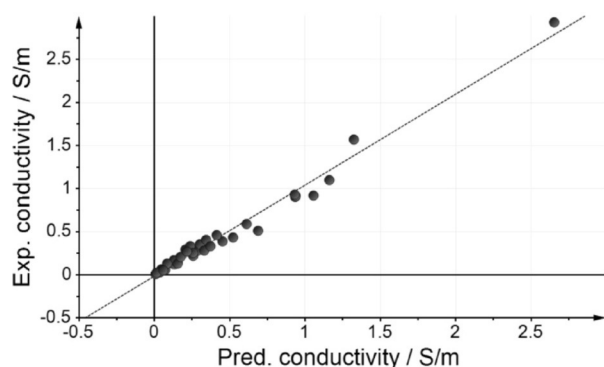
**Table 1.** Coefficients (scaled and centered) for the VolSurf+ descriptors in the decomposition-temperature OPLS model.

Var ID	Coeff. [+]	Var ID	Coeff. [–]
CW2_An	0.1093	IW3_An	–0.0306
CW3_An	0.1013	D5_CatxAn	–0.0312
CW4_An	0.0943	D7_CatxAn	–0.0320
CW1_CatxAn	0.0886	R_CatxAn	–0.0325
W1_An	0.0824	HL1_An	–0.0357
A_An	0.0752	D4_CatxAn	–0.0386
W2_An	0.0737	CW8_An	–0.0413
ID1_An	0.0713	W8_An	–0.0413
ID1_CatxAn	0.0711	D3_CatxAn	–0.0450
ID2_CatxAn	0.0706	R_Cat	–0.0465
W3_An	0.0632	CP_CatxAn	–0.0481
DD7_Cat	0.0602	S_CatxAn	–0.0493
CP_An	0.0582	CW7_An	–0.0503
D2_An	0.0576	V_CatxAn	–0.0512
D3_An	0.0545	IW1_An	–0.0530
ID3_An	0.0513	W7_An	–0.0550
ID4_An	0.0496	ID4_CatxAn	–0.0592
D1_An	0.0488	IW4_An	–0.0778
W4_An	0.0487	CW6_An	–0.1085
W5_CatxAn	0.0483	HL2_An	–0.1253
ID2_An	0.0471	W6_An	–0.1280
HL2_CatxAn	0.0464		
CW5_CatxAn	0.0463		
W6_CatxAn	0.0430		
CW6_CatxAn	0.0429		
A_CatxAn	0.0415		
CD2_An	0.0399		
CD3_An	0.0355		
CW5_An	0.0331		

$2.93 \text{ S m}^{-1}$ ). These values, reported by Li et al.,<sup>[60]</sup> were excluded from the dataset.

Recent studies<sup>[61]</sup> adopted the logarithm of conductivity for QSPR modeling. A preliminary PLS analysis by using nine PPs as descriptors and  $\log(\text{conductivity})$  as the dependent variable provided a statistically unreliable model (Table S12), probably because two ILs (IM01 1COO and IM01 CF<sub>3</sub>COO) are outside the confidence ellipse of the scores plot (Figure S8). This could be due to the fact that in these two ILs, only one imidazolium ring nitrogen has an alkyl substituent, whereas all other imidazolium rings have two alkyl substituents. Therefore, the above two ILs, which behave as outliers with respect to all other imidazoliums in the dataset, were excluded and the corresponding PLS model was derived (C2, see Table S12). However, the statistical parameters of this model were not satisfactory, probably due to the limited ability of compacted PPs to describe conductivity, which suggests that new OPLS analysis must be carried out by using all the original VolSurf+ descriptors and their product terms, which also take into account cation–anion interactions (C3, Table S13). It is perhaps worth mentioning here that the exclusion of four ILs was not arbitrary, but suggested by the adopted data-driven approach that used data examination and the inspection of structural features.

The statistical parameters for the soft OPLS model provided a good predicting ability ( $Q^2 = 0.833$ ) and a very satisfactory correlation plot, recorded in Figure 10 ( $R^2 = 0.98$ ). For external validation we randomly selected two sets of five structurally



**Figure 10.** Predicted vs. experimental log(conductivity) values from model C3 ( $R^2 = 0.98$ ).

different ILs that covered the experimental conductivity range. The statistical parameters for the resulting models are reported in Table S14 and the predicted versus experimental conductivity values for these test-set ILs are given in Table S15. The  $Q^2$  values for the validation models (0.782 and 0.746), similar to that of the previous model (0.833), provide external validation for the model.

Table 2 reports the OPLS coefficients higher than 0.03 and lower than  $-0.03$  for cation and anion VolSurf+ descriptors and for their interactions. Table 2 shows that experimental conductivity data are also influenced by cation–anion interactions,

Table 2. Coefficients (scaled and centered) for the VolSurf+ descriptors in the conductivity OPLS model.			
Var ID	Coeff. [+]	Var ID	Coeff. [–]
CW2_An	0.1062	V_An	−0.0320
ID4_CatxAn	0.1050	W1_Cat	−0.0324
CW3_An	0.1013	IW2_Cat	−0.0325
CW1_An	0.0965	CD1_Cat	−0.0329
CW4_An	0.0912	IW2_An	−0.0353
D3_CatxAn	0.0869	D2_Cat	−0.0373
CP_CatxAn	0.0863	D3_Cat	−0.0386
CW5_An	0.0857	DD6_Cat	−0.0433
IW3_An	0.0811	A_Cat	−0.0486
CW1_CatxAn	0.0742	CW1_Cat	−0.0500
D2_CatxAn	0.0696	PB_Cat	−0.0514
D1_Cat*An	0.0695	R_An	−0.0524
W2_An	0.0627	CD3_Cat	−0.0540
ID1_An	0.0601	IW1_An	−0.0546
DD2_Cat	0.0568	CD2_Cat	−0.0582
W3_An	0.0562	W8_An	−0.0693
DD5_Cat	0.0539	CW8_An	−0.0727
A_CatxAn	0.0498	DD3_Cat	−0.0737
ID2_Cat	0.0491	R_CatxAn	−0.0810
ID2_CatxAn	0.0474	FLEX_RB_Cat	−0.1110
G_CatxAn	0.0441	R_Cat	−0.1167
DD1_Cat	0.0429		
W1_An	0.0401		
W4_An	0.0374		
DIFF_Cat	0.0373		
CD5_Cat	0.0349		
CW2_CatxAn	0.0340		
D4_CatxAn	0.0331		
G_Cat	0.0304		

as demonstrated by the coefficients of the descriptors ID2, ID3, ID4, D1, D2, D3, D4, CP, and A for both cation and anion moieties. The interpretation is that hydrophobic spots on cation and anion partners (Dx) and their locations at the molecular surface (IDx), plus the amphiphilic moment and critical packing (A and CP), influence the IL packing and the viscosity of the mixture, and thus have an effect on the overall conductivity.

## 2.5. Melting Point and Glass-Transition Temperature

The use of ILs at an industrial scale requires knowledge of their melting points and glass-transition temperatures, which are needed to set feasible temperature operation ranges. The experimental determination of solid–liquid phase transitions of ILs cannot be clearly distinguished into melting points and glass-transition temperatures because many samples begin to melt after the glass transition and no distinct peaks can be observed.<sup>[62]</sup> As expected, no significant PLS models were obtained by using PPs or VolSurf+ descriptors, the derivation of which, and consequently the use of, are relative to the liquid phase.

## 3. Conclusions

Theory-driven approaches aim at the best fit of all available data by using a unique, often nonlinear, model. To demonstrate the superiority of the adopted model, most papers report the plot of predicted versus experimental data, which provides a better correlation than that of other models. However, interpretation of the results is not always easy and in many cases only the numerical values of experimental and predicted properties, but not the numerical values of the descriptors, are reported. This prevents other researchers from reproducing the results, a condition always required for the publication of papers that involve experimental procedures, but rarely for computational results.

Conversely, the data-driven approach presented herein starts with an overview of the raw data and then compacts it into data of higher relevance that is eventually modeled by using different soft models with local validity. The domain of validity of each model was determined by the “objects” included in the learning set, herein IL cations and anions present in the learning set. Descriptor availability allows the readers to reproduce the results and perform a simple interpretation of their relevance, and the methodology is more flexible because it may adopt different data-modeling techniques depending on the purpose of the investigation and on the structure of the data.

In conclusion, data-driven chemometrics and chemoinformatic approaches are an unexploited opportunity for experimentalists to model, predict, and design the physicochemical properties of ionic liquids. Modeling from data complements theory-driven approaches for interpretation and correlation purposes and may represent an alternative for experimental design in industrial applications.

## Computational Methods

Herein, chemometric tools available in the SIMCA software package,<sup>[39]</sup> namely partial least squares projections to latent structures (PLS) and orthogonal PLS (OPLS), were used. Relationships between in silico molecular properties ( $x$  matrix) derived by using VolSurf+ and the “response”  $y$  (IL PPs) can be achieved by using PLS analysis,<sup>[30,31]</sup> in which  $y$  can be described as a function of the  $x$  matrix. The PLS algorithm computes PLS components and simultaneously looks for a linear relationship between the  $x$  scores and  $y$  by using Equation (11), in which  $b_a$  is a proportionality coefficient:

$$y_{ia} = \sum b_a t_{ia} + h_{ia} \quad (11)$$

The algorithm used in SIMCA is iterative for each dimension and consists of finding the latent variables of the  $x$  matrix  $t_{ia}$  that maximize the relationship between  $y_i$  and  $t_i$ .

The predictive power of a PLS/OPLS model was assessed by using the  $Q^2$  value, which expresses the fraction of predicted variation. It is usually evaluated by using cross-validation (CV) techniques. For the cases studied herein, the CV process was performed by building reduced models (models for which some of the objects were removed) and using them to predict the  $y$  variables of the held-out objects. Then the predicted  $y$  was compared with the experimental  $y$  and for each model the following dimensionality index was computed [Eq. (12)]:

$$Q^2 = 1 - \left[ \frac{\sum (y - y')}{\sum (y - \bar{y})} \right] \quad (12)$$

in which  $y$  is the experimental value,  $y'$  is the predicted value, and  $\bar{y}$  is the average value. In particular, CV in SIMCA<sup>[39]</sup> was performed by dividing the dataset into seven groups, with an equal (or nearly equal) number of objects in each one, and applying the above CV procedure to them.

The PLS method identifies the variables in the  $x$  block that are relevant to determine the dependent variable  $y$  by using the VIP values. The latter reflects the importance of the variables both with respect to  $y$ , that is, its correlation to the response, and with respect to  $x$ .

The OPLS<sup>[55–57]</sup> method, a modification of the PLS method,<sup>[63]</sup> separates the systematic variation in  $x$  into two parts, one that is linearly related to  $y$  and one that is unrelated (orthogonal) to  $y$ . This partitioning of the  $x$  data facilitates model interpretation and improves model predictivity.<sup>[55–57]</sup> The OPLS model is comprised of two modeled variations, the  $y$ -predictive (TPP T) and the  $y$ -orthogonal (TOPO T) components. Only the  $y$ -predictive variation is used for the modeling of  $y$  (TPCP T), see Equations (13) and (14).

$$\text{Model of } x : x = \text{TPP T} + \text{TOPOT} + E \quad (13)$$

$$\text{Model of } y : y = \text{TPCP T} + F \quad (14)$$

in which  $E$  and  $F$  are the residual matrices of  $x$  and  $y$ , respectively.

## Acknowledgements

We thank Prof. Giorgio Montaudo for helpful discussions and the University of Catania for financial support (FIR project ECDF5E) and for a PhD grant (to A.P.).

**Keywords:** chemoinformatics • ionic liquids • molecular modeling • physicochemical properties • structure–activity relationships

- [1] H. Martens, *J. Chemom.* **2015**, *29*, 563–581.
- [2] REACH regulation available at: <http://www.hse.gov.uk/reach/>.
- [3] UFT-Merck Ionic Liquids Biological Effects Database available at: [http://www.il-eco.uft.unibremen.de/?page=home&chent\\_id=&view=intro&lang=en](http://www.il-eco.uft.unibremen.de/?page=home&chent_id=&view=intro&lang=en); accessed in September 2014.
- [4] A. Paternò, F. D'Anna, G. Musumarra, R. Noto, S. Scirè, *RSC Adv.* **2014**, *4*, 23985–24000.
- [5] A. Paternò, G. Bocci, L. Goracci, S. Scirè, G. Musumarra, *SAR QSAR Environ. Res.* **2016**, *27*, 1–15.
- [6] G. Cruciani, P. Crivori, P. A. Carrupt, B. Testa, *J. Mol. Struct. Theochem.* **2000**, *503*, 17–30.
- [7] VolSurf+ manual and software available at: <http://www.moldiscovery.com/docs/vsplus/>.
- [8] E. Carosati, S. Sciabola, G. Cruciani, *J. Med. Chem.* **2004**, *47*, 5114–5125.
- [9] P. J. Goodford, *J. Med. Chem.* **1985**, *28*, 849–857.
- [10] D. N. A. Boobbyer, P. J. Goodford, P. M. Mcwhinnie, R. C. Wade, *J. Med. Chem.* **1989**, *32*, 1083–1094.
- [11] R. Wade, K. J. Clerk, P. J. Goodford, *J. Med. Chem.* **1993**, *36*, 140–147.
- [12] A. Paternò, F. D'Anna, C. G. Fortuna, G. Musumarra, *Tetrahedron* **2016**, *72*, 3282–3287.
- [13] A. Paternò, G. Bocci, G. Cruciani, C. G. Fortuna, L. Goracci, S. Scirè, G. Musumarra, *SAR QSAR Environ. Res.* **2016**, *27*, 221–244.
- [14] A. Paternò, S. Scirè, G. Musumarra, *Tox. Res.* **2016**, *5*, 1090–1096.
- [15] R. N. Das, T. E. Sintra, J. A. P. Coutinho, S. P. M. Ventura, K. Roy, P. L. A. Popelier, *Tox. Res.* **2016**, *5*, 1388–1399.
- [16] M. Grzonkowska, A. Sosnowska, M. Barycki, A. Rybinska, T. Puzyn, *Chemosphere* **2016**, *159*, 199–207.
- [17] A. Sosnowska, M. Barycki, A. Gajewicz, M. Bobrowski, S. Freza, P. Skurski, S. Uhl, E. Laux, T. Journot, L. Jeandupeux, H. Keppner, T. Puzyn, *ChemPhysChem* **2016**, *17*, 1591–1600.
- [18] A. Paternò, R. Fiorenza, S. Marullo, G. Musumarra, S. Scirè, *RSC Adv.* **2016**, *6*, 36085–36089.
- [19] R. L. Gardas, J. A. P. Coutinho, *Fluid Phase Equilib.* **2008**, *266*, 195–201.
- [20] H. Matsuda, H. Yamamoto, K. Kurihama, K. Tokigi, *Fluid Phase Equilib.* **2007**, *261*, 434–443.
- [21] B.-K. Chen, M.-J. Liang, T.-Y. Wu, H. P. Wang, *Fluid Phase Equilib.* **2013**, *350*, 37–42.
- [22] K. Padaszynski, U. Domańska, *J. Chem. Inf. Model.* **2014**, *54*, 1311–1324.
- [23] R. Sure, J. G. Brandenburg, S. Grimme, *ChemistryOpen* **2016**, *5*, 94–109.
- [24] R. Bini, M. Malvaldi, W. R. Pitner, C. Chiappe, *J. Phys. Org. Chem.* **2008**, *21*, 622–629.
- [25] G. Yu, L. Wen, D. Zhao, C. Asumana, X. Chen, *J. Mol. Liq.* **2013**, *184*, 51–59.
- [26] G. Yu, D. Zhao, L. Wen, S. Yang, X. Chen, *AIChE J.* **2012**, *58*, 2885–2899.
- [27] C. Han, G. Yu, L. Wen, D. Zhao, C. Asumana, X. Chen, *Fluid Phase Equilib.* **2011**, *300*, 95–104.
- [28] R. Alcalde, G. Garcia, M. Atilhan, S. Aparicio, *Ind. Eng. Chem. Res.* **2015**, *54*, 10918–10924.
- [29] Y. Zhao, Y. Huang, X. Zhang, S. Zhang, *Phys. Chem. Chem. Phys.* **2015**, *17*, 3761–3767.
- [30] S. Wold, C. Albano, W. J. III Dunn, U. Edlund, K. Esbensen, P. Geladi, S. Hellberg, E. Johansson, W. Lindberg, M. Sjöström in *Chemometrics, Mathematics and Statistics in Chemistry* (Ed.: B. R. Kowalski), Reidel, Dordrecht, **1984**, pp. 17–94.
- [31] S. Wold, M. Sjöström, L. Eriksson in *The Encyclopedia of Computational Chemistry* (Ed.: P. v. R. Schleyer), John Wiley & Sons, Chichester, **1998**, pp. 2006–2020.

- [32] M. Barycki, A. Sosnowska, M. Piotrowska, P. Urbaszek, A. Rybinska, M. Grzonkowska, T. Puzyn, *J. Cheminf.* **2016**, *8*, 40–54.
- [33] A. Golbraikh, A. Tropsha, *J. Mol. Graph. Model.* **2002**, *20*, 269–276.
- [34] P. Gramatica, *QSAR Comb. Sci.* **2007**, *26*, 694–701.
- [35] V. Consonni, D. Ballabio, R. Todeschini, *J. Chem. Inf. Model.* **2009**, *49*, 1669–1678.
- [36] P. Gramatica, N. Chirico, E. Papa, S. Cassani, S. Kovarich, *J. Comput. Chem.* **2013**, *34*, 2121–2132.
- [37] P. Gramatica, S. Cassani, N. Chirico, *J. Comput. Chem.* **2014**, *35*, 1036–1044.
- [38] K. Roy, S. Kar, R. N. Das, *A Primer on QSAR/QSPR Modeling*, Springer, Heidelberg, **2015**, pp. 37–94.
- [39] SIMCA 13, MKS Umetrics AB, Malmo (Sweden), **2013**.
- [40] M. Marszalek, Z. Fei, D.-R. Zhu, R. Scopelliti, P. J. Dyson, S. M. Zakeeruddin, M. Gratzel, *Inorg. Chem.* **2011**, *50*, 11561–11567.
- [41] Y. Yoshida, K. Muroi, A. Otsuka, G. Saito, M. Takahashi, T. Yoko, *Inorg. Chem.* **2004**, *43*, 1458–1462.
- [42] D. Kuang, P. Wang, S. Ito, S. M. Zakeeruddin, M. Gratzel, *J. Am. Chem. Soc.* **2006**, *128*, 7732–7733.
- [43] J. A. Lazzús, *J. Phys. Org. Chem.* **2009**, *22*, 1193–1197.
- [44] M. Deetlefs, K. R. Seddon, M. Shara, *Phys. Chem. Chem. Phys.* **2006**, *8*, 642–649.
- [45] J. Palomar, V. R. Ferro, J. S. Torrecilla, F. Rodriguez, *Ind. Eng. Chem. Res.* **2007**, *46*, 6041–6048.
- [46] C. Shen, C. Li, X. Li, Y. Lu, Y. Muhammad, *Chem. Eng. Sci.* **2011**, *66*, 2690–2698.
- [47] A. Rybinska, A. Sosnowska, M. Barycki, T. Puzyn, *J. Comput. Aided Mol. Des.* **2016**, *30*, 165–176.
- [48] S. Zhang, X. Lu, Q. Zhou, X. Li, X. Zhang, S. Li, *Ionic Liquids-Physicochemical Properties*, Elsevier, Oxford, **2009**.
- [49] Ionic Liquids Database ILThermo available at: <http://ilthermo.boulder-nist.gov/>; last access June 2016.
- [50] M. T. Clough, C. R. Crick, J. Grasvik, P. A. Hunt, H. Niedermeyer, T. Welton, O. P. Whitaker, *Chem. Sci.* **2015**, *6*, 1101–1114.
- [51] B. González, E. Gómez, A. Domínguez, M. Vilas, E. Tojo, *J. Chem. Eng. Data* **2011**, *56*, 14–20.
- [52] A. Bhattacharjee, P. J. Carvalho, J. A. P. Coutinho, *Fluid Phase Equilibria* **2014**, *375*, 80–88.
- [53] N. Papaiconomou, J. Salminen, J.-M. Lee, J. M. Prausnitz, *J. Chem. Eng. Data* **2007**, *52*, 833–840.
- [54] S.-H. Yeon, K.-S. Kim, S. Choi, H. Lee, H. S. Kim, H. Kim, *Electrochim. Acta* **2005**, *50*, 5399–5407.
- [55] J. Trygg, S. Wold, *J. Chemom.* **2002**, *16*, 119–128.
- [56] J. Trygg, *J. Chemom.* **2002**, *16*, 283–293.
- [57] J. Trygg, S. Wold, *J. Chemom.* **2003**, *17*, 53–64.
- [58] I. Bandrés, R. Alcalde, C. Lafuente, M. Atilhan, S. Aparicio, *J. Phys. Chem. B* **2011**, *115*, 12499–12513.
- [59] M. A. Taige, D. Hilbert, T. J. S. Schubert, *Z. Phys. Chem. (Muenchen Ger.)* **2012**, *226*, 129–139.
- [60] J.-G. Li, Y.-F. Hu, C.-W. Jin, H.-D. Chu, X.-M. Peng, Y.-G. Liang, *J. Solution Chem.* **2010**, *39*, 1877–1887.
- [61] P. Eiden, S. Bulut, T. Kochner, C. Friedrich, T. Schubert, I. Krossing, *J. Phys. Chem. B* **2011**, *115*, 300–309.
- [62] E. S. Sashina, D. A. Kashirskii, G. Janowska, M. Zaborski, *Thermochim. Acta* **2013**, *568*, 185–188.
- [63] S. Wold, A. Ruhe, H. Wold, W. J. Dunn, *SIAM J. Sci. Stat. Comp.* **1984**, *5*, 735–743.

Received: September 29, 2016

Published online on January 9, 2017