Supplementary Information

Ionic Liquid Conductivity Models by Symbolic Regression

Isak Bengtsson*[a] and Patrik Johansson[a,b,c]

Note 1

Both for the small dataset, provided by Solvionic for the study by Nilsson-Hallén[1], and the large dataset collected from the database ILThermo[2,3], the ILs were identified by their IUPAC names. These were parsed to SMILES strings by either matching the name against a manually constructed lookup table or by using the freely available algorithm “Open Parser for Systematic IUPAC Nomenclature” (OPSIN).[4] The names were first compared against the lookup table, and if no match was found, OPSIN was used instead.

Note 2

The molecular descriptors used as variables in the discovered expressions were all related to molecular shape and charge distribution, computed using the open-source toolkit RDKit[5] and the open-source toolbox Open Babel[6]. Several of the descriptors were functions of a molecule´s principal moments of inertia, here called I_i for $i \in \{1,2,3\}$, with the numbers ordering the moments of inertia in terms of magnitude. Below are descriptions of each computed descriptor.

van der Waals volume

In RDKit, the van der Waals volume is estimated using the double cubic lattice method[7]. Each atom is represented as a sphere with an element-specific van der Waals radius taken from an internal table. The molecular van der Waals volume is computed as the total space occupied by the union of these atomic spheres, determined by counting grid points within the spheres across two offset cubic lattices to reduce discretization error.

Radius of gyration (R)

A function of a molecule´s principal moments of inertia and molecular weight M:

$$R = \begin{cases} \sqrt{\sqrt{\frac{I_3 * I_2}{M}}}, & \text{for planar molecules} \\ \sqrt{\frac{2\pi (I_3 I_2 I_1)^{1/3}}{M}}, & \text{for non-planar molecules} \end{cases} . \tag{S1}$$

Inertial shape factor (Z)

A function of a molecule´s principal moments of inertia:

$$Z = \frac{I_2}{I_1 I_3}. \tag{S2}$$

[a] *Department of Physics, Chalmers University of Technology, SE-41296 Göteborg, SWEDEN*
[b] *Department of Chemistry - Ångström, Uppsala University, 751 21 Uppsala, SWEDEN*
[c] *ALISTORE-ERI, FR CNRS 3104, Hub de l’Energie, 80039 Amiens, FRANCE*

Asphericity (A)

A function of a molecule´s principal moments of inertia:

$$A = \frac{\frac{1}{2}\left((I_3 - I_2)^2 + (I_3 - I_1)^2 + \left(I_2 - I_1^2\right)^2\right)}{I_1^2 + I_2^2 + I_3^2}. \tag{S3}$$

Eccentricity (E)

A function of a molecule´s principal moments of inertia:

$$E = \frac{\sqrt{I_3^2 - I_1^2}}{I_3}. \tag{S4}$$

Spherocity index (ζ)

A function of a molecule´s principal moments of inertia, when the moments are calculated without atomic weights:

$$\zeta = 3\frac{I_1}{I_1 + I_2 + I_3}. \tag{S5}$$

Norm of the dipole moment (μ)

The norm of the dipole moment is computed by first assigning partial charges to each atom in the molecule, using the charge equilibration method (Qec) from Open Babel. Then, we compute a dipole vector:

$$\vec{\mu} = \sum_{i}^{N} q_i \vec{r}_i \tag{S6}$$

where q_i is the partial charge of atom i and $\vec{r}_i$ is the position vector to atom i with respect to the molecule´s centre of mass. The norm of the dipole moment is then taken as the Euclidian norm of the dipole vector:

$$\|\vec{\mu}\| = \sqrt{\mu_x^2 + \mu_y^2 + \mu_z^2}. \tag{S7}$$

Gini coefficient (of charge distribution) (G)

The Gini coefficient is used to quantify how uneven the molecular charge distribution of a given ion is. This is done in the following steps. First, a 3D grid is defined around the molecule. Then, each atom contributes to the charge density with a normalized Gaussian centered at its atomic position, where the width of the Gaussian is the atom's van der Waals radius and the amplitude is proportional to the atomic partial charge. The resulting charge distribution is normalized such that the total sum over all grid points equals 1, forming a 3D probability distribution. Flattening and sorting the grid points from lowest to highest density allows the construction of the Lorenz curve, and the Gini coefficient is computed as

$$G = 1 - 2 \times \text{area under Lorenz curve}. \tag{S8}$$

If $G \to 0$ we have an evenly spread charge and if $G \to 1$ the charge is concentrated in small, localized regions.[8,9]

Table S1: Numerical parameters for the best found expression when considering the small dataset.

Parameter	Value
p_0	18.31
p_1	98.26
p_2	0.66
p_3	0.94
p_4	0.14
p_5	0.02
p_6	3.64

Table S2: Numerical parameters for the best found expression when considering the large dataset.

Parameter	Value
p_0	501.97
p_1	33.70
p_2	216.23
p_3	53.63
p_4	1110.55
p_5	1.98
p_6	36.41
p_7	76.21

References

[1] J. Nilsson-Hallén, B. Ahlström, M. Marczewski and P. Johansson, *Front. Chem.*, 2019, **7**, 126.

[2] A. Kazakov, J. Magee, R. Chirico, E. Paulechka, V. Diky, C. Muzny, K. Kroenlein and M. Frenkel, *NIST Standard Reference Database 147: NIST Ionic Liquids Database - (ILThermo).*

[3] Q. Dong, C. D. Muzny, A. Kazakov, V. Diky, J. W. Magee, J. A. Widegren, R. D. Chirico, K. N. Marsh and M. Frenkel, *Journal of Chemical & Engineering Data*, 2007, **52**, 1151–1159.

[4] D. M. Lowe, P. T. Corbett, P. Murray-Rust and R. C. Glen, *J. Chem. Inf. Model.*, 2011, **51**, 739–753.

[5] *RDKit: Open source cheminformatics*, `https://www.rdkit.org`.

[6] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, *J. Cheminf.*, 2011, **3**, 33.

[7] F. Eisenhaber, P. Lijnzaad, P. Argos, C. Sander and M. Scharf, *J. Comput. Chem.*, 1995, **16**, 273–284.

[8] P. M. Dixon, J. Weiner, T. Mitchell-Olds and R. Woodley, *Ecology*, 1987, **68**, 1548–1551.

[9] C. Damgaard and J. Weiner, *Ecology*, 2000, **81**, 1139–1142.