

Computed descriptors for the article: Ionic Liquid Conductivity Models by Symbolic Regression

Isak Bengtsson^{*a} and Patrik Johansson^{a,b,c}

The molecular descriptors in the datasets are related to molecular shape and charge distributions, computed using the open-source toolkit RDKit¹ and the open-source toolbox Open Babel.² Notably, Open Babel was responsible for computing the partial atomic charges using the charge equilibration method (Qeq).³

Several of the descriptors are 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 method⁴. 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.

In the dataset, the van der Waals volume is referred to as "V_rdkit_x".

Radius of gyration (R)

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

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

In the dataset, the radius of gyration is referred to as "RadiusOfGyration_rdkit_x".

Inertial shape factor (Z)

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

$$Z = \frac{I_2}{I_1 I_3}. \quad (\text{S2})$$

In the dataset, the inertial shape factor is referred to as "InertialShapeFactor_rdkit_x".

Asphericity (A)

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

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

In the dataset, the asphericity is referred to as "Asphericity_rdkit_x".

^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

Eccentricity (E)

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

$$E = \frac{\sqrt{I_3^2 - I_1^2}}{I_3}. \quad (\text{S4})$$

In the dataset, the aspericity is referred to as "Eccentricity_rdkit_x".

Sphericity index (ζ)

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

$$\zeta = 3 \frac{I_1}{I_1 + I_2 + I_3}. \quad (\text{S5})$$

In the dataset, the sphericity index is referred to as "SphericityIndex_rdkit_x".

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:

$$\vec{\mu} = \sum_i^N q_i \vec{r}_i \quad (\text{S6})$$

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:

$$\|\vec{\mu}\| = \sqrt{\mu_x^2 + \mu_y^2 + \mu_z^2}. \quad (\text{S7})$$

In the dataset, the norm of the dipole moment is referred to as "dm_norm_qeq_ob_x".

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

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

If $G \rightarrow 0$ we have an evenly spread charge and if $G \rightarrow 1$ the charge is concentrated in small, localized regions.^{5,6} In the dataset, the Gini coefficient is referred to as "G_qeq_ob_x".

References

[1] *RDKit: Open source cheminformatics*, <https://www.rdkit.org>.

- [2] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, *J. Cheminf.*, 2011, **3**, 33.
- [3] A. K. Rappe and W. A. Goddard, *J. Phys. Chem.*, 1991, **95**, 3358–3363.
- [4] F. Eisenhaber, P. Lijnzaad, P. Argos, C. Sander and M. Scharf, *J. Comput. Chem.*, 1995, **16**, 273–284.
- [5] P. M. Dixon, J. Weiner, T. Mitchell-Olds and R. Woodley, *Ecology*, 1987, **68**, 1548–1551.
- [6] C. Damgaard and J. Weiner, *Ecology*, 2000, **81**, 1139–1142.