# SOLVATION ENERGIES VIA MACHINE LEARNING WITH PHYSICALLY SOUND FEATURES

Sergei F. Vyboishchikov

[a] Institut de Química Computacional i Catàlisi and Departament de Química, Universitat de Girona, Spain
vyboishchikov@googlemail.com

We have proposed a family of artificial neural network methods for calculating solvation free energy $\Delta G^{\circ}_{\mathrm{solv}}$ of molecules and ions based on physically sound input features. An early version of the method, *ESE-EE-DNN*,[1,2] uses *COSMO*[3] electrostatic energy, atomic cavity surface areas, total cavity volume, and induced surface charges as input features. For the electrostatic calculation, a specially modified version of electronegativity-equalization atomic charges is employed. A more efficient method, *ESE-GB-DNN*,[4] avoids an explicit cavity construction by using generalized-Born terms, as well as atomic surface areas and the molecular volume. A slightly modified version of *ESE-EE-DNN*, dubbed *ESE-ΔH-DNN*,[5] yields both $\Delta G^{\circ}_{\mathrm{solv}}$ and $\Delta H^{\circ}_{\mathrm{solv}}$ for neutral solutes. The newest method of the ESE-DNN family, *AtomicESE* [6], calculates $\Delta G^{\circ}_{\mathrm{solv}}$ by summing atomic contributions $\Delta G^{\circ}_{\mathrm{solv}}(i)$ for each atom $i$, with each $\Delta G^{\circ}_{\mathrm{solv}}(i)$ evaluated by a dense neural network. This atomic network uses six local atomic features, two global charge-related molecular properties, and five solvent-specific properties. For neutral solutes, AtomicESE achieves an average RMSE for organic solvents below 0.6 kcal/mol, demonstrating strong performance across diverse classes of organic solvents.

To further improve the performance of the solvation-energy schemes, we introduced an atomic-charge scheme, *BoostCha*,[7] based on Gradient Boosting Decision Trees. The *BoostCha* model predicts pseudo-charges for individual atoms from their local environments, represented by 3-D Kocer–Mason–Erturk descriptors. The BoostCha charges are employed as input features in two independent machine-learning models for predicting solvation free energies in organic solvents: *ESE-Boost*, a gradient-boosting model, and *ESE-ANN*,[7] a dense artificial neural network. Both approaches yield robust and consistent predictive performance, with average RMSE of about 0.5 kcal/mol.

**Acknowledgements.** This work has been supported by the Spanish Ministerio de Ciencia, Innovación y Universidades (grant PID2023-147424NB-I00).

## REFERENCES

[1] A. A. Voityuk, S. F. Vyboishchikov, *Phys. Chem. Chem. Phys.* **2020**, *22*, 14591–14598. https://doi.org/10.1039/d0cp02667k

[2] S. F. Vyboishchikov, *J. Chem. Inf. Model.* **2023**, *63*, 6283–6292. https://doi.org/10.1021/acs.jcim.3c00922

[3] A. Klamt, G. Schüürmann, *J. Chem. Soc. Perkin Trans.* **1993**, *2*, 799–805. https://doi.org/10.1039/P29930000799

[4] S. F. Vyboishchikov, *J. Chem. Theory Comput.* **2023**, *19,* 8340–8350. https://doi.org/10.1021/acs.jctc.3c00858

[5] S. F. Vyboishchikov, *Liquids* **2024**, *4,* 525–538. https://doi.org/10.3390/liquids4030030

[6] S. F. Vyboishchikov, *J. Comput. Chem.* **2025**, *46,* e70104. https://doi.org/10.1002/jcc.70104

[7] S. F. Vyboishchikov, *J. Comput. Chem.* **2026**, *47,* e70310. https://doi.org/10.1002/jcc.70310