WLOGPnp: adjusting the atomic contribution model for the computational prediction of log*P* of natural products

Angelos Kollias[1], Antoine Daina[2], Vincent Zoete[1,2]

[1] Computer-aided Molecular Engineering Group, Department of Oncology, Faculty of Biology and Medicine, University of Lausanne, Ch. des Boveresses 155, CH-1066, Lausanne, Switzerland

[2] Molecular Modeling Group, SIB Swiss Institute of Bioinformatics, Quartier Sorge, Bâtiment Génopode, Lausanne, CH-1015, Switzerland

Abstract Text

Lipophilicity plays a crucial role in drug design and discovery[1], and reliable prediction of logP can determine the fate of a bioactive compound in its selection for drug discovery projects. However, this physicochemical property differs significantly between synthetic compounds and natural products[2], as it relies heavily on their structural characteristics. To date, no method has specifically focused on logP predictions for the latter. Computational methods that partially depend on the atomic contribution model[3–5] could therefore produce more reliable predictions for natural products when appropriately designed. In this study, the Natural product-likeness score developed by Ertl et al.[6] was leveraged to mitigate the data scarcity in the chemical space of natural compounds and the limited availability of reliable experimental logP measurements. Investigating the selected compounds using sliding windows of this 'NP-Score' continuum rather than applying a strict natural/non-natural products distinction allowed for the adjustment of the previous atomic contribution model established by Wildman and Crippen[7] with minimal refinements. The training of the multilinear regression was enhanced by introducing a regularization factor to reduce the negative effects of the model's inherent collinearity features. The fine-tuning of this hyperparameter through cross validation and the expansion of the training set to include more data points of natural products' logP measurements delivered a ridge regression model tailored for natural products. One of our goals is to implement this new model in the freely accessible tools of the SwissDrugDesign suite for computer-aided drug design and advance the efforts of the scientific community in natural products-based drug discovery.

Bibliography

[1] B. Testa ; P. Crivori ; M. Reist ; P.-A. Carrupt. KLUWER/ESCOM (2000).

[2] Y. Chen ; J. Kirchmair. *Mol. Inform.* 39 (2020) e2000171.

[3] A. Daina ; V. Zoete. *ChemMedChem* 11 (2016) 1117–1121.

[4] A. Daina ; O. Michielin ; V. Zoete. *Nucleic Acids Res.* 47 (2019) W357–W364.

[5] A. Daina ; O. Michielin ; V. Zoete. *Sci. Rep.* 7 (2017) 42717.

[6] P. Ertl ; S. Roggo ; A. Schuffenhauer. *J. Chem. Inf. Model.* 48 (2008) 68–74.

[7] S.A. Wildman ; G.M. Crippen. *J. Chem. Inf. Comput. Sci.* 39 (1999) 868–873.