Journal of Chemical Information and Modeling 2014 vol.54 N12, pages 3320-3329

How accurately can we predict the melting points of drug-like compounds?

Tetko I., Sushko Y., Novotarskyi S., Patiny L., Kondratov I., Petrenko A., Charochkina L., Asiri A.
Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© 2014 American Chemical Society. This article contributes a highly accurate model for predicting the melting points (MPs) of medicinal chemistry compounds. The model was developed using the largest published data set, comprising more than 47k compounds. The distributions of MPs in drug-like and drug lead sets showed that >90% of molecules melt within [50,250]°C. The final model calculated an RMSE of less than 33 °C for molecules from this temperature interval, which is the most important for medicinal chemistry users. This performance was achieved using a consensus model that performed calculations to a significantly higher accuracy than the individual models. We found that compounds with reactive and unstable groups were overrepresented among outlying compounds. These compounds could decompose during storage or measurement, thus introducing experimental errors. While filtering the data by removing outliers generally increased the accuracy of individual models, it did not significantly affect the results of the consensus models. Three analyzed distance to models did not allow us to flag molecules, which had MP values fell outside the applicability domain of the model. We believe that this negative result and the public availability of data from this article will encourage future studies to develop better approaches to define the applicability domain of models. The final model, MP data, and identified reactive groups are available online at http://ochem.eu/article/55638.

http://dx.doi.org/10.1021/ci5005288