PREDICTION OF pH-DEPENDENT AQUEOUS SOLUBILITY OF DRUG-LIKE MOLECULES

Niclas Tue Hansen, Irene Kouskoumvekaki, Flemming Steen Jørgensen[+], Søren Brunak and Svava Ósk Jónsdóttir

Center for Biological Sequence Analysis, BioCentrum, Technical University of Denmark, DK-2800 Lyngby, Denmark

[+]Department of Medicinal Chemistry, The Danish University of Pharmaceutical Sciences, Universitetsparken 2, DK-2100 Copenhagen, Denmark

The oral bioavailability of a drug is a highly complicated property that depends on the drug's solubility in the gastrointestinal tract and its absorption into the blood stream, among a variety of other factors. Several models have been developed for the prediction of the solubility of non-ionised compounds (intrinsic solubility). However, the solubility of ionisable molecules varies considerably with the pH, and as the pH takes different values throughout the gastrointestinal tract, a reliable model for prediction of pH-dependent solubility would significantly assist in improving the modelling of oral bioavailability of drugs.

In this work, the Henderson-Hasselbalch equation has been employed, which describes pH-dependent solubility as a function of the intrinsic solubility (S_o) and the acid / base dissociation coefficients (pK_a / pK_b). A predictive model for the intrinsic solubility has been developed based on artificial neural networks (ANN) that have been trained on a unique set of 378 drugs and drug-like molecules from the literature, as well as a drug-like PhysProp dataset of 4548 compounds (filtered from the original dataset of 41.040 compounds). For the prediction of the acid / base dissociation coefficients, a commercial tool has been used (MARVIN) after being validated on a dataset of 467 molecules from the PhysProp database. The best performing ANN has an RMSE of 0.61 logS-units using three-part validation, while the MARVIN module has an RMSE of 0.71 pH-units.

Furthermore, a dataset of 27 drugs with experimentally determined solubility curves has been assembled from the literature for the validation of the combined model. The resulting average *RMSE* is 0.81 log *S*-units, which suggests that pH-dependent solubility can be modelled with almost the same accuracy as the intrinsic solubility and shows that the combined model that we have developed in this work is a reliable predictive tool for the pH-dependent solubility for drug and drug-like compounds.