Big Data meets Physics

Combining data mining using thousands of solubilities with a physical model enables better solubility predictions.

Robust solubility predictions in a large variety of solvents are desirable at early stage of drug development for numerous purposes, e.g. manufacturing process development or purification. Current models that are based on structure-related properties, group contribution methods or heuristic rules become increasingly error-prone when facing complex large molecules that are currently in the development pipelines. Big-data models are usually not applicable to early stages of drug development as they require too much unavailable data.

Training a powerful physical model that considers fundamental physical basics of intermolecular interactions (hydrogen bonding, van-der-Waals, …) with a huge dataset enables deriving a unique interaction landscape for chemically complex molecules. In total 2000 solubility values (140 drug molecules in 60 solvents) enabled deriving a robust and predictive model for all commonly used solvents. Figure 1 depicts the modelling workflow for this approach, the model quality is shown in the parity plot in Figure 2.

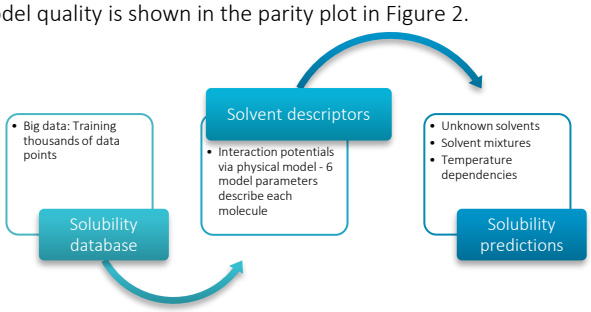

Figure 1. Modelling workflow for the data-based modelling approach

Our approach only requires few key experiments to characterize the individual interaction profile for a new compound, the chemical structure itself remains disclosed. Although the initial training was performed with thousands of solubility points, the combination with the physical model allows establishing the model for new molecules with only few data points.

Benefits

- Reliable solubility estimates - validated by 2000 solubility points in 60 organic solvents and water
- Applicable to complex drug structures
- Molecular structure of the drug molecule not required
- Benefit of physical model: Only five experimental solubilities to set up the model for a new drug molecule. The underlying data groundwork does not change.
- Seamless integration of the model in further formulation development pipeline, e.g., polymer selection for amorphous solid dispersions

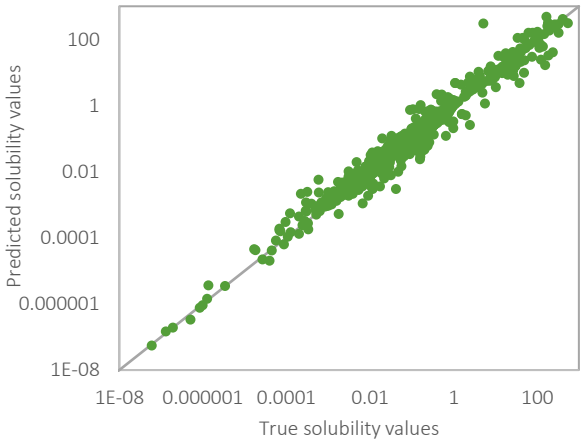

Figure 2. Comparison of predicted and true solubility data for 2000 solubility values (different concentration units and temperatures).

Contact us to find out more

amofor GmbH | Otto-Hahn-Str. 15 | 44227 Dortmund | Germany
www.amofor.de | info@amofor.de

amofor
enabling formulations