and eight additional descriptors. The dataset was divided into a training and a test set on the basis of Kohonen's self-organizing neural network. Good prediction results were obtained for back-propagation neural network models: with 18 topological descriptors, for the 936 compounds in the test set, a correlation coefficient of 0.92 and a SD of 0.62 were achieved; with three-dimensional descriptors, for the 866 compounds in the test set, a correlation coefficient of 0.90 and a SD of 0.73 were achieved. The models were also tested by using another dataset, and the relationship of the two datasets was examined by Kohonen's self-organizing neural network.

In recent years, efforts were made to improve simulation efficiency and make use of readily available open-source software for solubility prediction. Sellers et al. (2016) extended various free-energy methodologies to determine the chemical potential of the solid and liquid phases of a fully flexible molecule using classical simulation. They outlined an efficient technique to find the absolute chemical potential and melting point of a fully flexible molecule using one set of simulations to compute the solid absolute chemical potential and one set of simulations to compute the solid–liquid free energy difference. With this combination, only a handful of simulations are needed, whereby the absolute quantities of the chemical potentials are obtained for use in other property calculations, such as the characterization of crystal polymorphs or the determination of the entropy. Using the LAMMPS molecular simulator, the Frenkel and Ladd and pseudo-supercritical path techniques are adapted to generate 3rd order fits of the solid and liquid chemical potentials. Results yield the thermodynamic melting point $T_m = 488.75$ K at 1.0 atm.

Conceptually, the simplest way to compute solubilities from simulation is to carry out *brute-force* direct coexistence simulations (Espinosa et al., 2016; Kolafa, 2016). While this method has the advantage of simplicity, it generally requires long simulations (in some cases up to microseconds) to achieve solubility equilibrium, even for highly soluble compounds. In order to develop a general method but make use of only readily available open-source software, Li et al. (2017) developed a methodology for universal solubility prediction. They developed a numerical method that enables convenient solubility estimation of general molecular crystals at arbitrary thermodynamic conditions where solid and solution can coexist. The methodology is based on standard alchemical free energy methods, such as thermodynamic integration and free energy perturbation, and consists of two parts: (1) systematic extension of the Einstein crystal method to calculate the absolute solid free energies of molecular crystals at arbitrary temperatures and pressures, and (2) a flexible cavity method that can yield accurate estimates of the excess solvation free energies. The results show that via classical Molecular Dynamic simulations, their approach can predict the solubility of OPLS-AA-based (Optimized Potentials for Liquid Simulations All Atomic) naphthalene in SPC (Simple Point Charge) water in good agreement with experimental data at various temperatures and pressures.

The Modified Separation of Cohesive Energy Density Model (MOSCED) is an efficient, analytic method to predict infinite dilution activity coefficients over a range of temperatures. Its predictability makes MOSCED an attractive engineering design tool. However, its use is limited. When trying to model a novel compound, reference data must first be available to regress the necessary MOSCED parameters. Here, Ley et al. (2016) proposed the use of molecular simulation to generate the reference dataset. In this fashion, MOSCED can be made a truly predictive engineering design tool. This combines the predictive strength of molecular simulation with the efficiency of MOSCED to create a powerful new tool. By adopting the melting point temperature and enthalpy of fusion of these compounds from available experimental data, they were able to predict equilibrium solubilities. Predictions using the new predictive MOSCED are in good agreement with available experimental solubility data for acetaminophen in non-aqueous solvents, as shown in Figure 3.5.

Cox et al. (2017) demonstrated for the solutes methylparaben, ethylparaben, propylparaben, butylparaben, lidocaine, and ephedrine how conventional molecular simulation free energy calculations or electronic structure calculations in a continuum solvent can instead be used to generate the necessary reference data, resulting in a predictive flavor of MOSCED. Adopting the melting point temperature and enthalpy of fusion of these compounds from experiment, they found the method is able to well