In silico Models of Solubility in Simulated Gastrointestinal Fluids (FaSSIF, FeSSIF, and FaSSGF) for a diverse set of 160 molecules

Michael B. Bolger1, John Crison1, Robert Fraczkiewicz1, Marvin Waldman1, Walter S. Woltoasz1, Ece Gamsiz2, Mukul Ashtikar2, and Rebecca Carrier2
1Simulations Plus, Inc. Lancaster, California, 2Northeastern University, 457 Snell Engineering Center, Boston, MA 02115, USA

Purpose
To build in silico models based on our chemically diverse database of solubilities for 160 drugs and drug-like molecules in simulated gastrointestinal fluids for the stomach, fasted state small intestine, and fed state small intestine. Such models should have greater relevance to in vivo conditions used in gastrointestinal simulations.

Methods
A diverse set of 160 drugs and drug-like compounds were obtained from commercial sources. Most were in their free base or free acid form. Equilibrium solubilities were measured in fasted state simulated intestinal fluid (FaSSIF-V2, pH=6.5), fed state simulated intestinal fluid (FeSSIF-V2, pH=6.5), and fed state simulated gastric fluid (FaSSGF-V2, pH=1.6) [Jantratid, 2008]. Detailed methods are reported in another abstract at this meeting [R. Carrier, T2001, AAPS 2009]. We compared the accuracy of literature-based equations for the influence of octanol/water partition coefficient (logP) on the change in equilibrium water solubility due to the addition of sodium taurocholate [S. Mithani, 1996], to preliminary in silico models built using the model-building functions of ADMET Predictor™ (ver. 4.0.0007 Simulations Plus, Inc.).

Results
The original Mithani equation for the log of solubilization ratio (log SR = 2.234 + 0.606 logP) provided reasonably good estimates of solubility in biorelevant media (RMSE = 0.57, r² = 0.79). Addition of descriptors for number of hydrogen bond donors and acceptors to a multiple linear regression resulted in an improved model of log S SR (RMSE = 0.51, r² = 0.80) (graph not shown). Application of the model-building methods in ADMET Predictor to 157 molecules (with 3 outliers removed) resulted in artificial neural network ensemble (ANNE) models with additional improvements in performance (average RMSE = 0.41, r² = 0.82).

Conclusions
A rich and chemically diverse database of solubilities for drugs and drug-like molecules can provide a good training set for in silico estimates of solubility in biorelevant media.

References: