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

Michael B. Bolger[1], John Crison[1], Robert Fraczkiewicz[1], Marvin Waldman[1], Walter S. Woltosz[1],
Ece Gamsiz[2], Mukul Ashtikar[2], and Rebecca Carrier[2]
[1]Simulations Plus, Inc. Lancaster, California, [2]Northeastern 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 of160 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=5.8) and fasted 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, r^2 = 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_{FaSSIF} (RMSE = 0.51, r^2 = 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, r^2 = 0.82).

Conclusions

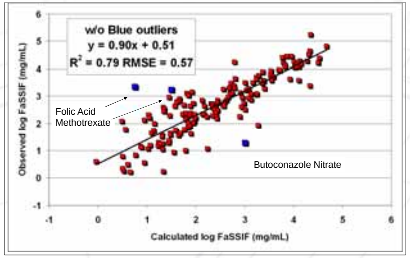
A rich and chemically diverse database of solubilities for drugs and drug-like molecules can provide a good training set for *in silico* estimates of solubility in biorelevant media.

References:

Mithani S.D., et al. Pharm. Res. 1996, 13(1):163

Jantratid E., et al., Pharm. Res. 2008, 25(7):1663

FaSSIF



Calculated vs. Observed log FaSSIF for 160 molecules using Mithani equation and S+logP. (Mithani, 1996). Blue points were not included in regression line.

Mithani developed a relationship between aqueous solubility and the octanol /water partition coefficient that explains the effect of bile salts on solubility.

$$S_B = S_W + SC_{aq} \times SR \times MWt \times [TauCh]$$

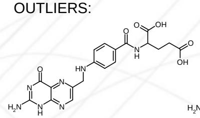
Where: S_B = Solubility in bile salt (Na^+ taurocholate), S_w=Solubility in water

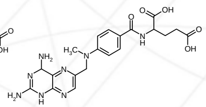
SC_{aq} = Solubilization capacity of water (moles Drug/moles Water)

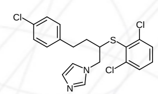
SR = Solubilization ratio = $10^{(2.23+0.61*\log P)}$, *MWt* = molecular weight

[TauCh] = Concentration of Na^+ taurocholate (3 mM for FaSSIF).

OUTLIERS:



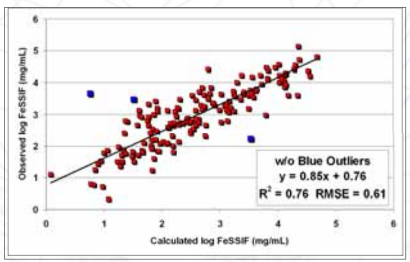




Folic Acid: acid pKa = 3.6 Methotrexate acid pKa = 4.6 Butoconazole base pKa = 5.8

Note: Both folic acid and methotrexate would carry -2 charge at pH 6.8.

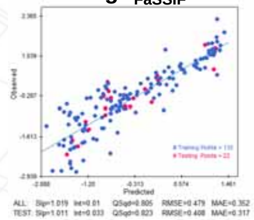
FeSSIF



Calculated vs. Observed $\log S_{FeSSIF}$ for 160 molecules using original Mithani equation and S+logP. (Mithani, 1996). Blue points were not included in regression line.

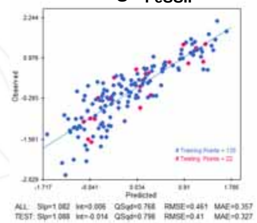
Here we used the ADMET Modeler™ module within ADMET Predictor to build artificial neural network ensemble (ANNE) models of solubility in FaSSIF, FeSSIF, and FaSSGF. Experimental aqueous solubility was included as a possible descriptor. A genetic algorithm using kernel partial least squares (GA-KPLS) was used for selection of descriptors from 325 possible descriptors. Molecules were separated into Training, Verification, and external Test sets using a Kohonen self-organizing map (SOM). 157 molecules were included in the model-building process.

$\log S_{FaSSIF}$



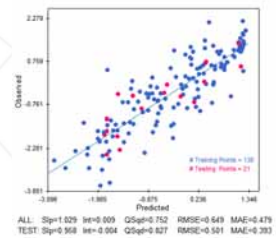
ANNE model of $\log S_{FaSSIF}$ using 2 inputs and 2 hidden nodes. Inputs were aqueous solubility and S+logD (ADMET Predictor)

$\log S_{FeSSIF}$



ANNE model of $\log S_{FaSSIF}$ using 6 inputs and 4 hidden nodes. Inputs were aqueous solubility, S+logD, Pi_Q3, EEM_AFnp, Pi_AQc, M_RNG (ADMET Predictor)

$\log S_{FaSSGF}$



ANNE model of $\log S_{FaSSGF}$ using 4 inputs and 2 hidden nodes. Inputs were Aqueous Solubility, S+logD, N_IoBaAt, F_AFRBWF, (ADMET Predictor)

Table of Model Statistics

Model	Architecture	R^2	RMSE (log units)
Mithani FaSSIF	Linear S+logP	0.79	0.57
Mithani FeSSIF	Linear S+logP	0.76	0.61
S+FaSSIF	ANNE 2I 2N	0.82	0.41
S+FeSSIF	ANNE 6I 4N	0.80	0.41
S+FaSSGF	ANNE 4I 2N	0.83	0.50

(ADMET Predictor) descriptors:
Pi_Q3 = *Third component of the autocorrelation vector of Hückel pi* ***atomic*** *charges.*
EEM_AFnp = *Sum of absolute values of sigma Fukui indices on nonpolar atoms*
Pi_AQc = *Sum of absolute values of Hückel pi atomic charges on carbon atoms.*
M_RNG = *Indicator variable for the presence of ring structures except benzene and its condensed rings (aromatic, heteroaromatic, and hydrocarbon rings.*
N_IoBaAt = *# of recognized ionizable groups that are basic.*
F_AFRBWF = *Average value of the freely rotatable bond weight factor. The FRBW factor is defined as the minimal atom count fraction per bond.*

simulations plus, inc.