White Paper

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

Looking beyond a few hard cut-offs in assessing ADMET Risk™

Dr. Robert D. Clark

Summarized by Wendy Warr, Ph.D.

42505 10th Street West • Lancaster, CA 93534 • USA • phone: +1-661-723-7723 • fax: +1-661-723-5524
email: info@simulations-plus.com • website: www.simulations-plus.com • NASDAQ: SLP

Looking beyond a few hard cut-offs in assessing ADMET risk

Robert D. Clark, bob@simulations-plus.com, Robert Fraczkiewicz, Michael B. Bolger, Walter S. Woltosz. Cheminformatics Sciences, Simulations Plus, Inc., Lancaster, CA 93534, United States

Drug discovery and development is a multi-objective problem. Unfortunately, it is easy to focus exclusively on *in vitro* activity against the therapeutic target and not consider absorption, distribution, metabolism, excretion, and toxicity (ADMET) until late in a development programme. The predictable result is late-stage failures.

The original Rule of Five (Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **2001**, *46* (1-3), 3–26) is widely considered to be an important development in modern drug discovery. It takes on numeric values from 0 to 4 depending on how many "potential problems" a compound might have with its absorption or permeation properties. It is a useful computational filter in drug candidate screening. In terms of Simulations Plus' ADMET Predictor (http://www.simulations-plus.com/Products.aspx?grpID=1&cID=10&pID=13) descriptors and models, the Rule of Five model rules can be formulated with the following set of conditions:

- MlogP > 4.15 (excessive lipophilicity)
- MWt > 500 (large size)
- HBDH > 5 (too many potential hydrogen bond donors)
- M_NO > 10 (too many potential hydrogen bond acceptors)

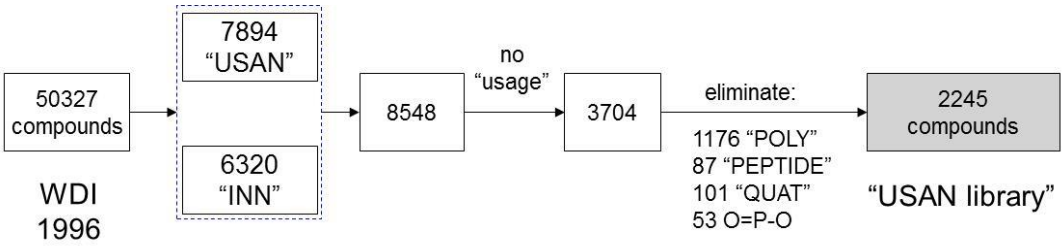
Most commercial drugs suitable for oral dosing violate no more than one of the rules these conditions represent. Researchers at Simulations Plus have used a newer, more refined reference set of drugs from the World Drug Index to develop an extended set of rules that go beyond potential absorption problems to include many other kinds of predictable ADMET liabilities. The focus is on avoiding "bad" compounds rather than suggesting "good" compounds to make.

Simulations Plus' ADMET Risk (http://www.simulations-plus.com/Definitions.aspx?lID=79&pID=13) has a flexible scoring system that encompasses many components of risk: absorption-related properties, pharmacokinetic properties, metabolism and toxicity. It includes a customisable set of rules (24 by default) based on thresholds for ADMET Predictor property value estimates for a reference set drawn from the World Drug Index (WDI). The thresholds are set to pass most commercial orally delivered drugs. ADMET Risk is an easy way for chemists to rapidly "see" in many dimensions by way of a single number, which is augmented by two-letter codes to identify which rules are broken. It is a rapid way for chemists to compare and filter very large numbers of candidate molecules early in design and discovery.

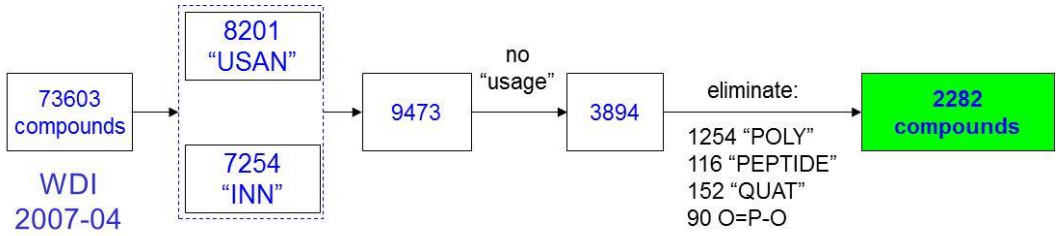
Rule of Five addresses only a narrow slice of the full gamut of hurdles a compound must pass before it can become a drug. In addition, it relies on "hard" thresholds: a compound with a molecular weight of 499 satisfies the molecular weight rule but a compound with a molecular weight of 501 violates it. Simulations Plus calculated a broad range of relevant molecular descriptors and ADMET property predictions for the focused subset of WDI and identified "soft" threshold ranges for each, along the lines suggested by Petit, J.; Meurice, N.; Kaiser, C.; Maggiora, G. Softening the Rule of Five -

where to draw the line? *Bioorg. Med Chem.* **2012**; *20*, 5343-5351, such that approximately 85% of the compounds in the dataset satisfy them completely and somewhat less than 10% violate them completely. The former contribute nothing to the overall risk, whereas the latter contribute the full amount (weight) specified for the corresponding rule. Predictions falling in the grey area in between contribute fractional amounts to the risk score.

The procedure described by Lipinski *et al.* was as follows:



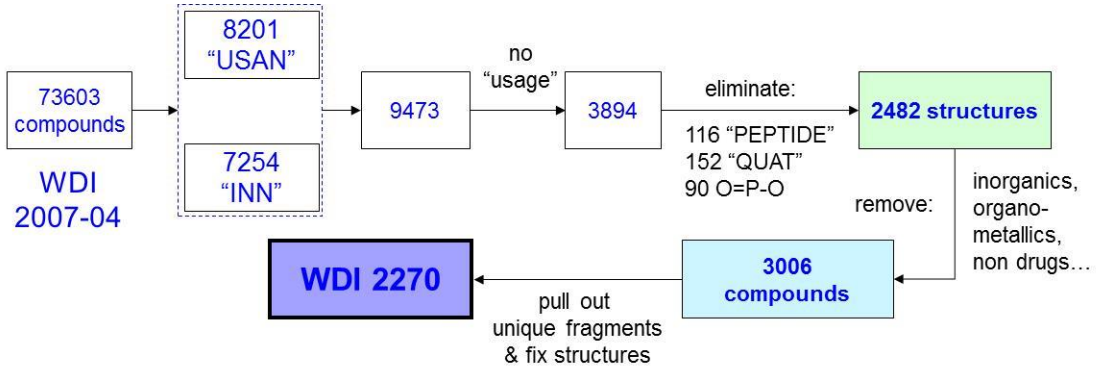
The initial procedure followed by Simulations Plus was as follows:



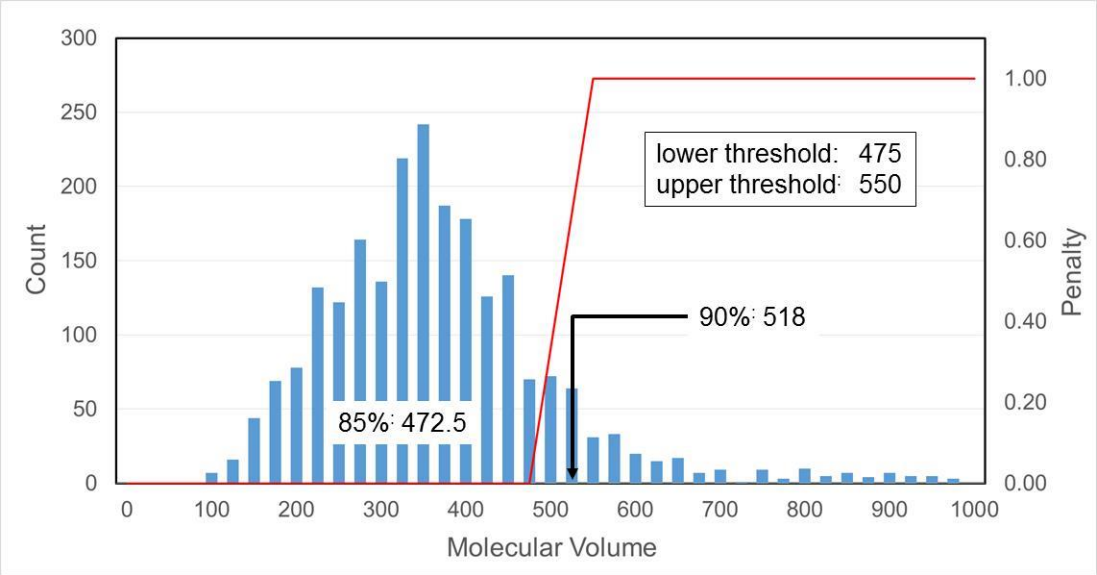
WDI annotation has evidently changed in the last 10 years, and the query "like %POLY%" now hits too many compounds, including polysaccharides, polyamines, poly-C-acids, polyphenols, polyalcohols, polyolefins, polyketones, and others. Some of these are drugs. Azithromycin is a "polyalcohol"; Aclarubcin is a "polyphenol; Amlodipine is a "poly-C-acid"; and Amikacin is a "polyamine". Polyamines, poly-C-acids, polyphenols, polyalcohols, polyolefins, polyketones, and others cannot be dropped if the dataset is to have about 10% of compounds with molecular weight greater than 500. Therefore Simulations Plus eliminated only polysaccharides and polypeptides. Other structure and activity classes eliminated were compounds with no carbon-carbon bond, compounds with a phosphorus-sulphur bond, insecticides, emollients, acidifiers, rubefacients, sweeteners, and cytostatic chemical mustards that contain the substructure N[CH][CH][Cl,Br]and have a molecular weight less than 310.

Also removed were flavourings, sweeteners, sedatives (e.g., bromoform), imaging agents, preservatives, laxatives, general anaesthetics, (e.g., 1,2-dichloro-1,1,2,2-tetrafluoroethane, cyclopropane), antioxidants (e.g., glycerol), escharotics (e.g., potassium nitrate), counterions from salts, antacids, antiseptics, solvents, dietary supplements (e.g., arginine), chelating agents, repellents, lubricants, dyes, emulsifiers, propellants, keratolytics (e.g., formaldehyde), protein conjugates, and any compound that did not have an associated United States Adopted Name (USAN) or International Non-proprietary Name (INN) identifier.

Simulations Plus' final procedure was as follows:



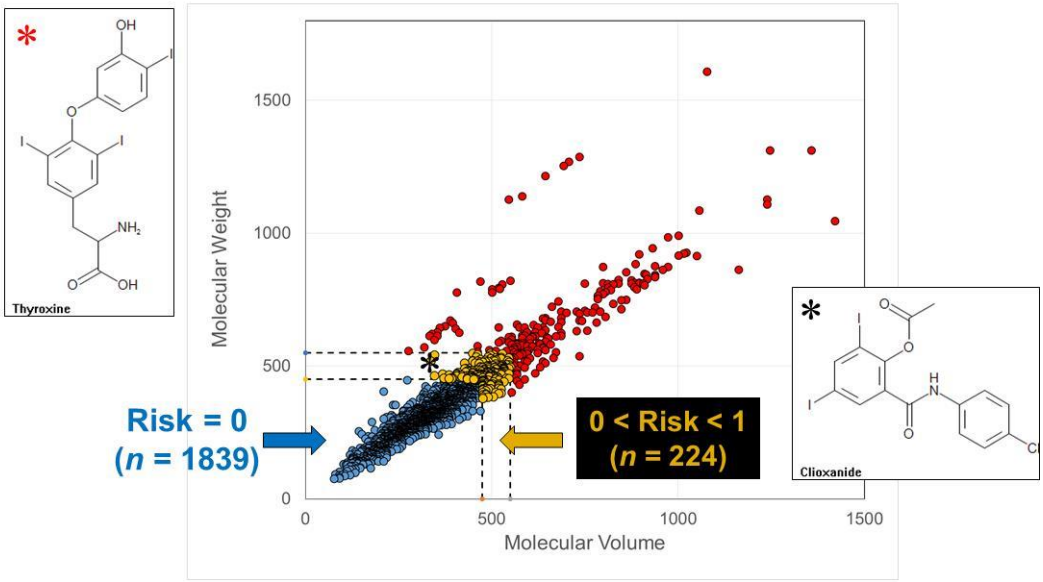
"Soft" threshold ranges are set such that approximately 85% of the compounds in the data set satisfy a property threshold completely and somewhat less than 10% violate them completely. Clark illustrated this by the penalties for molecular volume violation:



Criteria are combined to reduce risk inflation when different criteria tend to track each other. For example molecular weight, molecular volume, number of heavy atoms and number of bonds are all measures of molecular size (Sz). Combination keeps the risk codes as detailed as necessary, but no more complex. Combination is based on fuzzy logic rules developed by Zadeh (Zadeh, L. A. Fuzzy sets. *Information and Control* **1965**, *8*, 338-353; http://en.wikipedia.org/wiki/Fuzzy_set_operations):

- risk(A or B) = max(risk(A), risk(B))
- risk(A and B) = min(risk(A), risk(B))

A weight assigned to the rule is applied to the overall risk. Clark presented a simple example:

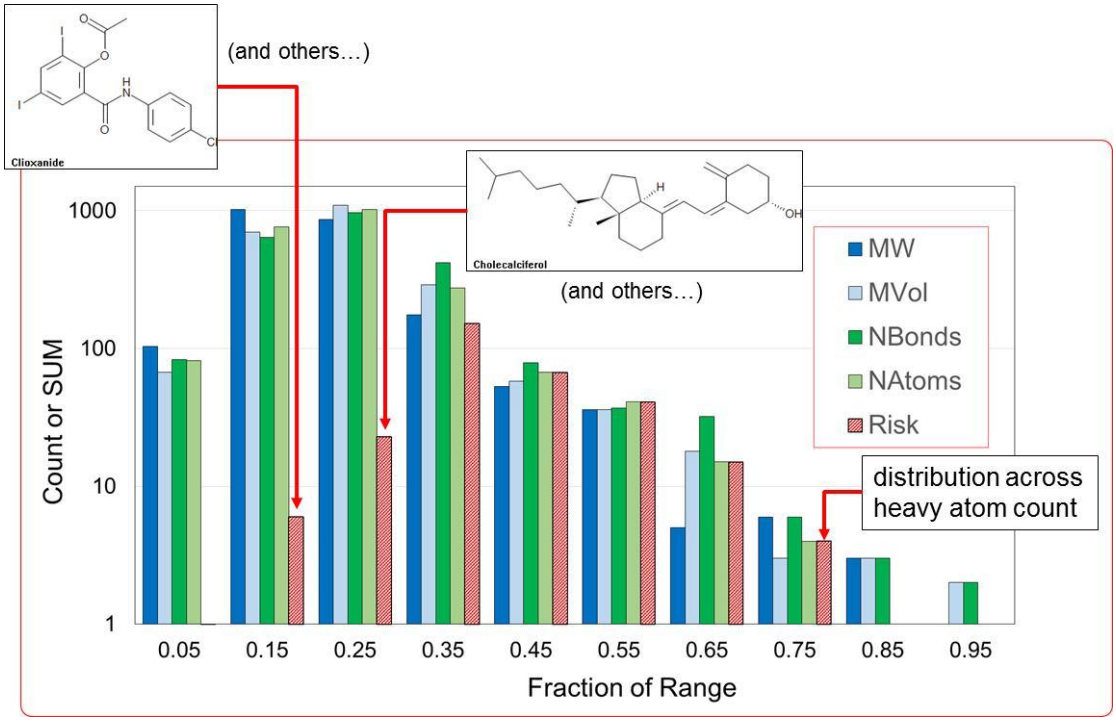


Rule: MWt > [450,550] **OR** MolVol > [475,550]

Thus highly correlated criteria are combined into single rules using Boolean operators. The rule for identifying overly large structures, for example, is:

size (Sz): MWt > [450,550] OR N_Atoms > [32,38] OR MolVol > [475,550] OR N_Bonds > [35,41]

The values within the brackets indicate the boundaries of threshold regions. The Sz rule includes four individual criteria, all of which use the ">" relational operator. Predictions falling below both threshold values contribute nothing to the risk, whereas predictions falling above both contribute 1 violation "point". Intermediate values represent intermediate risks: a compound of molecular weight 500 violates the first criterion and so would represent an incremental risk of 0.5 points for that criterion. Points are then combined across logical operators such as ORs and ANDs using Zadeh's method: criteria connected by the OR operator yield the maximum of the component violation points and those combined using the AND operator yield the minimum. The combined points from the criteria making up a rule then yield an overall value between 0 and 1, which is multiplied by the weight assigned to the rule as a whole. For the size rule the result is as follows.



The 24 default ADMET Risk rules, in three categories, are as follows.

Absorption and PK risks:

	Code	Criteria
S+Absn_Risk	Sz	MWt > [450,550] OR N_Atoms > [32,38] OR MolVol > [475,550] OR N_Bonds > [35,41] (too big)
	RB	N_FrRotB > [9,11] (too many rotatable bonds)
	HD	HBDH > [4,5] AND HBDch > [1.6,2.0] (too many strong H-bond donors)
	HA	HBA > [8,10] AND HBAch < [-5.8,-4.8] (too many strong H-bond acceptors)
	ch	NPA_ABSQ > [19,21] OR T_PSA > [120,155] (too polar or charged)
	ow	S+logP > [4.5,5.5] OR S+logD > [3.5,4.5] OR MlogP > [3.7,4.3] (too lipophilic)
	Pf	S+Peff < [0.25,0.5] OR MDCK < [20,30] (low permeability)
	Sw	S+Sw < [0.003,0.010] (low water solubility)
	fu	S+PrUnbnd < [1,3] (low percentage unbound in plasma)
	Vd	S+Vd > [3.7,5.7] L/kg (high steady-state volume of distribution)

Toxicity risks:

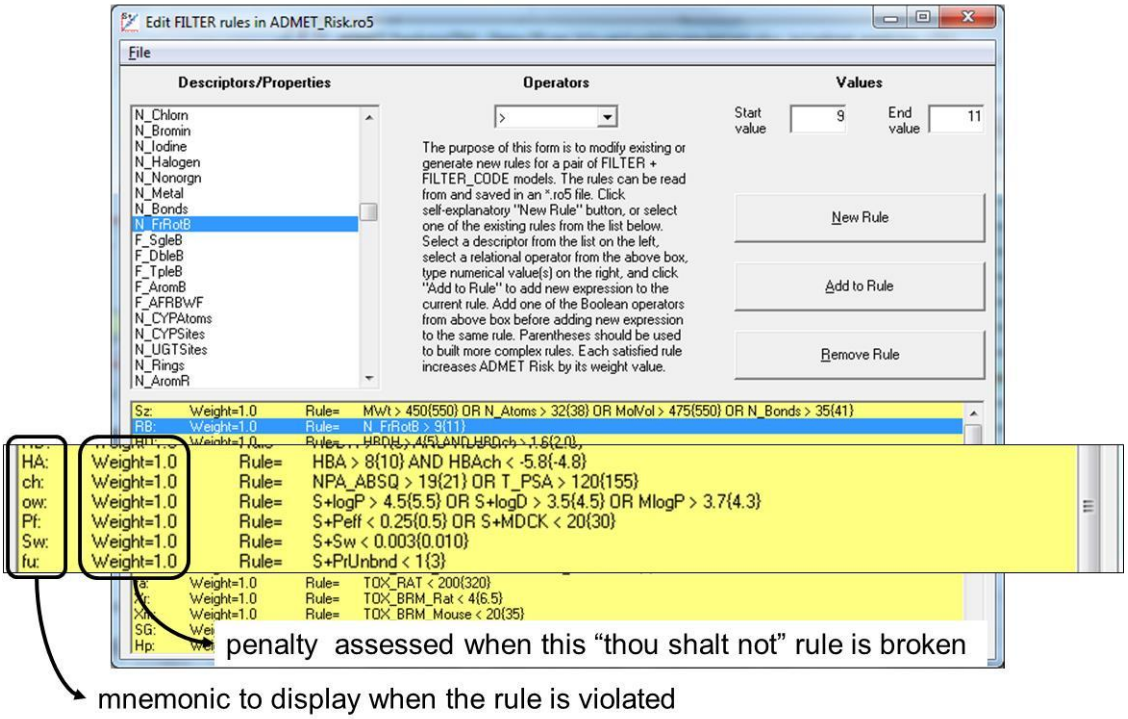
Code	Criteria
hE	TOX_hERG_FILTER = Yes AND TOX_hERG > [5.5,6] pIC50 (potential hERG liability)
ra	TOX_RAT < [200,320] (acute toxicity in rats)

Xr	TOX_BRM_Rat < [4,6.5] (carcinogenicity in chronic rat studies)
Xm	TOX_BRM_Mouse < [20,35] (carcinogenicity in chronic mouse studies)
Hp	(TOX_AlkPhos = Toxic OR TOX_GGT = Toxic OR TOX_LDH = Toxic) AND (TOX_SGOT = Toxic OR TOX_SGPT = Toxic) (hepatotoxicity)
SG	TOX_SGOT = Toxic AND TOX_SGPT = Toxic (elevated liver enzymes)
Mu	TOX_MUT_Risk > 1 (mutagenicity – *in silico* Ames test)

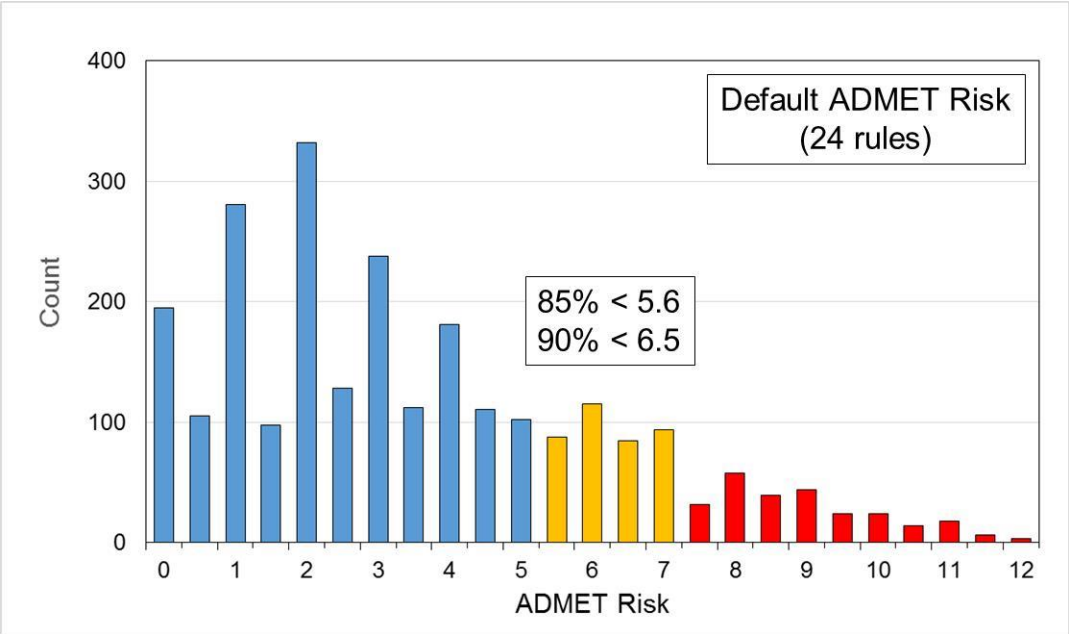
CYP Risks:

Code	**Criteria**
1A	MET_1A2_CLint > [15,30] μL/min/mg (excessive CYP 1A2 clearance)
C9	MET_2C9_CLint > [15,30] μL/min/mg (excessive CYP 2C9 clearance)
19	MET_2C19_CLint > [15,30] μL/min/mg (excessive CYP 2C19 clearance)
D6	MET_2D6_CLint > [15,30] μL/min/mg (excessive CYP 2D6 clearance)
3A	MET_3A4_CLint > [15,30] μL/min/mg AND MET_3A4_HLM_CLint > [15,30] μL/min/mg (excessive CYP 3A4 clearance)
mi	MET_3A4_Ki_mid < 1.5 μM AND (MET_3A4_I_mid=Yes OR MET_3A4_Inh=Yes)
ti	MET_3A4_Ki_tes < 1.0 μM AND (MET_3A4_I_tes = Yes OR MET_3A4_Inh = Yes)

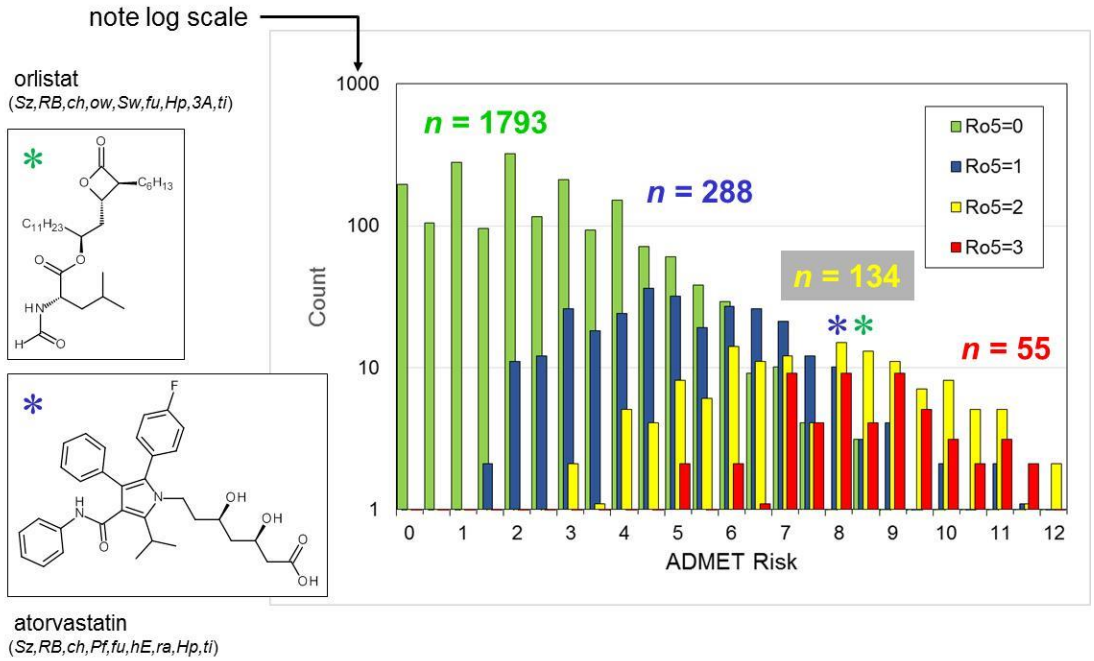
Clark showed how ADMET risk is defined in ADMET Predictor:



About 90% of the reference set has an ADMET Risk score below 6.5. Clark displayed a chart of distribution of risk across the WDI reference set:

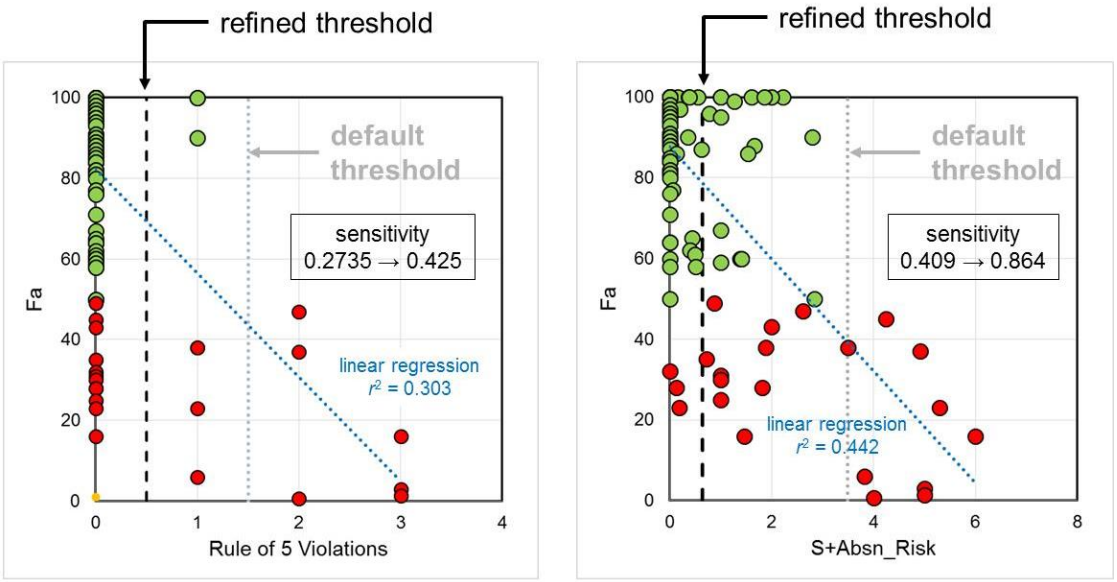


ADMET Risk can provide essential details that Rule of Five does not:

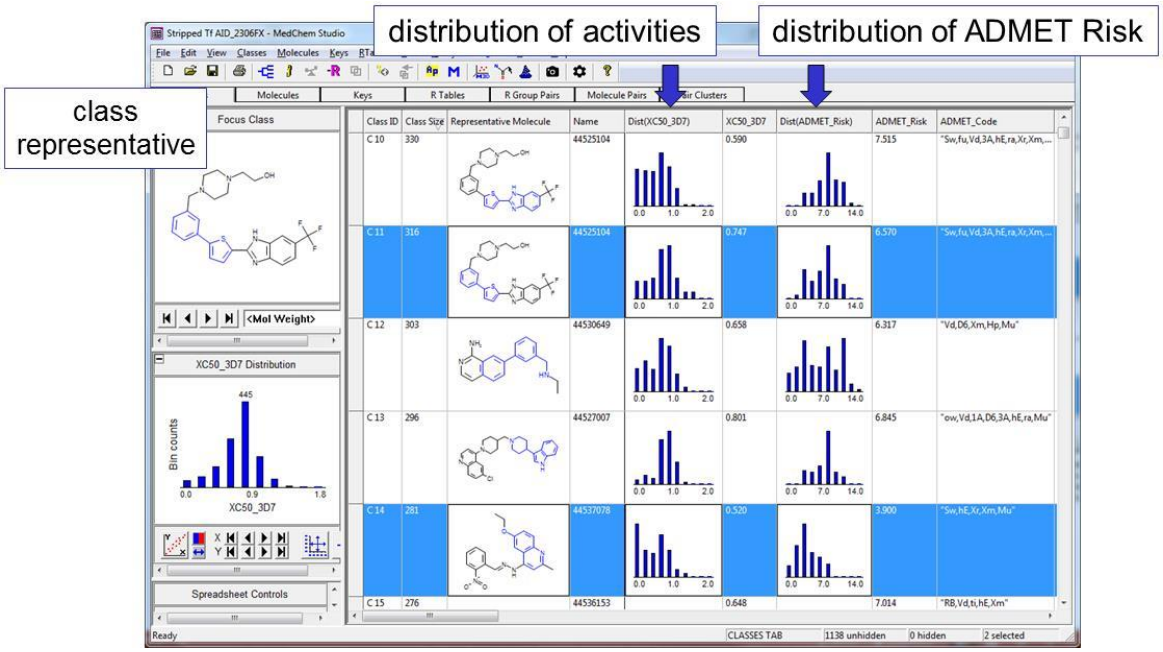


Orlistat is an obesity treatment that works by inhibiting lipases in the lumen of the stomach and small intestine, preventing hydrolysis of dietary fat into absorbable free fatty acids and monoglycerides. Undigested triglycerides are not absorbed. According to DrugBank (http://www.drugbank.ca/drugs): "*Systemic absorption of orlistat is minimal, however systemic absorption of the drug is not needed for activity*." Atorvastatin is 98% protein bound in plasma. It inhibits and is "extensively metabolised" by CYP 3A4 to metabolites that are themselves active. According to DrugBank: "…*Possible side effects include myotoxicity…and…hepatotoxicity*."

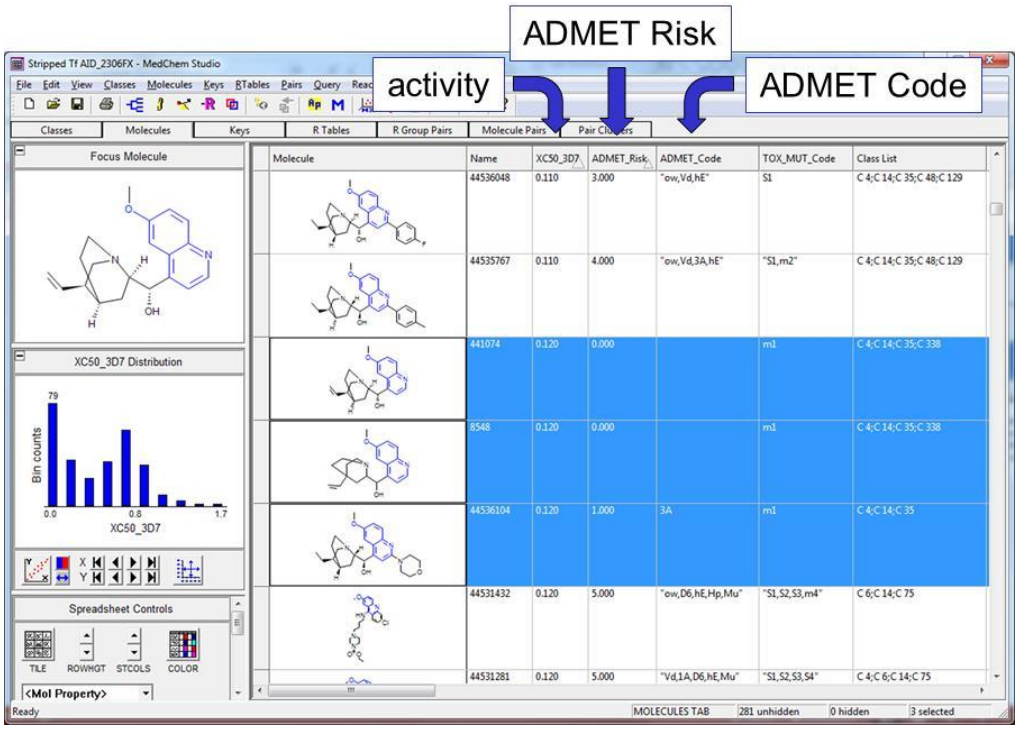
Using data taken from Zhao, Y. H.; Le, J.; Abraham, M. H. *et al.* Evaluation of Human Intestinal Absorption Data and Subsequent Derivation of a Quantitative Structure-Activity Relationship (QSAR) with the Abraham Descriptors. *J. Pharm. Sci.* **2001**, *90*(6), 749-784, Simulations Plus has shown that ADMET Risk is better than Rule of Five for refining absorption risks (F_a = fraction absorbed):



The scores and mnemonic codes can be used together with the series identification, combinatorial, and graphical analysis tools in MedChem Studio (http://www.simulations-plus.com/Products.aspx?pID=12) to prioritise drug development leads in a truly multi-dimensional manner. Clark grouped GSK antimalarials from PubChem AID 2306 by shared substructure and analysed them using MedChem Studio:



He also analysed individual compounds within one particular class:



ADMET Risk makes it possible to evaluate many different potential ADMET liabilities simultaneously. Used alongside measured activity, it can help identify lead series and leads likely to exhibit the best trade-off among multiple development objectives earlier in the discovery and development process than relying on activity alone. Combining ADMET Risk scores and codes with series identification, combinatorial, and graphical analysis tools like those in MedChem Studio makes it easy for medicinal chemists to prioritise drug development leads in a truly multi-dimensional manner.