# Integration of *in silico* and *in vitro* tools for scaffold optimization during drug discovery: Predicting P-glycoprotein efflux

**Prashant V. Desai[$]*, Geri A. Sawada[@], Ian A. Watson[#], Thomas J. Raub[@$]**

[$]Computational ADME and [@]Drug Disposition; [#]Advanced Analytics; Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, IN 46285

# Supporting Information

**Automated P-gp Substrate Screen**

Bidirectional transport of test compounds in the presence or absence of P-gp inhibitor LSN335984 also was performed robotically using a Tecan (Tecan Group AG, Mannendorf, Switzerland) with minor modifications detailed below. Cells were plated at 100,000 cells/well in 500-μL growth medium onto a Falcon 24-well HTS Multiwell insert system equipped with a 1.0-μm pore-size, polyethylene terephthalate (PET) membrane (Becton-Dickinson, Franklin Lakes, NJ). Growth medium (20-25 mL) was added to the common-tray basal chamber. Growth medium was changed 24 hours prior to experimentation on day 4. Compounds were tested at 10 μM with and without the P-gp inhibitor. Assay buffer was Hanks Balanced Salt Solution with 10 mM HEPES, pH 7.4 (HBSSH). Individual test compound aliquots were transferred to a 96-well polymerase chain reaction (PCR) plate and prepared for analysis by diluting them 1:10 in 33% acetonitrile/67% HBSSH. Amprenavir (Moravek Biochemicals, Brea, CA), dexamethasone (Sigma, St.Louis, MO), and sulfasalazine (Sigma, St.Louis, MO) were measured at each run as strong and weak P-gp substrates, and a very slow passive permeability standard, respectively. A singular, standardized LC/MS/MS method was used to analyze all test compounds.

**Assay Screen for P-gp Inhibition**

MDCK-MDR1 cells were plated at a density of 40,000 cells/well in 96-well, flat-bottom cell culture plates in a growth medium volume of 200 μL, which was replaced on day 3. On day 4, cells were washed once with PBSH, and incubated with test compound in PBSH at 37°C under room atmosphere. For the two-dose inhibition screen, cells were washed once with PBSH and pre-incubated with 80 μL of 5 μM or 25 μM test compound in PBSH at 37° C for 30 min. Incubation solutions were then changed to include 0.5 μg/mL calcein-AM, a P-gp substrate, and incubated for another 20 min. Intracellular calcein fluorescence was measured on a Cytofluor series 4000 multiwell plate reader (PerSeptive Biosystems, Framingham, MA) with $\lambda_{ex}$ and $\lambda_{em}$ set to 485 and 530 nm, respectively. Percent inhibition was determined for each test compound by comparing relative fluorescence units to that of cells inhibited 100% with 5 μM LSN335984.

**Molecular descriptors used to build the models**

We calculated more than 1800 locally-implemented descriptors, a brief explanation of only the most relevant descriptors that were incorporated in the models is provided here. The number of descriptors for each class is listed in the parenthesis.

1. Abraham Descriptors (6)
2. clogP (3)
3. MEDV (60)

4. Molecular connectivity (71)
5. Abraham and Platts (2)
6. E-state (316)
7. CATS (94)
8. Interpretable (205)
9. Crippen (165)
10. E-state-variant (278)
11. Heteroatom patterns (32)
12. Hydrogen bond (H-bond) patterns (73)

The Abraham descriptors[1] are the specific interaction terms used to describe solvation characteristics of a molecule. Biobyte version 4.3 (http://www.biobyte.com) was used to calculate clogP, augmented by a local implementation of offsets for logD (R.F. Bruns, I.A. Watson, manuscript in preparation). The Molecular Electronegativity Distance Vector (MEDV) descriptors by Liu *et al*[2] utilize an e-state-related measure for assigning atom electronegativies, and topological distances for distance measures. Additionally, the atomic attributes based on E-state are also implemented as descriptors.[2] Molecular connectivity descriptors have been described by Kier and Hall.[3] The Abraham and Platts descriptors are designed to more finely characterize hydrogen bonding capabilities within molecules.[4] CATS descriptors are a local implementation of the Chemically Advanced Template Search described by Schneider et. al.[5] A locally derived set of readily interpretable 2-D molecular descriptors were also included. These descriptors are generally easy to compute and easy to understand including parameters like heavy atom count, aromatic density, electron rich sections, molecular diameter, number of aromatic rings, atoms with pi electrons, TPSA and others. Crippen descriptors are a local, modified implementation of the group contribution methodology for estimating logP and molar refraction described by Wildman and Crippen.[6] E-state variant is a locally developed alternative determination of e-state values (unpublished). Heteroatom patterns are a substructure-based set of descriptors that perform an exhaustive search for all possible arrangements of heteroatoms in linear and/or T-shaped arrangements up to a given size. H-bond patterns are analogous to the Heteroatom patterns, but this time the features are either donors or acceptors as defined by a set of locally developed rules.

**Selection of the optimum QSAR modeling method**

A variety of methods for building classification models were evaluated including: (a) recursive partitioning (tree) - based algorithms such as Bagging[7] and Random Forest (RF)[8] and (b) support vector

machine (SVM) models, either based on descriptors (SVM-D)[9-10] or fingerprints (SVM-FP).[11] The Bagging models were built using 100 trees while RF models used 500. To compare the Bagging, RF and SVM-D models, compounds were split into training and test sets with a range of relative proportions for training: test sets such as 90:10, 65:35, 35:65 and 10:90, respectively. For each split, five different random sets were created to assess average performance of each model (see the model performance evaluation criteria described in the manuscript) built using the split training set and evaluated against the corresponding test set. Moreover, to assess whether these models provided a statistically significant improvement compared to random predictions, a random builder was used that predicts random responses derived from the training set. Such comparison is especially valuable when evaluating classification models.

For the SVM-FP models, a variety of fingerprints and fingerprint combinations were evaluated. Besides the more 'traditional' fingerprints such as extended connectivity circular fingerprints, atom pairs, ring substitution fingerprints, 2-D pharmacaphore (CATS), ErG,[12] path- and fragment-based fingerprints, relatively complex hybrid fingerprints incorporating various physicochemical properties were also utilized. These properties can include things like clogP, PSA and other whole molecule properties represented in fingerprint form. To select the optimum fingerprint, the original set was split into training: test sets in 60:40 proportions. This was repeated 50 times with stratified random assignment to the training or the test set to assess average performance for each fingerprint. The fingerprint that provided the highest average value of sum of PPV, NPV, sensitivity and specificity was selected to build the final model and its performance was compared with other models.

In general, all models performed significantly better (PPV and NPV values exceeding 65%) than the random builder. Based on this 'internal' assessment where the test sets were derived from the original set, Bagging, RF and SVM-FP models exhibited similar performance to each other with PPV and NPV values between 85 – 90 %. These three models were superior to SVM-D models which provided PPV and NPV values between 65-70%.

Next we compared the Bagging, RF and SVM-FP models based on predictions of chronological test sets consisting of data generated after the models were built. In this analysis both recursive partition-based models, Bagging and RF, produced better PPV, NPV, overall accuracy and kappa values than the SVM-FP model. The Bagging and the RF models were similar in performance, both for 'internal' assessment and chronological evaluation. Since the Bagging models were built using a relatively small number of trees (100) compared to the RF model (500), it takes significantly less time to return predictions for a large test set using the former. Based on the anticipated use of these models across multiple projects in parallel for a large set of compounds (including virtual libraries), we decided to implement the Bagging

models given their efficiency. Accordingly, for the purpose of this paper, detailed results are provided only for various versions of the Bagging models, utilized for supporting design and prioritization of synthesis and experimental measurement across a variety of drug discovery projects over a period of several years at Eli Lilly and Company.

**Table S1: Comparison of the automated P-glycoprotein substrate screen and the QSAR model against the outcome from the manual P-gp Substrate Assay.**[a]

| Evaluation Parameter | Automated P-gp Substrate Screen | QSAR model |
|---|---|---|
| Sensitivity | 70% | 86% |
| Specificity | 88% | 88% |
| PPV | 83% | 86% |
| NPV | 77% | 88% |
| Overall accuracy | 79% | 87% |
| kappa | 0.58 | 0.73 |

[a.] Values for various evaluation parameters are based on agreement between the binary results (substrates or non-substrates) from either the automated P-gp Substrate Screen or the QSAR model and those from the manual P-gp Substrate Assay

**References cited in the Supporting Information**


1. Platts, J. A.; Butina, D.; Abraham, M. H.; Hersey, A. Estimation of Molecular Linear Free Energy Relation Descriptors Using a Group Contribution Approach. *J. Chem. Inf. Comput. Sci.* **1999,** *39*, 835-845.
2. Liu, S. S.; Yin, C. S.; Li, Z. L.; Cai, S. X. QSAR study of steroid benchmark and dipeptides based on MEDV-13. *J. Chem. Inf. Comput. Sci.* **2001,** *41*, 321-329.
3. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Analysis;* John Wiley and Sons: New York, 1986.
4. Abraham, M. H.; Platts, J. A. Hydrogen Bond Structural Group Constants. *J. Org. Chem.* **2001,** *66*, 3484-3491.
5. Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. "Scaffold-Hopping" by Topological Pharmacophore Search: A Contribution to Virtual Screening. *Angew. Chem. Int. Ed. Engl.* **1999,** *38*, 2894-2896.
6. Wildman, S. A.; Crippen, G. M. Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.* **1999,** *39*, 868-873.
7. Breiman, L. Bagging predictors. *Mach. Learn.* **1996,** *24*, 123-140.
8. Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J. C.; Sheridan, R. P.; Feuston, B. P. Random forest: a classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Comput. Sci.* **2003,** *43*, 1947-1958.
9. Ovidiu, I. Applications of support vector machines in chemistry. In *Rev. Comput. Chem.,* Lipkowitz, K. B.; Cundari, T. R., Eds.; Wiley-VCH: Weinheim, 2007; Vol. 23, pp 291-400.
10. Vapnik, V. N. *The Nature of Statistical Learning Theory;* Springer: New York, 2000.
11. Erickson, J. A.; Mader, M. M.; Watson, I. A.; Webster, Y. W.; Higgs, R. E.; Bell, M. A.; Vieth, M. Structure-guided expansion of kinase fragment libraries driven by support vector machine models. *Biochim. Biophys. Acta* **2010,** *1804*, 642-652.
12. Stiefl, N.; Watson, I. A.; Baumann, K.; Zaliani, A. ErG: 2D pharmacophore descriptions for scaffold hopping. *J. Chem. Inf. Model.* **2006,** *46*, 208-220.