Effective Use of *In-Silico* Tools in Lead Optimization

Pranas Japertas[1], Andrius Sazonovas[1],
Kiril Lanevskij[1,2]

[1] ACD/Labs, Inc., A.Mickeviciaus g. 29, LT-08117 Vilnius, Lithuania,
[2] Department of Biochemistry and Biophysics, Vilnius University, M.K.Ciurlionio g. 21/27, LT-03101 Vilnius, Lithuania.

INTRODUCTION

Of all the challenges facing medicinal chemists in general, one of the most significant must be transforming an active molecule into a viable drug. Lead optimization efforts are guided by a combination of factors, such as potency, ease of synthesis, patentability concerns, specific synthetic constrains of the interaction with the target, as well as the lead's toxicity and ADME properties. Physicochemical profiling is carried out very early to get a jump-start in the drug discovery process. Even during hit-to-lead and lead optimization many individuals are involved in the determination of these properties to keep projects focused on molecules that are more likely to be good drug candidates. The advent of various *in silico* techniques has led many to believe these methods would become the 'holy grail' of the future drug discovery. However, despite constant advances in the field and the huge number of available models enabling predictions of a multitude of properties, computational approaches in general still fail to meet high initial expectations.
A rational approach to computer-aided lead optimization should involve a variety of techniques including predictions of ADME/Tox properties with a mechanistic insight, and routines that address other issues associated with the compound's suitability for use as a drug.

GLOBAL PHYSICOCHEMICAL RULES

The majority of existing QSAR models are statistics-driven, i.e., they are derived in pursuit of the best possible fit to available data. However, in lead optimization projects straightforward mechanistic interpretation of the models is a much more powerful feature than good statistical performance since it may help guide the efforts towards more promising candidates. Many biological properties in fact represent an interplay of the simplest physicochemical characteristics and can be successfully modeled by mechanistic approaches.

Brain delivery and intestinal absorption serve as good examples of such endpoints. Compounds entering the brain by passive diffusion across the blood-brain barrier can be classified as penetrating (BBB+) or non-penetrating (BBB–) on the basis of two quantitative parameters: the extent of brain/plasma partitioning at equilibrium and equilibration rate. As outlined in Table 1, they represent a multitude of processes mostly governed by lipophilicity and ionization [1]. The same property may have opposite effects on different processes, so achieving good BBB penetration means finding an optimal balance between the two properties.

Figure 1 illustrates the sigmoid relationship between the fraction absorbed in the intestine (f_a) and octanol/water log*P*. Sizes of the bars indicate the range of variation in f_a of equilipophilic drugs with varying H-bond donating potential (N_{HD}) and molecular size (V_x). f_a is very sensitive to slight changes in physicochemical characteristics on the steep part of the curve. Absorption of hydrophilic drugs also varies considerably, which can be explained by possible contribution of paracellular transport routes. [2].

BBB+/– category			
	Brain/plasma equilibration rate	↑ BBB permeability (log *PS*)	↑ LogP
			↓ Ionization
	Brain/plasma partitioning ratio (log *BB*)	↓ ↑ Brain tissue binding	↑ LogP
			↓ Acid pKa
		Plasma protein binding ↓	↑ LogP
			↑ Acid pKa

TABLE 1. Influence of physicochemical properties on processes involved in drug permeation across BBB.

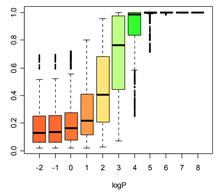


FIGURE 1. Fraction absorbed vs. log*P* for acids with N_{HD} = 4-6, V_x = 0.5-1.5.

Simple physicochemical trends are evident for complex properties characterizing protein-ligand interactions. Figure 2 illustrates a clear distinction between substrates (inhibitors) and non-substrates (non-inhibitors) of human P-glycoprotein (P-gp) on the basis of the compounds' molecular weight, lipophilicity (log*P*), and H-bond accepting potential represented by Abraram's *B* parameter. [3] Similar findings are observed when hERG channel inhibition is considered. Additionally, susceptibility to interactions depends on the ionization state: presence of basic center enhances hERG inhibition, while strong acidic groups (pK_a < 4) are detrimental for binding to both hERG and P-gp.

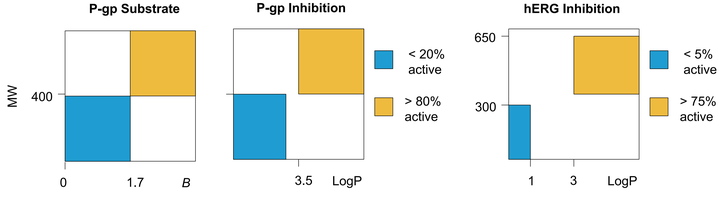


FIGURE 2. Global physicochemical rules for identification of potential ligands of P-gp and hERG.

PRE-SCREENING STRUCTURAL GROUPS FOR TOXICITY

Although accurate prediction of a toxicological profile of candidate compounds is a challenge on its own, insight on a compound's hazardous potential may also be gained from knowledge-based rules. It is known from literature that certain structural fragments are likely to be associated with high acute toxicity or genotoxic (carcinogenic) activity. Filtering out known alerting groups from the database of substituents, used in lead optimization, reduces the risk that the final optimized compound would have serious toxicity findings. Numerous literature reports suggest that even well-known structural moieties, such as an aromatic nitro-group, should be avoided [4] (see Figure 3).

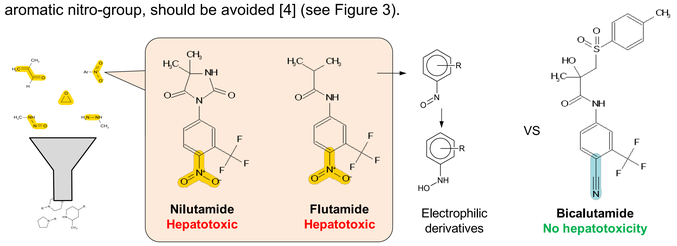


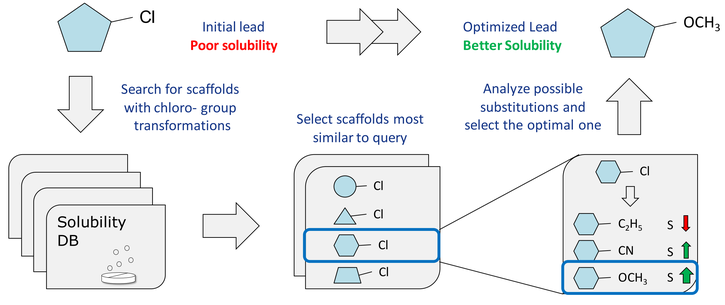
FIGURE 3. Biotransformation pathways of nonsteroidal antiandrogens leading to hepatic adverse effects.

LOCAL ACTIVITY CLIFFS: PAIRWISE QSAR

Although physicochemical rules and simple statistical models capture general trends in the relationship between the property of interest and structural features of compounds, in many situations such global rules are insufficient. This occurs in the case of local 'anomalies' –the so-called 'activity cliffs' –when small structural changes lead to unproportionally large changes in observed effects. Therefore, use of advanced statistical techniques such as pairwise QSAR (also referred to as molecular match pairs) is desirable in order to obtain a reliable estimate of the influence of changing substituent in the particular chemical context. The essence of pairwise QSAR method is as follows [5,6]:

- The property database is searched for the pairs of molecules that represent a particular structural modification.
- Among these pairs, the ones with core scaffolds most similar to the target query are selected. The effect of the proposed modification in the local chemical environment is evaluated on the basis of available data for selected transformations.

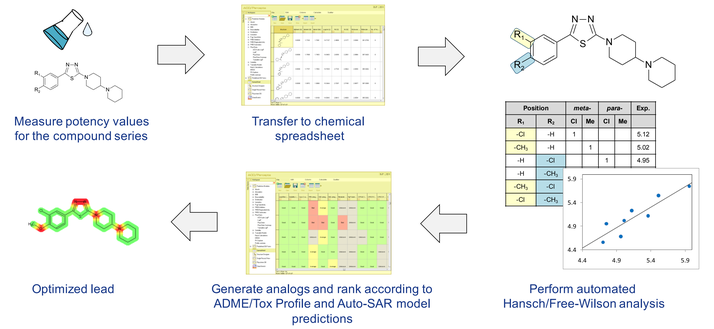
Scheme 1 shows how a pairwise QSAR can be used to identify the optimal structural modification that would result in the most favorable change of the analyzed property.



SCHEME 1. A workflow demonstrating the use of a pairwise QSAR in solubility optimization.

INTEGRATING POTENCY DATA: AUTO-SAR

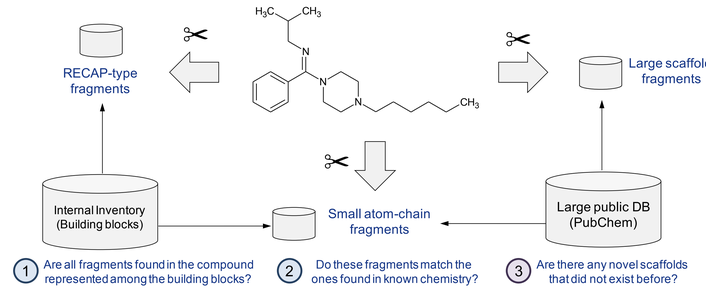
A natural extension of all above concepts related to ADME/Tox profiling would be to include available potency data in the analysis. This task can be accomplished by the means of automatic Hansch and/or Free-Wilson type QSAR analysis. A small dataset of measured potency values for 20+ compounds with substituent alteration performed in at least two sites would suffice for derivation of a simple Hansch type model describing the substituent contributions to the compound's overall potency in terms of their electronic effects. Alternatively, a Free-Wilson approach may be employed, directly relating the changes in activity levels to the presence or absence of a particular structural element [7]. The modified lead optimization workflow would then look like that depicted in Scheme 2. This approach would solve the imminent issue that candidates suggested by the software solely on the basis of their ADME/Tox profiles may fail the potency requirements.



SCHEME 2. A proposed workflow of *in silico* lead optimization involving ADME/Tox profiling combined with Auto-SAR utilizing available potency data.

COMMERCIAL FATE: SYNTHESABILITY AND PATENTABILITY

The factors that will impact the commercial fate of a particular analog are not limited to its potency and ADME/Tox profile. There are also a number of aspects that are not associated with a compound's suitability to be used as a drug, namely synthetic feasibility and patentability prospects. Although conceptually different, these are still inherent parts of the lead optimization process. Computational methods used to evaluate whether the compound is readily synthetically accessible and whether it belongs to a substantially novel chemical class rely on fragmenting the molecules according to a specific set of rules and comparing the frequencies of found fragments with those in known databases [8].



SCHEME 3. Possible approaches to assess the compound's ease of synthesis and structural novelty.

CONCLUSIONS

A key element for successful utilization of *in silico* tools in lead optimization is mechanistic interpretability of predictive models. Rational strategy of lead optimization should aim at achieving balanced physicochemical profiles of candidate compounds that would translate into the desired ADME properties. In those cases when a successful compound does not follow general rules, local models (pairwise QSAR) might help in achieving the desired balance. Compound synthetic accessibility and novelty can also be assessed using chem-informatic approaches, comparing substructures of different sizes in the compounds with the ones in available chemistry.

REFERENCES

[1] Lanevskij K et al. *J Pharm Sci.* **2011**, 100, 2147.
[2] Reynolds DP et al. *J Pharm Sci.* **2009**, 98, 4039.
[3] Didziapetris R et al. *J Drug Target.* **2003**,11, 391.
[4] Walsh JS & Miwa GT. *Annu Rev Pharmacol Toxicol.* **2011,** 51, 145.
[5] Papadatos G et al. *J Chem Inf Model.* **2010**, 50,1872.
[6] Leach AG et al. *J Med Chem.* **2006**, 49, 6672.
[7] Kubinyi H. QSAR: Hansch Analysis and Related Approaches. *Wiley VCH*, **1993**, 240 p.
[8] Ertl P & Schuffenhauer A. *J Cheminform.* **2009,** 1, 8.

ACD/Labs
Advanced Chemistry Development, Inc.
Tel: (416) 368-3435
Fax: (416) 368-5596
Toll Free: 1-800-304-3988
Email: info@acdlabs.com
www.acdlabs.com