Review

Leveraging chromatography based physicochemical properties for efficient drug design

Gilles H. Goetz*, Marina Shalaeva

Molecular Properties Group, Worldwide Medicinal Chemistry, Pfizer Worldwide Research & Development, Pfizer, Inc., Groton, Connecticut 06340, United States

*Corresponding Author E-mail: gilles.h.goetz@pfizer.com; Tel.: +1-860-715-6311

Received: March 28, 2018; Revised: May 26, 2018; Available online: June 04, 2018

Abstract

Applications of chromatography derived lipophilicity, polarity, and 3D concepts such as conformational states, exposed polarity and intramolecular hydrogen bonds (IMHB), are discussed along with recently developed methods for incorporating these concepts into drug design strategies. In addition, the drug design process is described with examples and practices used at Pfizer, as well as experimental and computed parameters used for parallel optimization of properties leading to drug candidate nominations.

Keywords

ElogD, EPSA; lipophilicity, exposed polarity

Introduction

An active molecule with favorable pharmacokinetic/pharmacodynamic (PK/PD) properties is the goal of drug design. This means a biologically active molecule can be developed into a drug only if it fits strict pharmacological profile requirements. Tablets are the preferred form of administration due to convenience, patient compliance and ease of manufacturing. Therefore most drug design projects are focused on creating drugs suitable for oral dosage.

In order to be absorbed in the gastrointestinal (GI) tract, a solid formulation must be ingested, solubilized. Only then can the solvated drug be absorbed into the blood stream. Here drug distribution, metabolism and excretion affect its clearance, defining the pharmacokinetic parameters, half-life and Therapeutic Index (TI). Balancing Absorption, Distribution, Metabolism, Excretion and Toxicity (ADMET) properties is a major challenge in drug discovery and development. Failure to achieve the right ADMET balance was cited in 1997 as the reason for the attrition of 40% of drug candidates during development [1]. Here we present and discuss recent advances and approaches, mostly developed at Pfizer, using chromatographically determined physicochemical properties for the optimization of ADMET properties.

Lipophilicity

Lipophilicity (as defined by IUPAC) represents the affinity of a molecule or a moiety for a lipophilic environment. It is commonly measured by its distribution behavior in a biphasic system, either liquid-liquid...
Lipophilicity may be expressed as $\log P$ for neutral molecules and $\log D$ for ionized molecules at a particular pH ($\log D_{\text{pH}}$), where permeation $P$ and distribution $D$ are defined by the neutral and ionized species concentration ratios:

$$P = \frac{[\text{neutral}]_{\text{octanol}}}{[\text{neutral}]_{\text{water}}} \quad D = \frac{[\text{ionized + neutral}]_{\text{octanol}}}{[\text{ionized + neutral}]_{\text{water}}}.$$

Successful drug designs require careful balancing of in vivo and in vitro properties. Lipophilicity is a descriptor intrinsic to many of these properties. Computational models of ADMET properties generally include $\log P$ as a significant descriptor of the modelled parameter.

An extensive analysis of the literature linking ADMET and lipophilicity in drug discovery was published by Waring [3]. It details important, but obscure aspects of the $\log D$ parameter. For example, many studies rely on compilations of $\log P$/log $D$ values from a variety of sources with poorly described experimental details. These non-homogeneous datasets lead to discrepancies in analyses. There is also confusion between $\log P$ and log $D$; terms often used interchangeably.

For drug compounds ionized at physiological pH, $\log D_{\text{pH}}$ is more relevant than $\log P$, however, it deviates because of its dependence on pKa. This is especially flagrant, in case of predicted $\log D_{\text{pH}}$, due to unreliable pKa predictions. The widely cited publication by Mannhold et al. [4] compares several computed $\log P$ models to over 96,000 measured values. Unfortunately, actual experimental details were not provided.

Many research organizations have developed their own methods for lipophilicity measurements. They use their particular values for drug design and optimization, i.e. chromatographic hydrophobicity index (CHI) at GlaxoSmithKline [5], and Elog $D$ at Pfizer [6].

Early stage drug design strives to create the most active molecule for a particular therapeutic target. The ultimate goal is the discovery of a molecule possessing “drug-like” properties. The reliance on lipophilicity to understand molecules' drug-like properties was defined in 1997 Lipinski’s “rule of five” (RO5) [7]. Now it is ingrained into drug design that oral drugs require a cLogP $< 5$. Rules, depending mostly on log $D$, help guide designs towards optimal drug-like chemistry space. Log $D$ is used to calculate ligand efficiency (LE) [8,9], metabolic efficiency (MetE) [3,10,11], central nervous system multiparameter optimization (CNS MPO) [12,13], and toxicity scores [14-16].

Recent developments are reviewed in the application of log $D$ for balancing potency and ADME properties.

**Lipophilicity and ligand efficiency**

The review by Hopkins et al. [17] concluded that recently marketed oral drugs “frequently have highly optimized ligand efficiency values for their targets”. Early stage design efforts often lead to series of hundreds of newly synthesized compounds. These structures need to be rapidly evaluated for projects to proceed. Promising leads may be prioritized using lipophilic ligand efficiency (LLE) [9] or LipE [10]:

$$\text{LLE (or LipE)} = pIC_{50} - \text{cLogP}.$$

The LLE/LipE concept has been developed following the introduction of ligand efficiency (LE) [8]. The LE parameter allows direct comparison of molecules through their target binding energy. LE can be expressed
by the following simplified equation in which N is the number of non-hydrogen atoms:

\[ \text{LE} = 1.37(-\log \text{IC}_{50})/N \]

Project progress may be tracked using both LE and LipE; LipE is expected to reach values of 6 or 7 for a potential drug candidate to be chosen for further development [18-20].

Target binding increases with the addition of a lipophilic fragment, at the expense of potential for off-target binding and decreased solubility. An efficient drug design requires potency to be achieved with a minimal increase in MW or lipophilicity [21-24].

**Lipophilicity and oral absorption**

Oral administration, the most desirable route of medicine delivery, requires a drug to be absorbed while in the GI tract, with cell membrane permeation being the major determinant of absorption.

There are several processes involved in crossing the intestinal mucosa, including active transport [15], paracellular transport, and drug efflux. Passive membrane diffusion is considered to be the major contributing process [25]. Recent analyses continue to demonstrate the significance of the non-specific permeation mechanisms governed by molecular properties (namely molecular size and polarity) [16].

The favorable property space for an orally administered drugs is defined by the aforementioned Ro5 [26,27]. In order to achieve absorption, an oral drug needs to have a cLogP < 5; MW < 500; hydrogen bond donor (HBD) count (NH+OH) < 5; and a hydrogen bond acceptor (HBA) count (N+O) < 10.

More specific rules were developed for central nervous system (CNS) projects, where crossing the blood-brain barrier is challenging due to tighter junctions and higher efflux. The CNS MPO score introduced by Wager et al. [12] was derived by analysis of drugs and candidates with regard to their ability to cross the blood-brain barrier. The CNS MPO includes ionization constant (pKa), \( \log P \), and \( \log D_{7.4} \). This emphasizes the importance of ionization and lipophilicity adjusted for ionization [13,23,28,29]. Recently Z. Ranković further refined the CNS desirability criteria in CNS MPOv2 [30].

Lipophilicity measurements also help to estimate solubility in discovery settings [31]; solubility decreases with \( \log P > 3 \).

**Current trend towards and beyond Ro5 compounds**

Recently, the introduction of innovative “flat” targets and the potential for higher potency and selectivity provided by larger molecules, has led to a renewed interest in peptide-based drugs. These fill the gap between small molecules and biotherapeutics, as illustrated in Figure 1 [32,33].

The exploration of new targets required new design concepts including “an update” to Ro5. Bunnage et al. defined “four pillars of target validation”, the first pillar being “exposure at the site of action”, defined largely by permeability, where tactics such as increasing lipophilicity could help improve both potency and permeability [34].

Failing one or more of the Ro5 may decrease oral bioavailability, due to low solubility, and or poor membrane permeability, in addition to metabolic instability, synthetic complexity and other disadvantages of large drug molecules.

The study by Guimarães et al. [35] highlighted the importance of 3-D properties for passive permeation of beyond Ro5 (bRo5) molecules, where taking molecular conformation into account, as well as size and polarity assessments, are needed to aid design in bRo5 space.
The Ro5 defines properties of drug-like molecules, essentially, via molecular size and polarity, where MW describes the size, HBD and HBA counts describe polarity and log D describes both. Notably, MW and NH, OH, O and N are just simple counts and only roughly describe interactions with the environment, especially for molecules of any complexity. For example, intramolecular hydrogen bonding (IMHB) in large, flexible molecules could lead to shielding of HBD and HBA and, consequently, to changes in molecular shape and size, which are not accounted for by MW or HBD and HBA counts. These simple counts, including polar surface area (PSA), are indifferent to changes in molecular shape and size, the major determinants of passive diffusion [35].

Experimental log D, on the other hand, accounts for molecular size and polarity in solution and for 3D conformational changes adopted by the molecule. Therefore, the “extension” of the Ro5 to large molecules depends even more on the “effectiveness” of the log D parameter. While the accuracy of log D prediction is important for large, flexible molecules, it is also much more difficult to predict, as it deals with the assessment of multiple conformations and their populations. Measured log D values are necessary to build and test such computational models for bRo5 compounds. However, log D values > 3 are challenging to obtain via conventional shake-flask log D (SFlog D), due to limits in quantitation for the concentration of above 1000 between phases.

Kihlberg et al. analyzed orally available drugs in bRo5 space [36] and observed significant extension, almost doubling in values of Ro5 parameters (Figure 2). They proposed an “extended Ro5 space” (MW < 700, 0 < cLogP < 7.5, HBD < 5, PSA < 200 Å², and the number of rotatable bonds (NRotB) < 20. They also proposed a substantially larger “possible to be oral” bRo5 space with the limits for oral bioavailability extended to approximately MW < 1000, -2 < cLogP < 10, HBD < 6, HBA < 15, PSA < 250 Å², NRotB < 20 (dashed box). Notably, HBD has hardly changed (increased by only 1 count) highlighting the significance of limiting HBD exposure for oral absorption of large molecules.
ADMET & DMPK 6(2) (2018) 85-104  
Leveraging chromatography physchem properties in drug design

Figure 2. Physicochemical property space of drugs and clinical candidates with MW > 500. Solid box marks “extended Ro5” space. Dashed box marks “bRo5” space. Reprinted from [36] Copyright (2014), with permission from Elsevier.

Wang et al. [37] observed a correlation of RP HPLC log \( k' \) with passive permeability determined by human colorectal adenocarcinoma cells (Caco-2) and the parallel artificial membrane permeability assay (PAMPA) for a set of macrocyclic peptides. Bockus et al. used Elog \( D \) (5.5 to 7.5) to study macrocycles. The lipophilicity values correlated with cell-based permeability values by Madin-Darby canine kidney (MDCK)-low efflux cells (MDCK-LE) [38].

Lokey et al. investigated Ro5 and bRo5 compounds with MW > 1000 and demonstrated a sharp decline in apparent passive permeability for compounds with molecular size above 750 Å [39]. They also concluded that bulk physical properties contributing to passive permeability could be approximated by lipophilicity and molecular size (\( \log K_{\text{hydrocarbon/water}} \) and MW respectively). Molecular size relates to passive permeability with pKa and aqueous solubility also very important properties to consider. The lipophilicity parameter used by Lokey et al. (\( \log K_{\text{hydrocarbon/water}} \)) is lipophilicity measured in a non-polar environment and is different from \( \log D_{\text{octanol/water}} \). It has been shown to correlate well with passive permeability measured by PAMPA and MDCK-LE.

The use of \( \Delta \log P = \log P_{\text{alkane/water}} - \log P_{\text{octanol/water}} \) was first introduced by Seiler [40] and applied to improve absorption [41], as a guide in the design of novel brain-penetrating H2 antagonists [42], or as a measure of HBD acidity [43,44]. Hydrogen bonds characteristically feature binding energies and contact distances that can lead to large variations even for a single donor-acceptor pair [45].

The 96-well plate-based shake-flask \( \log P_{\text{toluene/water}} \) method (pH 1 to 11) was developed along with the IMHB interpretation scheme based on \( \Delta \log P \) [46]. This method allows verification of molecular conformations predicted by COSMO-RS software [47] which describes the geometry of virtual molecules’ interactions in both polar and non-polar environments.

Furthermore, an RP HPLC method using a polystyrene-divinylbenzene stationary phase (PLRP-S) was developed to simulate a non-polar lipidic membrane environment [48], and to obtain experimental lipophilicity values.

Kihlberg et al described errors in cLogP; \( \log P_{\text{oct}} \) predictions that were far from the experimentally determined values [36]. Posaconazole has a cLogP of 5.4, while the experimentally determined \( \log P \) is 2.4 [49]. However, in this particular case, the measured value may be incorrect.

These studies reiterate the need for an extended range of lipophilicity measurements for bRo5 compounds that would accommodate the large lipophilic molecules designed for current targets. They highlight the importance of lipophilicity as an approximation of molecular size and conformations in specific environments.

doi: 10.5599/admet.529
Lipophilicity and Clearance

Avoiding highly lipophilic compounds is a design principle used to improve clearance [50]. The analysis of matching pairs conducted by Stepan et al. [11] confirmed that changes in metabolic stability largely come from changes in lipophlicity (Figure 3). To determine how structural changes affect the relative clearance between analogues, the lipophilic metabolism efficiency parameter (LipMetE) was introduced. LipMetE relates lipophilicity to the in vitro metabolic clearance measured by the human liver microsomes (HLM) assay. LipMetE and unbound clearance (CL_{int,u}) are defined as follows:

\[
\text{LipMetE} = \log D_{7.4} - \log (\text{CL}_{\text{int,u}}),
\]

\[
\text{CL}_{\text{int,u}} = \frac{\text{CL}_{\text{int,app}}}{f_{u,\text{mic}}}. \]

![Figure 3](image.png)

Figure 3. The plot of unbound clearance, CL_{int,u} versus experimental log D, Elog D. The 45° lines represent different values of LipMetE. Compounds that parallel the same 45° line offer the same ratio of metabolic clearance to lipophilicity. Reprinted with permission from [11]. Copyright (2013) American Chemical Society.

Stepan et al. [11] suggested choosing high LipMetE compounds as a starting point for optimization due to their wider \( \log D \) range. Thus, potency and permeability could be improved while maintaining low clearance. LipE and LipMetE have become complementary opposites ("yin and yang") in medicinal chemistry decision making at Pfizer and the authors provide specific guidance on simultaneously optimizing these lipophilicity driven parameters.

Optimization of ADME properties, such as cell membrane permeability and metabolic stability, often comes to an act of balancing these orthogonal parameters. Gleeson [14] describes the contribution of MW, ionization state, and cLogP to in vivo clearance. Log clearance (log Cl) was largest for a cLogP > 5, less pronounced for cLogP 3–5, and most favorable for a cLogP < 3. Johnson et al. [16] analyzed permeability and clearance data for 47,018 Pfizer compounds and observed broad trends for favorable clearance and permeability compared to MW and log D. Using lipophilicity determination, they identified a “golden triangle” of MW and log D values for optimal permeability and metabolic stability. Figure 4 shows permeable and stable compounds in blue and compounds failing permeability or metabolic stability in grey against MW and log D. Compounds with favorable clearance and permeability properties are clustered within this “golden triangle” of MW and log D.
Leveraging chromatography physchem properties in drug design

**Figure 4.** Combined *in vitro* permeability and clearance trends across MW and log $D$. Reprinted from [16] Copyright (2009), with permission from Elsevier.

**Lipophilicity and Toxicity**

Avoiding high lipophilicity is also critical for reducing the probability of adverse safety findings [3, 51]. Hughes *et al.* [52] demonstrated that lipophilic compounds (cLogP > 3) with a low polar surface area (PSA < 75 Å²) have a 6-fold greater risk of toxicity findings in preclinical toxicology studies [53].

An alternative to using log $P$, which could oversimplify *in vivo* behavior of compounds, was proposed by Wenlock [54]. This criterion is based on the amount of compound in the body at a steady state. The relationship of this criterion at an acceptable human dose of 0.5 mg/ kg for 242 oral drugs with different *in vivo* plasma clearances was established with regard to their safety profiles.

**Experimental log $D$ determinations**

The optimal range of log $D$ to satisfy ADMET properties centers on 2 (+/-1) [3], but expands far beyond that range for some marketed drugs [36]. The wide range of lipophilicity employed in modern drug design requires experimental methods and computational models to support exploitation of these chemical spaces.

The chromatography based Elog $D$ method [6, 55] has been developed at Pfizer to alleviate limitations the of the classical shake-flask technique and to extend to compounds bRo5. Since its introduction, it has been used successfully in a traditional chemistry space [22, 56-60] as well as in the bRo5 programs [38].

The simultaneous optimization of potency and clearance using both measured and calculated Elog $D$ has been described for the Takeda-G-protein-receptor-5 (TGR5) program seeking an orally available compound for improved glycemic control via glucose-dependent insulin secretion. This has led to significantly improved clearance (HLM < 100 ), log $D$ = 2, and equivalent potency compared to competitor compounds with high clearance HLM > 200 and log $D$ > 2 [50].

Lipophilicity (Elog $D$) has also been utilized in clinical studies to understand the specific and non-specific binding of an active ingredient to the beta-amyloid plaques in Alzheimer’s disease patients which has been observed with high contrast positron emission tomography (PET) imaging [61].

At Pfizer, log $D$ at pH 7.4 is routinely measured for most compounds using either SFlog $D$ or Elog $D$ or both methods; each method has limitations. While Elog $D$ cannot be applied to acids, SFlog $D$ has a limitation in measuring values above 3.5 due to the increasing errors in quantification of concentration.
extremes (indeed, log $D = 4$ implies that the concentration in 1-octanol is 10,000 times higher than in the aqueous phase). Additionally, an accurate prediction of log $D$ at pH = 7.4 is still an issue for novel structures and therefore comparisons of SFlog $D$ and Elog $D$ values are often useful.

With project X, compounds were evaluated using SFlog $D$ (and occasionally Elog $D$), but significant discrepancies were observed for “low” SFlog $D$ samples (Figure 5. (a)). However, if Elog $D$ – SFlog $D$ differences (residuals) were plotted against Elog $D$ it became apparent that the largest discrepancies were in the range above 3.0. (Figure 5(b)). This illustrates how the SFlog $D$ procedure may underestimate lipophilicity values above 3.

**Figure 5. (a)** Comparison of SFlog $D$ and Elog $D$ values for project X; **(b)** SFlog $D$ - Elog $D$ residuals vs. Elog $D$

In summary, if analyzed by Elog $D$, both SFlog $D$ and Elog $D$ give, as expected, very similar values (within +0.5) in the Elog $D < 3$ range. However, in the Elog $D > 3$ range, many SFlog $D$ values are under-evaluated (by up to 2 log units). Therefore, it is important to recognize and respect the experimental methods applicability domain when using log $D$ for SAR or building computational models based on such data.

**Computed log $D$**

Day-to-day experience in measuring log $D$ demonstrated that despite great efforts to improve log $D$ calculations, there were still instances where log $D$ predictions are inaccurate; especially if ionization is involved and an accurate pKa is needed to adjust the calculated log $P$ of a neutral form to a log $D$ at a particular pH, usually 7.4.

The cPFLogD model built at Pfizer takes into account both SFlog $D$ and Elog $D$ experimental values while respecting the limits of each method, i.e. SFlog $D < 3$ and Elog $D$ for bases and neutrals only. Consequently, the model gives preference to SFlog $D$ values in the log $D$ range below 3 and to Elog $D$ values in the log $D$ range above 3. The cPFLogD model is built using the Cubist non-linear in-silico regression methodology and it is regularly updated with the latest experimental data, thus providing improved cPFLogD predictions for the next design cycle.

Performance of the cPFLogD model is demonstrated in Figure 6 on a rather challenging subset of diverse “natural-product” like compounds. This subset of about 300 molecules is more complicated than encountered in most small molecule projects. It encompasses MW from 200 to 2000, has compounds violating up to four Ro5, with high NRotB, HBD and HBA.

For many compounds in the log $D > 4$ range a SFlog $D$ value could not be acquired due to extremely low concentration in the aqueous phase and therefore Elog $D$ values were measured on all compounds, excluding acids.
Figure 6. (a) Measured ElogD vs. cPFLogD; (b) Measured SFlog D values vs. cPFLogD. A diverse subset of “natural product”-like library, about 300 compounds. The size of the markers represents Molecular Weight (193 to 2013 Da). The colours represent violations of the Ro5.

As expected, cPFLogD predicted SFlog D more accurately in the < 2 range and, Elog D more accurately in the > 2 range. Several outliers in the upper left corner of the Elog D vs. cPFLogD Figure 6(a) represent the underestimated SFlogD results used by the cPFLogD model leading to underestimated prediction values.

Comparison of cPFLogD to commercially available ACDlabs log D_{7.4} predictions for the same subset are shown in Figure 7, where many compounds violating three Ro5 criteria (yellow markers) fell far outside cPFLogD, especially in the low lipophilicity range.

Figure 7. Lipophilicity calculations using log D_{7.4} by ACD labs vs. cPFLogD for a subset of a “natural product”-like library.

It should be emphasized that project teams usually design around an active lead molecule and work on a few series, where “diversity” in a general chemistry sense is rather limited. Therefore, measured log D values determined on a few compounds in a series introduced into the model often allow significant improvement in accuracy of cPFLogD predictions in that chemical space.

Polarity

In the otherwise comprehensive IUPAC Compendium of Chemical Terminology, also known as the Gold Book, (unlike lipophilicity, hydrophobicity and hydrophilicity), the compound polarity is not defined [2]. The only reference to polarity in the Gold Book is coupled with the concept of solvent rather than solute (see below).
When applied to solvents, the term polarity covers their overall solvation capability (solvation power) for solutes (i.e. in chemical equilibria: reactants and products, in reaction rates: reactants and activated complex, in light absorptions: ions or molecules in the ground and excited state). A solvent’s solvation power depends on the action of all possible intermolecular interactions between solute ions or molecules and solvent molecules. Those interactions can be both nonspecific and specific but exclude the interactions leading to definite chemical alterations of the ions or molecules of the solute.

**TPSA**

In the context of drug design, the polarity of a molecule has been redefined as its polar surface area (PSA). Originally computed using molecular mechanics calculations, the fragments of the dynamic van der Waals’ surface area associated with oxygen, nitrogen and their attached hydrogen atoms were considered as the polar portions of the molecular surface area. This showed correlation to cell permeability in Caco-2 cells for a homologous series of beta-adrenoreceptor antagonists [62].

The popularity of PSA has recently increased significantly, due to its performance in predicting intestinal adsorption as well as blood-brain barrier penetration [63-70].

In 2000 Ertl et al. introduced the topological polar surface area (TPSA)[71], an index based on the addition of tabulated surface contributions of polar fragments. Based on ~35,000 drug-like molecules, the surface contributions of 43 fragments centered on polar oxygen, nitrogen, phosphorus and sulfur were determined by the least square fit and tabulated. This methodology proved 2-3 orders of magnitude less computationally intensive than the previous PSA calculations and gave similar results [71]. The use of PSA in medicinal research has since been reviewed by Ertl [72].

In addition to dynamic, molecular and topological polar surface areas, quantum mechanical polar surface area (QMPSA) was recently introduced and showed a good correlation with the fraction absorbed (FA) after oral administration for a set of 18 drugs when carboxyl groups were deprotonated, suggesting adsorption to be strongly related to polar interactions of molecules in water solution [73]. It should be noted that significant increases in computing resources are needed to generate QMPSA, compared to the table-entry recalling TPSA.

While mostly used for barrier crossing prediction in an ADME context, TPSA has also been recently investigated as a descriptor for 2D-QSAR for diverse pharmacological activity data [74], as well as for active drug transport by multidrug resistance associated protein 1 (MRP1) [75].

The first study pointing to a hard PSA limit dates back to 1997, in which Palm et al. [65] showed on a diverse set of 20 model drugs, that fully absorbed drugs (FA > 90 %) had a PSA ≤ 60 Å² while drugs that are less than 10 % absorbed had a PSA > 140 Å². Later, Veber at all [76] showed that a PSA ≤ 140 Å² and the number of rotatable bonds ≤ 10 is as efficient and selective a criterion as the Lipinski’s Ro5 for selecting oral bioavailability of > 20-40 %. As far as the blood-brain barrier was concerned, it was found that the upper limit for PSA for a molecule to penetrate the brain was around 90 Å² [69,77].

**TPSA in MPO**

In 2010 Wager at al introduced the CNS MPO (multiparameter optimization) desirability tool, which incorporated TPSA along with 5 other fundamental physicochemical properties: cLogP, cLogD, MW, pKa, and HBD. A monotonic decreasing function was used for cLogP, cLogD, MW, pKa, and HBD, and a hump function was used to define TPSA (Figure 8). The CNS MPO desirability method is quite simple, with parameters derived from medicinal chemistry best practices, and it is able to balance multiple variables...
while avoiding hard cut-offs. It demonstrated a good correlation between a high score (> 4 out of a possible max of 6) and good in vitro ADME properties [12,78].

Figure 8. Components of CNS multiparameter optimization desirability tool (MPO)

EPSA

Having access to TPSA, a simple, inexpensive and efficient way to assess polarity from a 2D structure, which correlates reasonably well with cell permeability, the need for a polarity measurement was never really felt by drug design professionals.

While historically relevant for the targets pursued in the late 1990s, criteria such as the Ro5 or the Veber PSA-Rotatable bonds rule [76], tend to lose traction due to the rarity of easy targets, resulting in drug discovery programs having to foray into more challenging chemical spaces. In Jürgen Drews’ words “one truth is there for all to see: many ‘easy’ targets or molecules have been found and developed” [79].

Over the time, simplification of the polar surface area calculations down to TPSA may have led to the false equivalency of certain considerations such as the actual position of a polar group in a molecule with regard to its immediate and adjacent environment. For example, the TPSA values for ortho-, meta- and para- analogues of any aromatic polar compound are identical, and similarly, the TPSA values of any regioisomers are also identical, due to the nature of the TPSA calculation, solely based on adding fragment contributions to overall polarity. We contend here that each polar fragment does not contribute equally to the polar surface area of a molecule; in essence advocating a return to a more “3D-PSA”-like descriptor [35,62].

Utilizing separation sciences expertise, a team of Pfizer scientists developed an assay which outputs a polarity readout termed EPSA (not an acronym) [80]. The main driver behind EPSA was to have access to a fast, robust method capable of identifying the potential for a compound to form intramolecular hydrogen bonds (IMHB) enabling it to hide polarity that might otherwise reduce its passive permeability. Indeed, the identification of compounds likely to form IMHBs is an important drug design consideration given the correlation of intramolecular hydrogen bonding with increased membrane permeability [81-85].

The EPSA method provides, under controlled supercritical fluid chromatography (SFC) conditions, an exposed/hidden polarity readout that is derived from the retention time of a compound on a specific column. The stationary phase, (Phenomenex Chirex 3014) was selected for its balance of lipophilic and polar attributes and its capacity to separate compounds with wide polarity differences. SFC, which essentially uses normal phase-like conditions, provides an environment with a low dielectric constant conducive to IMHB formation. Polar compounds are retained more under these conditions, and a low-slope gradient of methanol achieves elution based on the increasing polarity of the mobile phase. In the cases that have been studied, using matched molecular pairs with or without IMHBs, compounds with IMHB resulted in a significant reduction in polarity and eluted earlier in the chromatogram. Results are normalized through the use of calibration standards generating a linear relationship between retention times and EPSA values [80].

This is possible to confirm or disprove the presence of an intramolecular hydrogen bond (IMHB) in small
molecules or peptides using EPSA and pairwise analysis. Such IMHBs are a critical factor affecting the conformation of molecules and can have a direct impact on their absorption into the human body, their permeability into target cells, and even their potency against therapeutic targets.

This method has been successfully used in a small molecule medicinal chemistry project to assess exposed polarities of their lead compounds. Cheng et al. [86] determined EPSA for their mechanistic target of rapamycin (mTOR), phosphatidylinositol-3 kinase (PI3K) dual inhibitors series and found no direct correlation between EPSA and TPSA. However, they showed that the formation of IMHB can significantly reduce the effective polar surface area. Suggesting that for designs with high TPSA, the introduction of an IMHB can help achieve good permeability and cellular potency by modulating the EPSA to a desirable range. Indeed, several compounds, with TPSA greater than 135 Å², exhibit EPSA values between 82 and 101, which is a range capable for this series of yielding permeability and cellular potency.

Wakenhut et al. [87] determined EPSA for non-structural protein 5A (NS5A) inhibitors and notably compared two analogues: 12 and 13. Compound 12 has less HBD and HBA, less rotatable bonds, lower TPSA (105 Å²) and, by all standard medicinal chemistry principles, 12 was expected to be more permeable than 13 (TPSA 175 Å²). It turned out that the opposite was true; compound 12 was not permeable while 13 was permeable. Compound 12 had an exposed polarity value (EPSA) of 128, while the EPSA value of 13 was 103, significantly lower than for 12.

These facts can be explained by the peripheral functionality capable of forming IMHBs in 13 within a membrane environment, thereby masking H-bonding (HBD and HBA) character and contributing to passive permeability, as compared with the more extended conformations available to compounds such as 12 incapables of forming IMHBs (Figure 9). The team based its subsequent designs on enabling IMHB formation within the analogues being developed.

![Figure 9](image)

Figure 9. Representation of compound 12 and 13. Compound 13 initially in solution phase (6 HBA, 4 HBD, PSA = 175 Å²) and as it traverses a hydrophobic membrane (4 HBA, 2 HBD, 3D-PSA~120 Å²) as a possible explanation of the good membrane permeability of 13. Adapted from [87] Copyright (2014), with permission from Wiley.

**EPSA and Peptides**

Once EPSA was established as a viable polarity monitoring tool for small molecules, its remit was expanded to peptides, notably cyclic peptides, which have generally demonstrated insufficient permeability to be used as oral drugs. Guiding the design of such peptides without a reliable permeability monitoring method is a challenge. Since for peptides, the main obstacle to permeability is polarity, the EPSA experiment was a perfect fit.
Goetz et al. [88] have reported a project team design strategy illustrating the suitability of EPSA as a surrogate for cyclic peptide permeability. While working towards improving permeability of their lead compounds, the team was faced with a flat structure-permeability relationship (no significant improvement in permeability as measured by RRCK cells was detected with any modification). Since the lead compound had an EPSA value in the high 160s, reducing measured polarity became the surrogate objective. EPSA measurements drove each new design. Plans that increased EPSA were abandoned in favor of those decreasing EPSA. With each design cycle it was observed that a gradual decrease in polarity was obtained but without any significant improvement in permeability as measured by RRCK, until a threshold in EPSA was reached (EPSA < 100) upon which permeability became measurable and increased to reach acceptable levels (EPSA < 80).

At Pfizer, and across multiple academic collaborations [38,89-91], project teams routinely use EPSA data to inform peptide design as a predictor of permeability and as an indicator of IMHB patterns in cyclic peptides. Furthermore, EPSA is now used in a prospective way after the development of an in-house computational model [92]. It was built using the following procedure. Compounds undergo a thorough conformational analysis (Macro model) followed by 3D descriptors calculation (VolSurf+) based on molecular interaction fields (MIFs). A partial least squares (PLS) regression algorithm is then used to establish a quantitative relationship between the matrix of descriptors and the EPSA endpoint. Implementation of the computational EPSA prediction model, combined with other physicochemical properties, has enabled virtual compound design, selection of better candidates for synthesis and has accelerated efficiency of design cycles.

Figure 10. Illustration of polarity based design cycle enabled by the Pfizer computational EPSA prediction model

In the particular example depicted in Figure 10, the computational EPSA prediction model was successfully implemented by the project team, first designing virtual compound libraries, and then testing their EPSA predictions, as a permeability surrogate. Only compounds with predicted EPSA values < 95 were prioritized for synthesis. Compounds with reasonable potency and measured EPSA < 100 were prioritized for cell-based passive permeability evaluation and HLM stability assessment. Learnings from this data the next design cycle was refined [93].

The EPSA technique and its application are not reserved or limited to Pfizer and its collaborators, indeed, EPSA has been implemented across the pharmaceutical industry, i.e. at AbbVie [94], Novartis [95], and Merck [96], EPSA services are available through an established analytical CRO [97]. Additionally, Peter Wipf in his January 2017 editorial [98] selected EPSA as one of the 6 technical innovations published in ACS Medicinal Chemistry Letters over the past 3 years that facilitate the practice of medicinal chemistry and are likely to become essential items in the medicinal chemist’s toolbox.
Conclusions

Across the pharmaceutical industry, each R&D department has its own philosophy. Based on historical approaches, resource allocations and leadership advocacy, drug design strategies essentially follow the same guidelines with regards to lead optimization: exquisite potency along with favorable ADMET profiles in the least amount of time and at the lowest possible cost. Guiding those design strategies through chromatography based physicochemical properties (i.e. Elog D and EPSA) is one of the more successful approaches developed in order to accelerate the process.

Acknowledgements: Laurence Philippe for her continuous support and mentorship, Patrick Mullins and Jason Ramsay for critical editing and Justin Montgomery for Figures 6 and 7.

References


Leveraging chromatography physchem properties in drug design


