

Original scientific paper

## Defining desirable natural product derived anticancer drug space: optimization of molecular physicochemical properties and ADMET attributes

Deepika Singh

Medicinal Chemistry Division, Central Institute of Medicinal and Aromatic Plants, PO CIMAP, Lucknow 226015, India

Corresponding author: E-mail: [deepika.sh25@yahoo.com](mailto:deepika.sh25@yahoo.com).

Received: March 25, 2016; Revised: May 14, 2016; Published: June 29, 2016

---

### Abstract

*As part of our endeavor to enhance survival of natural product derived drug candidates and to guide the medicinal chemist to design higher probability space for success in the anti cancer drug development area, we embarked on a detailed study of the property space for a collection of natural product derived anti cancer molecules. We carried out a comprehensive analysis of properties for 24 natural products derived anti cancer drugs including clinical development candidates and a set of 27 natural products derived anti cancer lead compounds. In particular, we focused on understanding the interplay among eight physicochemical properties including like partition coefficient (log P), distribution coefficient at pH=7.4 (log D), topological polar surface area (TPSA), molecular weight (MW), aqueous solubility (log S), number of hydrogen bond acceptors (HBA), number of hydrogen bond donors (HBD) and number of rotatable bonds ( $n_{Rot}$ ) crucial for drug design and relationships between physicochemical properties, ADME (absorption, distribution, metabolism, and elimination) attributes, and in silico toxicity profile for these two sets of compounds. This analysis provides guidance for the chemist to modify the existing natural product scaffold or designing of new anti cancer molecules in a property space with increased probability of success and may lead to the identification of druglike candidates with favorable safety profiles that can successfully test hypotheses in the clinic.*

### Keywords

Anticancer, ADMET, Natural Products, Physicochemical property, pharmacokinetic-pharmacodynamic

---

### Introduction

Cancer is one of the major disease causes of mortality worldwide and the numbers of cancer cases are increasing gradually [1]. Cancer is a main public health burden in both developed and developing countries and affects the lives of millions of people. Cancer is an abnormal growth of cells in the body, which underlies a collection of multiple genetic abnormalities through a multistep, mutagenic process. Cancer cells usually invade and destroy normal cells in the body. Factors responsible for cancers includes genetic predisposition, smoking, incorrect diet, infectious diseases and environmental factors. American Cancer Society has predicted ~27 million newly diagnosed individuals and ~17 million cancer related deaths globally by 2050 [2]. The key problem to cancer treatment is the recurrence of tumor and the side effects of chemotherapy drugs. Hence, there is a potential demand to develop new and efficient anti-cancer drugs

[3]. Natural products have received increasing attention in the past 30 years for the discovery of novel cancer preventive and therapeutic agents [4]. Natural products have been used for centuries for the treatment of several ailments. There are many basic ancient medicinal systems derived from dietary sources. Nature has provided a plenty of natural products with potential anti-cancer activity in the last few decades. Since 1940, approximately 175 small molecules have been approved as anti-cancer agents, of these, 48.6 % were a natural product or derivative [5].

Currently, pharmaceutical industry faces large attrition rates of preclinical and clinical candidates due to toxicity or lag of optimal pharmacokinetics properties, resulting in high costs and increased timelines for the drug discovery process [6]. Lead structures are compounds that typically exhibit suboptimal target binding affinity. Pharmacological studies have shown that there is a difference exists between leads and drugs [7]. The present study is an approach to establish the difference between some selected potent anticancer natural compounds (leads) and FDA approved natural product derived anti cancer drugs considering the distribution of physicochemical, ADME (absorption, distribution, metabolism, and elimination) attributes and *in silico* toxicity endpoints. This data was examined with the goal of identifying trends and defining a set of property values that would best define the anticancer drug space associated with a higher probability of clinical success. Several critical physicochemical properties of compounds like  $\log P$ ,  $\log D$ , TPSA,  $M_w$ ,  $\log S$ , HBA, HBD and  $n_{\text{Rot}}$  proposed by various research groups should be considered for compounds with oral drug delivery as a concern [8].

The information obtained from this analysis could, in turn, be utilized to design anticancer drug molecules with optimum bioavailability and less or no toxicity based on the alignment of a set of key properties. The multi-parameter optimization (MPO) approach is very popular for providing guidance on how to design preferred molecules to reduce the attrition rate and increase the probability of prospectively designing molecules that survive preclinical safety studies and that possess optimal pharmacokinetic and pharmacodynamic properties to test hypotheses in the clinic [9]. Tremendous progress has been made in recent years in terms of enabling the development of robust pharmacokinetic-pharmacodynamic (PK/PD) relationships for anticancer agents as well as in understanding how these relationships are influenced by molecular physicochemical properties.

Key physicochemical properties related to a drug like molecules have been described previously by various research groups [10]. The most important and well-known rule of five (RO5) was given by Lipinski et al. in 1997 based on the database of clinical candidates that had reached phase II trials or further [11]. RO5 provided the end points for four crucial physicochemical properties that described 90 % of orally active drugs: (a) molecular weight,  $MW < 500$  Da; (b) calculate of 1-octanol/water partition coefficient,  $\text{ClogP} < 5$ ; (c) number of hydrogen-bond donors, (OH plus NH count)  $< 5$ ; and (d) number of hydrogen-bond acceptors, (O plus N atoms)  $< 10$ . These four physicochemical properties and their endpoints are associated with acceptable aqueous solubility and intestinal permeability, the important first step of oral bioavailability [11]. After the Lipinski's RO5, various other ways of predicting the drug-like space for rational design purposes have been introduced by other people. Veber et al. 2002, showed that molecular weight cutoff at 500 Da does not itself significantly separate compounds with poor oral bioavailability from those with acceptable values based on the oral bioavailability measurements in rats for GlaxoSmithKline database of almost 1100 drug candidates [12]. He suggested that compounds having two criteria of (1) number of rotatable bonds ( $n_{\text{Rot}} \leq 10$ ), (2)  $\text{TPSA} \leq 140 \text{ \AA}^2$  will have a high probability of good oral bioavailability. Another effective range of physicochemical properties provided by Ghose et al. 1999, based on Comprehensive Medicinal Chemistry (CMC) database can be used in the design of drug-like

combinatorial libraries [13].

To go beyond the properties associated with the RO5 and other drug-like filters, we became interested in developing a holistic understanding of physicochemical property space for anticancer molecules by carrying out a thorough analysis of properties for natural product derived anticancer drugs and a set of natural product lead anticancer molecules, as most of the anticancer drugs have been derived from the natural products [14]. Herein, we present our efforts to develop a prospective MPO design tool for anti cancer molecules that does not focus on hard cutoffs or single end points but utilizes the eight essential physicochemical properties to prospectively align drug like attributes such as high permeability, low P-gp efflux liability, low metabolic clearance, and high safety into one molecule. In order to increase the flexibility in design and probability of identifying candidates with optimal pharmacokinetic and safety profile, we should not use hard cutoffs or focus on a single property, as it may restrict design space and may not align multiple attributes at once.

## Experimental

### *ADMET property calculations*

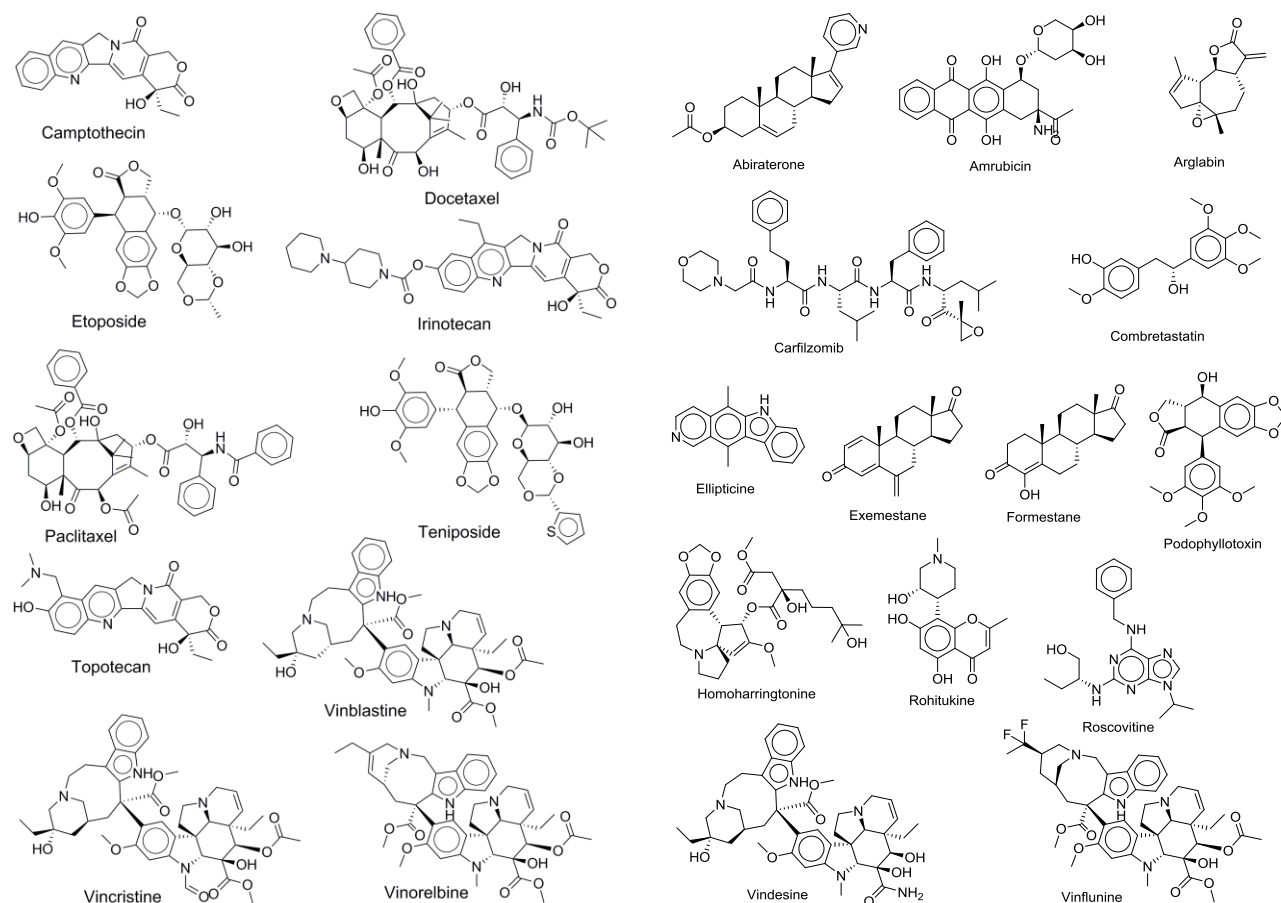
Poor pharmacokinetic properties are one of the main reasons for terminating the development of drug candidates. Computed physicochemical properties associated with compounds that have good oral bioavailability, less or no toxicity and optimum values of physicochemical properties are key parameters for the anti cancer drug discovery, and we need compounds with good pharmacokinetic properties [15, 16]. The drug set used in this study includes 24 natural product derived anti cancer drugs and structure of these drugs are mentioned in **Figure 1**. To the best of our knowledge, all compounds in the drug set could be used as oral agents [17]. The lead candidates included in our analysis consisted of 27 natural products derived anti cancer compounds that collected from the literature belong to the several chemical classes as shown in **Figure 2** [18]. A complete list of the drugs and lead candidates used in the analysis appears in **Table 1**.

ADMET related physicochemical properties for 24 natural product derived anticancer drugs including clinical development candidates and 27 natural product lead anticancer compounds were predicted using OSIRIS Datawarrior Version 4.2.2 software on a Windows XP operating system [19]. *DataWarrior* is able to calculate physicochemical properties, lead- or drug-likeness related parameters, ligand efficiencies, various atom and ring counts, molecular shape, flexibility and complexity as well as indications for potential toxicity.

After calculating properties, these are automatically added as new columns to the data table. Chemical structures for 24 Natural Product derived anticancer drugs including clinical development candidates and 27 Natural product lead compounds were downloaded and saved individually in 3D SDF format from pubchem ([www.pubchem.org](http://www.pubchem.org)). *DataWarrior* is unable to optimize structures; therefore, geometry optimization of the molecules was performed in Avogadro software prior to the prediction of physicochemical properties [20]. *DataWarrior software* calculates the descriptors as inputs to independent mathematical models to estimate a range of ADMET values at relevant pH 7.4.

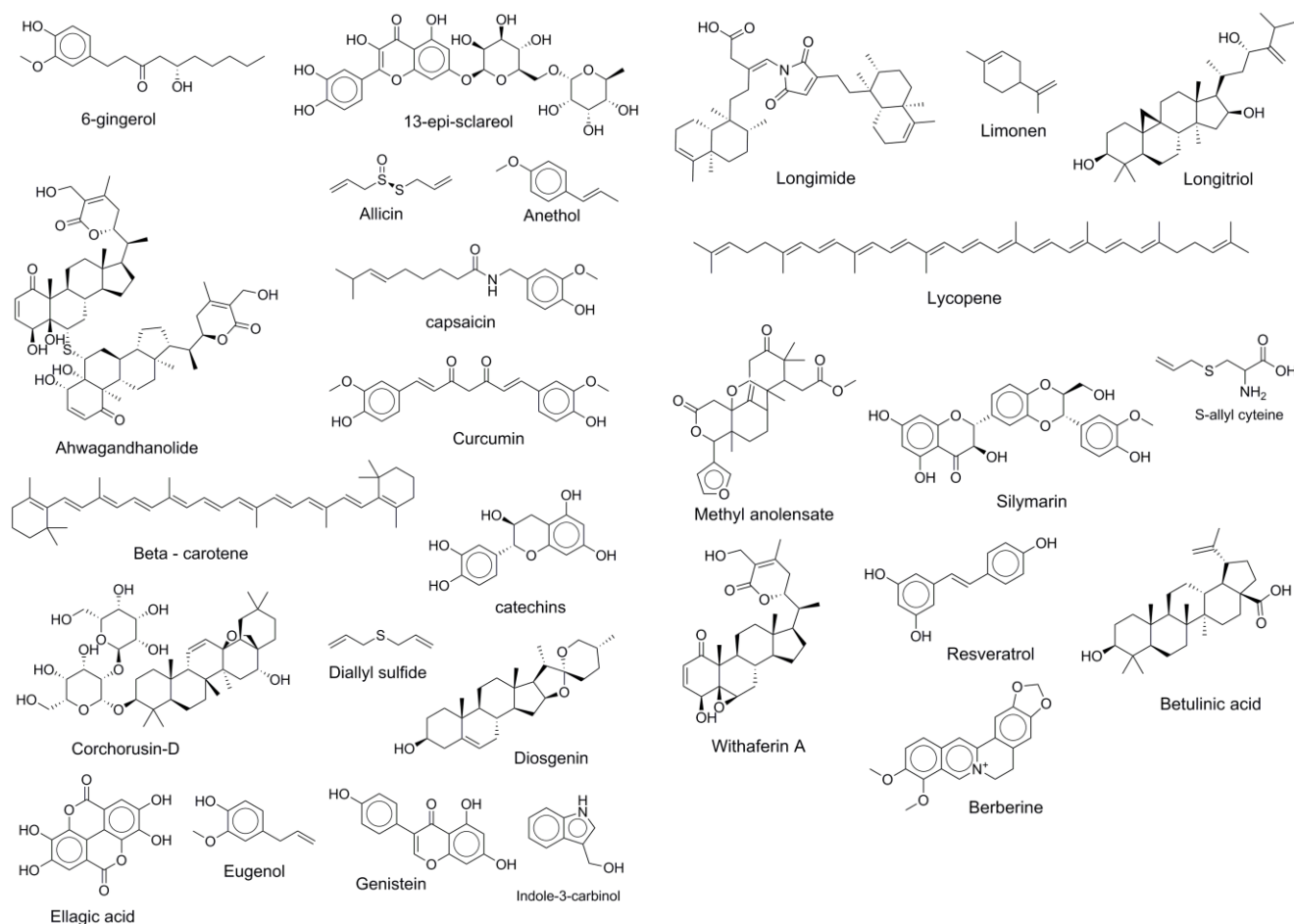
Physicochemical properties of interest included predicted lipophilicity ( $\log P$ ), predicted aqueous solubility ( $\log S$ ), topological polar surface area (TPSA), molecular weight (MW), hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), and number of rotatable bonds ( $n_{\text{Rot}}$ ). Specific ADME properties of interest included predicted distribution coefficient at pH=7.4 ( $\log D$ ) (value predicted by ACD/Labs,

www.chemspider.com), predicted aqueous solubility ( $\log S$ ), quantitatively predicted apparent permeability ( $P_{app}$  Caco-2 cell), predicted effective permeability ( $P_{eff}$ ), and predicted Human Intestinal Absorption (HIA). In order to evaluate the distribution of drugs and leads, we considered two important parameters including a fraction of unbound to plasma proteins ( $F_u$ ), and volume of distribution ( $V_{Dss}$ ), a requirement of all clinical candidate through recently developed online ADMET calculation tool pkSCM (<http://bleoberis.bioc.cam.ac.uk/pkscm/>) [21].



**Figure 1.** Chemical structures of 24 natural products derived anti cancer drugs.

To determine excretion routes, natural products anticancer drugs and leads we quantitatively predicted the total clearance and qualitatively predicted renal OCT2 substrate. The safety profile of compounds is one of the most common factors in drug attrition (1). As part of our analysis of properties for natural products anticancer drugs and leads, we assess some of the major toxicity endpoints. We generated in silico data to assess potential for the following safety risks: drug-drug interactions (CYP inhibitions) including CYP3A4, CYP2C9, and CYP2D6, hERG liability (inhibition of dofetilide binding), predicted LD50, predicted hepatotoxicity, predicted skin sensitization, cellular toxicity through pkSCM tool and mutagenicity, tumorigenicity and irritant effects through *DataWarrior software*. To access the likelihood of binding to transporter permeability-glycoprotein (P-gp), we used Pgp\_Substrate model. We also calculated the three most crucial drug-likeness filters including Lipinski, Ghose, and Veber rules as well as the quantitative estimate of drug-likeness (QED) with the Drug Likeness Tool (DruLiTo) software ([http://www.niper.gov.in/pi\\_dev\\_tools/DruLiToWeb/DruLiTo\\_index.html](http://www.niper.gov.in/pi_dev_tools/DruLiToWeb/DruLiTo_index.html)).



**Figure 2.** Chemical structures of 27 natural products derived anti cancer lead molecules.

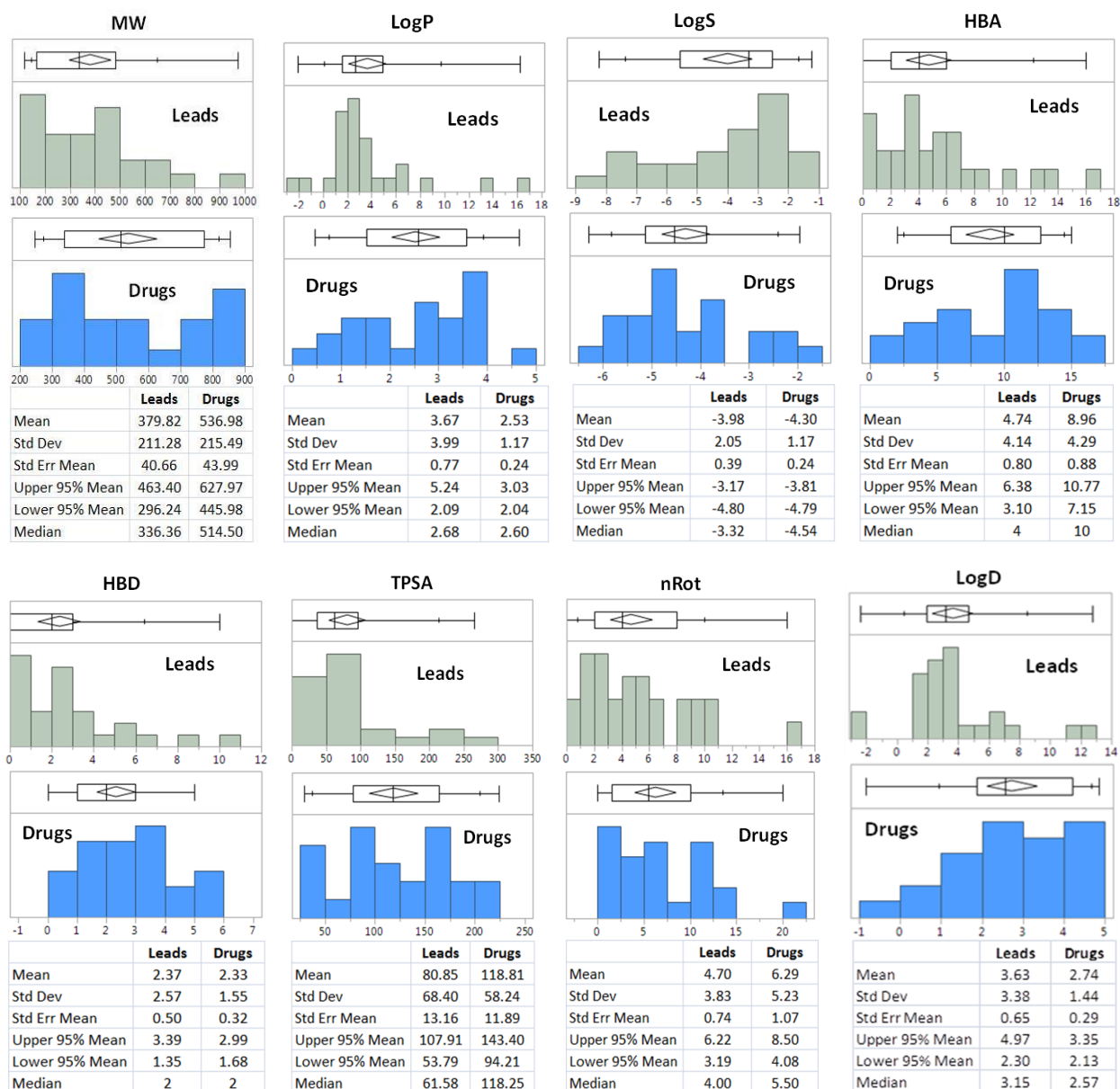
## Results and Discussion

### *Optimum physicochemical property space for anticancer molecules*

The 24 natural product derived anticancer drugs including clinical development candidates and 27 natural product lead anticancer compounds were evaluated against a set of eight calculated fundamental physicochemical properties that have gained wide acceptance as key parameters for drug design and development: (a) lipophilicity, calculated partition coefficient ( $\log P$ ); (b) distribution coefficient at  $\text{pH}=7.4$  ( $\log D$ ); (c) molecular weight ( $M_w$ ); (d) topological polar surface area (TPSA); (e) number of hydrogen bond donors (HBD); (f) hydrogen bond acceptor (HBA), (g) number of rotatable bonds ( $n_{\text{Rot}}$ ) and predicted aqueous solubility ( $\log S$ ) [11, 22]. The calculated physicochemical properties value for the drugs and leads are mentioned in **Table 1**.

Physicochemical property space as captured by these eight parameters was quite broad (**Figure 3**). The  $M_w$  values for the drugs varied from 246 to 853 with a median  $M_w$  value of 514, while  $M_w$  range for leads is much broad and varied from 114 to 975 with a median value of 336. The  $\log P$  value of the drugs varied from 0.46 to 4.67 with a median  $\log P$  value of 2.6. Molecules in the lead set having the  $\log P$  value from -2 to 16, which is quite broad with a median of 2.68. There was no significant difference in the median  $\log P$  values between the two sets, although the drug set had a lower span of  $\log P$  values. Low hydrophilicity (e.g. high  $\log P$ ) values can cause poor absorption or permeation. Our analysis suggests that for anticancer drugs, there may be a need to design compounds with further reduced  $\log P$  or  $M_w$  to better match the corresponding properties in the drug set. Polarity, as described by polar surface area (TPSA), ranged from

about  $29 \text{ \AA}^2$  to  $224 \text{ \AA}^2$  with a median value of  $118.25 \text{ \AA}^2$  for the drug set, while the polar surface area (TPSA) for lead molecules span from  $0 \text{ \AA}^2$  to  $266 \text{ \AA}^2$ . There was a significant difference in the TPSA values between lead candidates and drugs, almost  $\sim 60\%$  of the lead candidates having  $\text{TPSA} < 80 \text{ \AA}^2$ , oppositely  $75\%$  drugs have  $\text{TPSA} \geq 80 \text{ \AA}^2$ , which clearly suggests there is a huge need for the optimization of TPSA of lead candidates. The drugs and the lead candidates had a minimal number of hydrogen bond donors (HBD), with the median value of 2 for both sets. Almost  $\sim 78\%$  of lead candidates and  $83\%$  of drugs having the HBD value  $\leq 3$ . Lipinski's RO5 identified HBD as a critical component of the drug property analysis and targets a HBD count (OH plus NH count) of  $< 5$ . Based on the number of HBD associated with anticancer drugs and lead candidates, optimization of HBD to  $\leq 3$  may increase the probability of identifying better anticancer molecules.



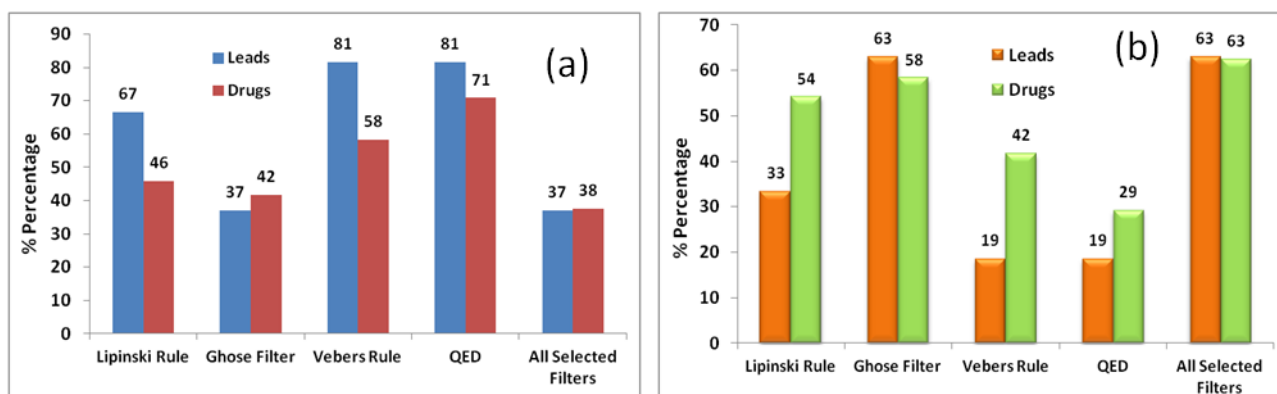
**Figure 3.** Physicochemical property distribution and statistics of drugs and lead candidates are shown for  $M_w$ ,  $\log P$ ,  $\log S$ , HBA, HBD, TPSA,  $n_{\text{Rot}}$  and  $\log D$ .

Hydrogen bond acceptor is another valuable physicochemical property in RO5, drugs and the lead candidates having the median value of 10 and 4 respectively. Only 11 out of 24 drugs in this study are following the  $\text{HBA} < 10$  rule of Lipinski, on the other hand, most of the lead candidates  $\sim 89\%$  having the  $\text{HBA} < 10$ , so anticancer lead candidates well following this RO5 compare to anticancer drugs. Aqueous

solubility is another very important parameter for the oral bioavailability. The recommended range for a molecule to be good oral bioavailable is  $(-6 \leq \log S \leq 0.5)$ , here all drugs are following this rule and almost ~95 % of lead candidates are also falling within the recommended range. The majority of this increased TPSA should originate from an increased number of HBA, as HBD must be strictly controlled at  $\leq 6$  to avoid reducing oral bioavailability.

Log  $D$  provides a better measurement of lipophilicity for ionizable compounds; we know that hydrophilic molecules have higher solubility, but are less equipped to readily cross the cell membrane. Hence, a compound is considered to be hydrophilic, if  $\log D < 0$ , lipophilic if  $\log D > 0$ , and molecule excessive lipophilic by molecules with  $\log D \geq 3.5$ . Nearly ~85 % of the lead candidates having a  $\log D$  value from 1 to 8 ( $1 \leq \log D \leq 8$ ), with a median of 3.15, which suggest maximum lead candidates are hydrophobic in nature. As expected for anticancer drugs, a similar but narrower range existed for  $\log D$ , which varied from -0.86 to 4.88 with a median value of 2.57.

In order to accurately access the differences between anticancer drugs and leads, four drug-like indices were utilized for comparison; Lipinski's RO5, Ghose filter, Vebers's selective criteria for oral bioavailable drugs, and QED by DruLiTo. The detailed results in percentage (%) of lead and drug molecules are following and violating the above mentioned most promising oral bioavailable rules are mentioned in the **Figure 4(a)** and **Figure 4(b)** respectively.



**Figure 4.** Bar graph for percentage (%) of lead and drug molecules are (a) following, and (b) violating the Lipinski's RO5, Ghose filter, Vebers's rule, QED, and all selected filters.

The inspection of the bar graph shown in Figure 4(a) reveals that leads are following most of the bioavailable rules greater than the drugs except the Ghose filter. Drugs and leads are showing almost equal percentage of molecules ~38 % following the all selected filters, which is obtained by considering all 4 bioavailable, filters (Lipinski's RO5, Ghose filter, Vebers's rule, and QED) together. Similarly, greater percentage of drugs are violating the bioavailable rules except the Ghose filter. While the natural product derived anticancer drug space as defined by  $M_w$ ,  $\log P$ ,  $\log S$ , HBA, HBD, TPSA,  $n_{\text{Rot}}$  and  $\log D$ , is broad. Hence, our emphasis has been on defining physicochemical property rules for compounds to reduce attrition and increase the likelihood of candidates at various stages of anticancer drug development based on our analysis as well as earlier published various oral bioavailability rules by different research groups.

Our analysis shows the optimal property ranges (covering almost ~80 % or more of the anticancer drugs) used to select drug-like anticancer molecule for these properties are  $200 < M_w \leq 800$  Da,  $1 < \log P \leq 5$ ,  $-6 \leq \log S \leq -1$ ,  $5 \leq \text{HBA} \leq 13$ ,  $1 \leq \text{HBD} \leq 5$ ,  $50 \leq \text{TPSA} \leq 180 \text{ \AA}^2$ ,  $0 \leq n_{\text{Rot}} \leq 10$ ,  $\log D = 2.8$ , which may be very

Table 1. Important computed physicochemical properties for anticancer lead candidates and drugs.

| Lead Candidates   | $M_w$ | $\log P$ | $\log S$ | HBA | HBD | TPSA | $n_{Rot}$ | $\log D$ |
|-------------------|-------|----------|----------|-----|-----|------|-----------|----------|
| 13-epi-sclareol   | 611   | -1.41    | -2.58    | 16  | 10  | 266  | 6         | -2.42    |
| 6-gingerol        | 294   | 3.56     | -3.25    | 4   | 2   | 67   | 10        | 3.64     |
| Ahwagandhanolide  | 975   | 5.36     | -8.25    | 12  | 6   | 233  | 8         | 6.75     |
| Allicin           | 162   | 1.84     | -1.22    | 1   | 0   | 62   | 5         | 2.01     |
| Anethol           | 148   | 2.68     | -2.54    | 1   | 0   | 9    | 2         | 2.77     |
| Berberine         | 336   | 0.52     | -4.67    | 5   | 0   | 41   | 2         | 3.96     |
| Beta carotene     | 537   | 13.87    | -7.33    | 0   | 0   | 0    | 10        | 12.00    |
| Betulinic acid    | 457   | 6.37     | -6.28    | 3   | 2   | 58   | 2         | 6.52     |
| Capsaicin         | 305   | 3.80     | -3.32    | 4   | 2   | 59   | 9         | 3.90     |
| Catechins         | 290   | 1.51     | -1.76    | 6   | 5   | 110  | 1         | 1.90     |
| Corchorusin-D     | 781   | 1.54     | -5.21    | 13  | 8   | 208  | 6         | 1.31     |
| Curcumin          | 368   | 2.95     | -3.62    | 6   | 2   | 93   | 8         | 3.55     |
| Diallyl sulfide   | 114   | 2.16     | -2.01    | 0   | 0   | 25   | 4         | 2.03     |
| Diosgenin         | 415   | 4.88     | -5.58    | 3   | 1   | 39   | 0         | 4.63     |
| Ellagic acid      | 302   | 1.28     | -3.29    | 8   | 4   | 134  | 0         | 1.29     |
| Eugenol           | 164   | 2.27     | -2.05    | 2   | 1   | 29   | 3         | 2.58     |
| Genistein         | 270   | 1.63     | -2.73    | 5   | 3   | 87   | 1         | 1.12     |
| Indole-3-carbinol | 147   | 1.10     | -2.03    | 2   | 2   | 36   | 1         | 1.52     |
| Limonen           | 136   | 3.36     | -2.54    | 0   | 0   | 0    | 1         | 3.50     |
| <b>Drugs</b>      |       |          |          |     |     |      |           |          |
| Abiraterone       | 392   | 4.67     | -4.75    | 3   | 0   | 39   | 3         | 4.72     |
| Amrubicin         | 483   | 0.84     | -4.62    | 10  | 5   | 177  | 3         | 1.59     |
| Arglabin          | 246   | 1.48     | -2.78    | 3   | 0   | 39   | 0         | 2.13     |
| Camptothecin      | 348   | 1.18     | -2.74    | 6   | 1   | 80   | 1         | 2.02     |
| Carfilzomib       | 720   | 2.51     | -4.57    | 12  | 4   | 158  | 20        | 4.43     |
| Combretastatin    | 334   | 2.17     | -2.48    | 6   | 2   | 77   | 7         | 1.84     |
| Docetaxel         | 808   | 2.61     | -5.81    | 15  | 5   | 224  | 13        | 2.59     |
| Ellipticine       | 246   | 3.90     | -5.14    | 2   | 1   | 29   | 0         | 3.14     |
| Etoposide         | 589   | 0.67     | -3.95    | 13  | 3   | 161  | 5         | 0.94     |
| Exemestane        | 296   | 3.61     | -3.95    | 2   | 0   | 34   | 0         | 3.01     |
| Formestane        | 302   | 3.14     | -4.04    | 3   | 1   | 54   | 0         | 2.55     |
| Homoharringtonine | 546   | 3.65     | -4.38    | 10  | 2   | 124  | 11        | 1.34     |
| Irinotecan        | 587   | 3.56     | -4.50    | 10  | 1   | 113  | 5         | 3.25     |
| Paclitaxel        | 854   | 3.19     | -6.29    | 15  | 4   | 221  | 14        | 3.06     |
| Podophyllotoxin   | 414   | 1.79     | -3.84    | 8   | 1   | 93   | 4         | 1.95     |
| Rohitukine        | 305   | 1.62     | -2.30    | 6   | 3   | 90   | 1         | -0.86    |
| Roscovitine       | 354   | 2.58     | -3.93    | 7   | 3   | 88   | 8         | 4.45     |
| Teniposide        | 657   | 1.35     | -4.85    | 13  | 3   | 189  | 6         | 2.52     |
| Topotecan         | 421   | 0.46     | -1.96    | 8   | 2   | 103  | 3         | 0.93     |
| Vinblastine       | 811   | 3.30     | -5.08    | 13  | 3   | 154  | 10        | 4.88     |
| Vincristine       | 825   | 2.98     | -5.53    | 14  | 3   | 171  | 10        | 3.97     |
| Vindesine         | 754   | 1.99     | -4.62    | 12  | 5   | 165  | 7         | 2.38     |
| Vinflunine        | 817   | 3.95     | -5.87    | 12  | 2   | 134  | 10        | 4.30     |
| Vinorelbine       | 779   | 3.59     | -5.15    | 12  | 2   | 134  | 10        | 4.66     |

helpful in prospective design of anticancer molecules from natural products or identification of lead candidates from natural products that can successfully progress to the clinic and becomes better anticancer drug. For some of the physicochemical property, specifically  $M_w$ , HBA and  $n_{Rot}$  lead molecules



are showing better optimal range compare to the drug candidates according to the RO5, which are quite significant for shaping the ADMET properties of potential anticancer drug candidates.

#### *Profiling ADME space of anticancer molecules*

Potential therapeutic compounds are useless without having a good ADMET profile, and thus, it is essential to find the source of such diminished potency for developing a drug. Significant advances in the development of HT in vitro ADME assays have enabled computational scientist to make robust computational models to the earlier assessment of potential liabilities (low permeability, susceptibility to efflux transporters, etc.) associated with new potential lead compounds. In order to gain a better perspective on the ADME properties of drugs and lead candidates, we evaluated the in silico profiling of these compounds to assess Caco2 cell permeability, human intestinal absorption, and P-gp efflux liability [23]. We can classify the permeability of a molecule as low, or high based on the predictive model and its relative range of  $\log P_{app}$  (in  $10^{-6}$  cm/s) rates are, as follows:  $\log P_{app} > 0.9$ , considered to be high permeability, while  $\log P_{app} < 0.9$ , considered to be low permeability of the molecule. The Caco2 cell permeability values for lead candidates and drugs are mentioned in **Table 2**.

In the Caco2 cell permeability prediction, 70 % of the lead candidates show high  $\log P_{app}$  values; surprisingly the drugs had a lower percentage (30 %) with high  $\log P_{app}$  values. A similar discrepancy was observed when we assessed P-gp efflux liabilities for drugs and lead candidates. The P-gp efflux liability was assessed utilizing preADMET's [<https://preadmet.bmdrc.kr/>] P-gp\_Substrate model. Prediction of the likelihood of Pgp efflux shows that all drugs and 60% of the lead candidates are considered to be P-gp efflux substrates; the predicted values from the Pgp\_Substrate model for both drugs and lead candidates dataset are mentioned in **Table 2**. An optimal clinical candidate could be achieved if it is possessed both high  $\log P_{app}$  and low P-gp efflux liability.

The intestine is the primary site of absorption for the orally administered drugs; hence, we predicted the percentage (%) of human intestinal absorption of the drugs and lead candidates. Systemic oral dosage requires compound properties that allow for dissolution and stability in the gastrointestinal (GI) tract, including the acidic environment of the stomach (pH 1–2 in fasted state, 3–7 in fed state) and the close to neutral environment (pH 4.4–6.6) of the small intestine [28]. The % human intestinal absorption of the drugs and lead candidates was assessed by using pkCSM web server model [21]. All the drugs are showing good predicted human intestinal absorption  $> 60$  %; while 93 % of the lead candidates predicted  $>70$  % human intestinal absorption. The predicted % human intestinal absorption values for both drugs and lead candidates dataset are mentioned in **Table 2**.

Many of the drugs in plasma will exist in equilibrium between an unbound state and a bound to serum proteins or whole blood proteins at various affinities. It is commonly accepted that only unbound drug may interact with anticipated molecular targets [24]; hence, the efficacy of a drug might affect by the degree to which it binds whole blood proteins. We have predicted the fraction unbound of both drugs and lead candidates through the predictive model of pkCSM, which was built using the measured free proportion of 552 compounds in human blood ( $F_u$ ). We also evaluated the steady-state volume of distribution ( $V_{Dss}$ ) of drugs and lead candidates; another important parameter, which suggests the total dose of a drug would be required to be uniformly distributed to provide the similar concentration as in blood plasma. The values of predicted fraction unbound ( $F_u$ ) and  $V_{Dss}$  values for both drugs and lead candidates dataset are mentioned in **Table 2**. Evaluation of individual ADME properties ( $P_{app}$ , P-gp,  $F_u$ ) suggested that to increase the probability of success, the design should focus on optimizing all properties of a molecule.

Table 2. Computed ADME properties for anticancer lead candidates and drugs.

| Lead candidates   | Caco2 permeability<br>(log $P_{app}$ in $10^{-6}$ cm/s) | Intestinal absorption<br>(human)<br>(% Absorbed) | $V_{Dss}$<br>(human)<br>(log L/kg) | Fraction<br>unbound<br>(human) | P-gp<br>Substrate<br>(Yes/No) |
|-------------------|---|--|------------------------------------|--------------------------------|-------------------------------|
| 13-epi-sclareol   | -0.791  | 28.495   | -1.597                             | 0.419                          | Yes                           |
| 6-gingerol        | 0.959   | 93.293   | -0.061                             | 0.248                          | Yes                           |
| Ahwagandhanolide  | 0.197   | 82.511   | -0.244                             | 0.112                          | Yes                           |
| Allicin           | 1.366   | 98.525   | 0.084                              | 0.502                          | No                            |
| Anethol           | 1.391   | 98.814   | 0.066                              | 0.322                          | No                            |
| Berberine         | 1.732   | 100  | 0.134                              | 0.196                          | No                            |
| Beta carotene     | 1.397   | 91.234   | 1.319                              | 0                              | No                            |
| Betulinic acid    | 1.294   | 91.58  | 0.418                              | 0                              | Yes                           |
| Capsaicin         | 1.429   | 92.542   | 0.084                              | 0.172                          | Yes                           |
| Catechins         | -0.38   | 71.562   | -0.79                              | 0.326                          | Yes                           |
| Corchorusin-D     | -0.193  | 55.926   | -0.401                             | 0.318                          | Yes                           |
| Curcumin          | 0.556   | 81.7   | -0.677                             | 0.103                          | Yes                           |
| Diallyl sulfide   | 1.252   | 98.6   | 0.282                              | 0.487                          | No                            |
| Diosgenin         | 1.245   | 96.426   | 0.931                              | 0.09                           | Yes                           |
| Ellagic acid      | -0.273  | 80.032   | -1.214                             | 0.27                           | Yes                           |
| Eugenol           | 1.48  | 96.594   | -0.071                             | 0.375                          | No                            |
| Genistein         | 1.07  | 90.14  | -0.836                             | 0.192                          | Yes                           |
| Indole-3-carbinol | 1.31  | 94.062   | -0.027                             | 0.406                          | No                            |
| Limonen           | 1.248   | 98.048   | 0.503                              | 0.422                          | No                            |
| Longimide         | 1.02  | 89.104   | 0.525                              | 0                              | Yes                           |
| Longitriol        | 1.277   | 90.662   | 0.857                              | 0                              | Yes                           |
| Lycopene          | 1.444   | 90.326   | 1.115                              | 0                              | No                            |
| Methyl anolensate | 1.241   | 100  | 0.04                               | 0.14                           | Yes                           |
| Resveratrol       | 1.294   | 89.885   | -0.403                             | 0.198                          | No                            |
| S-allyl cyteine   | 0.375   | 88.593   | 0.026                              | 0.711                          | No                            |
| Silymarin         | -0.363  | 74.315   | -1.327                             | 0.142                          | Yes                           |
| Withaferin A      | 1.475   | 95.986   | 0.337                              | 0.171                          | Yes                           |
| <b>Drugs</b>      |   |  |                                    |                                |                               |
| Camptothecin      | 1.118   | 99.412   | -0.694                             | 0.195                          | Yes                           |
| Docetaxel         | -0.337  | 62.988   | -1.068                             | 0.138                          | Yes                           |
| Etoposide         | -0.32   | 82.748   | -1.185                             | 0.194                          | Yes                           |
| Irinotecan        | 0.986   | 93.659   | -0.036                             | 0.179                          | Yes                           |
| Paclitaxel        | -0.239  | 71.535   | -1.08                              | 0.105                          | Yes                           |
| Teniposide        | -0.255  | 90.274   | -1.267                             | 0.098                          | Yes                           |
| Topotecan         | 0.668   | 83.641   | -0.21                              | 0.31                           | Yes                           |
| Vinblastine       | 0.257   | 81.36  | 0.123                              | 0.241                          | Yes                           |
| Vincristine       | 0.186   | 78.558   | -0.034                             | 0.261                          | Yes                           |
| Vinorelbine       | 1.413   | 90.783   | 0.282                              | 0.193                          | Yes                           |
| Abiraterone       | 1.20  | 98.16  | 0.67                               | 0.05                           | Yes                           |
| Amrubicin         | -0.26   | 66.22  | -0.48                              | 0.38                           | Yes                           |
| Arglabin          | 1.59  | 100.00   | 0.49                               | 0.41                           | Yes                           |
| Carfilzomib       | 0.26  | 52.92  | 0.01                               | 0.23                           | Yes                           |
| Combretastatin    | 1.06  | 94.13  | -0.71                              | 0.21                           | Yes                           |
| Ellipticine       | 1.35  | 94.47  | -0.15                              | 0.09                           | Yes                           |
| Exemestane        | 1.52  | 100.00   | 0.69                               | 0.18                           | Yes                           |
| Formestane        | 1.39  | 96.29  | 0.50                               | 0.21                           | Yes                           |
| Homoharringtonine | 0.57  | 79.16  | 0.11                               | 0.35                           | Yes                           |
| Podophyllotoxin   | 0.78  | 95.31  | -0.82                              | 0.13                           | Yes                           |
| Rohitukine        | -0.22   | 73.76  | 0.18                               | 0.49                           | Yes                           |
| Roscovitine       | 1.01  | 90.21  | 1.97                               | 0.59                           | Yes                           |
| Vindesine         | 0.36  | 70.43  | 0.24                               | 0.29                           | Yes                           |
| Vinflunine        | 1.31  | 91.20  | 0.16                               | 0.19                           | Yes                           |

In order to understand the effect of physicochemical properties on ADME attributes of the molecules, we analyzed ADME attributes for both drugs and leads against all eight fundamental physicochemical properties. Interestingly, TPSA, HBD, and HBA showed good correlation with the Caco2 cell permeability, with the correlation coefficient of 0.83, 0.8, and 0.7 respectively for the drug molecules. Similarly, anticancer leads also showed slightly better correlation of physicochemical properties TPSA, HBD and HBA with Caco2 cell permeability, with the correlation coefficient of 0.83, 0.84, and 0.78, respectively. Although, all three physicochemical properties (TPSA, HBD, HBA) are inversely correlated with the Caco2 cell permeability suggesting that lipophilicity is important for molecule to have good Caco2 cell permeability and by optimizing the TPSA, HBD, and HBA cell permeability of a molecule can be enhanced. Human intestinal absorption also showed good correlation with the HBD, TPSA, and HBA physicochemical properties for both anticancer drug and lead molecules. The correlation coefficient of % human intestinal absorption with HBD, TPSA, and HBA was 0.88, 0.74, and 0.67 respectively for drugs and 0.91, 0.81, and 0.8 respectively for leads. This results clearly revealed the influence of physicochemical properties on ADME attributes of the molecule for oral bioavailability and the key physicochemical properties especially HBA, HBD, and TPSA need to be consider for further improvement in the ADME profile of natural product derived anticancer leads.

#### *Determining potential safety end points for anticancer molecules*

Early prediction of the safety endpoints through *in silico* techniques screening have become regular practice for both designing new molecule and screening of the large chemical databases within pharmaceutical industries [25]. As part of our analysis of properties for anticancer drugs and lead candidates, we determine the potential toxicity end points for through pkCSM. Most frequently measured end points to evaluate potential safety issues include inhibition of cytochrome P450 (CYPs) monooxygenase enzymes to determine potential for drug-drug interactions [26], inhibition of hERG potassium ion channel effects [27], lethal rat acute toxicity (LD50) and other crucial toxicity (AMES toxicity, skin sensitization, and hepatotoxicity). All toxicity predictions for both drugs and lead candidates are presented in **Table 3**.

We qualitatively predicted the inhibition of CYP2D6 and CYP3A4 through pkCSM, which suggests the potential for drugs and leads candidates to mediate drug-drug interactions (DDI) through perturbation of clearance mechanisms for other drug substances. Inhibition of the potassium hERG channel might cause in prolongation of the QT interval of cardiac rhythm, which has resulted in the withdrawal of many clinical candidates from the market [28]. Therefore, we have qualitatively predicted the potassium hERG channel inhibition potential of drugs and lead candidates. The data obtained suggests that all of these drugs and lead candidates are non-inhibitor of the hERG channel as mentioned in **Table 3**. Analysis of the inhibition data of CYP2D6 and CYP3A4 revealed that all the drugs are non-inhibitor of both CYP's and all lead candidates are non-inhibitor of CYP2D6, while 89 % of lead candidates are non-inhibitor of CYP3A4.

This data suggests that most of these drugs and lead candidates occupy desirable, low-risk space for DDI. Drug metabolism and the drug excretion also have a significant role in the drug design process. Issues related to metabolism have been commonly associated with the compounds failure in the clinical. Understanding the metabolic pathways of drugs would be very helpful in predicting drug-drug interactions (DDI), toxicities, and pharmacokinetics [29]. Many relationships between CYP family enzymes and *in silico* molecular properties have been available in the literature; the primary concern is inhibition of CYP3A4, which is correlated to increasing  $M_w$  and  $\log P$  [30]. This may lead to issues with clearance as well as drug-drug interactions. We have predicted the total clearance for drugs and lead candidates measured by the

proportionality constant and primarily occur as a combination of hepatic and renal clearance mentioned in **Table 3**.

Hughes *et al.* have done the most considerable work with regard to the impact of molecular properties on in vivo toxicity, led to the “3/75 rule”, derived from an analysis of exploratory or dose-finding toxicology studies of 245 compounds at Pfizer [31]. Key finding emerged from this analysis was that compounds with a  $\log P < 3$  and  $\text{TPSA} > 75 \text{ \AA}^2$  were 2.5 times more likely to be non-toxic at the same total exposure. Reversely, compounds those with high lipophilicity ( $\log P > 3$ ) and low polar surface area ( $\text{TPSA} < 75 \text{ \AA}^2$ ) had an increased risk of widespread toxicities in short-term animal studies. One crucial elucidation of these results would be that promising lipophilic compounds with small polar functionality likely to have an increased chance of toxicity. A similar study by AstraZeneca [32] on their compound failures showed a different profile, with the majority of failure happening with  $\text{TPSA} > 75 \text{ \AA}^2$  and  $\log P < 3$ . Though, attrition in the high- $\log P$ -low-TPSA space can readily be rationalized via consideration of promiscuity and interactions across a range of systems. Further Eli Lilly Company study of > 400 (Eli Lilly) compounds supported the influence of compound lipophilicity on toxicology in rat toxicological studies [33]. In this analysis, there was a three-fold enrichment in toxic compounds when  $\log P > 3$ , but TPSA had little or no influence. Clearly, the benefits of establishing a link between important clinically relevant end points and simple descriptors such as  $\log P$  and PSA (which can be easily calculated before synthesis) are highly attractive.

We also analyzed our natural product derived drugs and lead compounds predicted toxicity endpoints, to establish meaningful correlations between physicochemical properties and toxicity profile of compounds. Predicted toxicities of drugs and leads have been categorized as “yes” or “no”. Most of anticancer drugs in our dataset having the low lipophilicity ( $\log P < 3.5$ ) are showing the hepatotoxicity (e.g. camptothecin, rohitukine, carfilzomib, docetaxel, etc.), out of which some drugs also having  $\log P < 3.5$  and  $M_w > 700$  also showing hepatotoxicity (e.g. vinblastine, vincristine, vindesine, carfilzomib, and docetaxel, etc.). Similarly, four drugs, showed the AMES toxicity, also having the low lipophilicity ( $\log P < 2$ ). This link between the low lipophilicity of compounds and toxicity is in line with the results of Hughes *et al.* [31] and other research groups. Furthermore, some drugs showing the toxicity but no specific correlations with physicochemical properties was found, possibly the toxicity was a consequence of the primary drug target mechanism or of a specific off-target pharmacology. Examination of the relationship between physicochemical properties and other predicted toxicity end points, we found very good correlation for the drug molecules between the physicochemical properties and Oral Rat Chronic Toxicity (LOAEL). The correlation coefficient of LOAEL with  $M_w$ , HBA, HBD, TPSA, and  $n_{\text{Rot}}$  was 0.85, 0.84, 0.68, 0.84, and 0.77, respectively. All five physicochemical properties are positively correlated with the LOAEL, suggesting the need for the optimization of these physicochemical parameters to avoid the LOAEL toxicity. On the other hand, no correlation was observed between the LOAEL and physicochemical properties for lead molecules, which is evident from the **Table 3**, that drug molecules showed more toxicity endpoints compare to lead molecules. Hence, establishing meaningful correlations between the physicochemical properties and toxicity of natural product derived oral anticancer drugs and leads might be useful for future anticancer drug discovery.

1 **Table 3.** Computed safety end points for anticancer lead candidates and drugs

|                   | CYP2D6<br>inhibitor | CYP3A4<br>inhibitor | Total<br>Clearance | Renal<br>OCT2<br>substrate | AMES<br>toxicity | hERG<br>inhibitor | Oral Rat<br>Acute<br>Toxicity<br>(LD50) | Oral Rat<br>Chronic<br>Toxicity<br>(LOAEL) | Hepatotoxicity | Skin Sensitisation |
|-------------------|---------------------|---------------------|--------------------|----------------------------|------------------|-------------------|---|--|----------------|--------------------|
| 13-epi-sclareol   | No                  | No                  | 0.183              | No                         | No               | No                | 1.526                                   | 2.231                                      | No             | No                 |
| 6-gingerol        | No                  | No                  | 1.369              | No                         | No               | No                | 1.861                                   | 2.381                                      | No             | No                 |
| Ahwagandhanolide  | No                  | Yes                 | -0.711             | No                         | No               | No                | 2.011                                   | 3.113                                      | Yes            | No                 |
| Allicin           | No                  | No                  | 0.721              | No                         | No               | No                | 2.423                                   | 1.317                                      | No             | Yes                |
| Anethol           | No                  | No                  | 0.279              | No                         | No               | No                | 2.03                                    | 2.204                                      | No             | Yes                |
| Berberine         | No                  | No                  | 1.324              | No                         | No               | No                | 2.587                                   | 1.998                                      | No             | No                 |
| Beta carotene     | No                  | No                  | 1.024              | No                         | No               | No                | 1.766                                   | 0.666                                      | No             | No                 |
| Betulinic acid    | No                  | No                  | 0.076              | No                         | No               | No                | 2.298                                   | 2.161                                      | No             | No                 |
| Capsaicin         | No                  | No                  | 1.269              | No                         | No               | No                | 1.998                                   | 2.373                                      | Yes            | No                 |
| Catechins         | No                  | No                  | 0.215              | No                         | Yes              | No                | 2.101                                   | 2.076                                      | No             | No                 |
| Corchorusin-D     | No                  | No                  | -0.069             | No                         | No               | No                | 2.094                                   | 2.341                                      | No             | No                 |
| Curcumin          | No                  | Yes                 | 0.014              | No                         | No               | No                | 1.93                                    | 2.421                                      | No             | No                 |
| Diallyl sulfide   | No                  | No                  | 0.562              | No                         | No               | No                | 2.274                                   | 1.689                                      | No             | Yes                |
| Diosgenin         | No                  | No                  | 0.287              | No                         | No               | No                | 2.408                                   | 1.496                                      | No             | No                 |
| Ellagic acid      | No                  | No                  | 0.539              | No                         | Yes              | No                | 2.201                                   | 1.947                                      | No             | No                 |
| Eugenol           | No                  | No                  | 0.28               | No                         | No               | No                | 1.994                                   | 2.304                                      | No             | Yes                |
| Genistein         | No                  | No                  | 0.241              | No                         | Yes              | No                | 2.309                                   | 2.075                                      | No             | No                 |
| Indole-3-carbinol | No                  | No                  | 0.555              | No                         | No               | No                | 2.301                                   | 1.836                                      | No             | Yes                |
| Limonen           | No                  | No                  | 0.224              | No                         | No               | No                | 2.257                                   | 2.204                                      | No             | Yes                |
| Longimide         | No                  | No                  | 0.036              | No                         | No               | No                | 2.347                                   | 1.445                                      | Yes            | No                 |
| Longitriol        | No                  | No                  | 0.239              | No                         | No               | No                | 2.218                                   | 2.035                                      | No             | No                 |
| Lycopene          | No                  | No                  | 1.912              | No                         | No               | No                | 1.461                                   | 0.764                                      | No             | No                 |
| Methyl anolensate | No                  | Yes                 | 0.265              | No                         | No               | No                | 2.594                                   | 1.088                                      | No             | No                 |
| Resveratrol       | No                  | No                  | 0.147              | No                         | Yes              | No                | 2.072                                   | 2.605                                      | No             | No                 |
| S-allyl cyteine   | No                  | No                  | 0.613              | No                         | No               | No                | 2.153                                   | 2.479                                      | No             | No                 |
| Silymarin         | No                  | No                  | -0.092             | No                         | No               | No                | 2.184                                   | 2.593                                      | No             | No                 |
| Withaferin A      | No                  | No                  | 0.37               | No                         | No               | No                | 2.294                                   | 1.983                                      | No             | No                 |

2 **Table 3.** Conituned

| Drugs             |    |     |        |    |     |    |       |       |     |    |
|-------------------|----|-----|--------|----|-----|----|-------|-------|-----|----|
| Camptothecin      | No | No  | 0.564  | No | Yes | No | 2.465 | 1.845 | Yes | No |
| Docetaxel         | No | Yes | -0.172 | No | No  | No | 1.617 | 2.808 | Yes | No |
| Etoposide         | No | No  | 0.041  | No | No  | No | 2.071 | 2.331 | No  | No |
| Irinotecan        | No | Yes | 1.213  | No | No  | No | 2.766 | 1.693 | Yes | No |
| Paclitaxel        | No | Yes | -0.121 | No | No  | No | 1.708 | 3.081 | No  | No |
| Teniposide        | No | Yes | 0.409  | No | No  | No | 2.162 | 2.689 | No  | No |
| Topotecan         | No | No  | 1.196  | No | Yes | No | 2.58  | 1.867 | Yes | No |
| Vinblastine       | No | Yes | 0.618  | No | No  | No | 2.251 | 2.588 | Yes | No |
| Vincristine       | No | Yes | 0.739  | No | No  | No | 2.11  | 2.67  | Yes | No |
| Vinorelbine       | No | Yes | 0.622  | No | No  | No | 2.343 | 2.5   | Yes | No |
| Amrubicin         | No | Yes | 0.4    | No | No  | No | 2.531 | 1.721 | No  | No |
| Arglabin          | No | Yes | 1.059  | No | Yes | No | 2.022 | 2.103 | No  | No |
| Carfilzomib       | No | No  | 0.839  | No | No  | No | 1.986 | 1.482 | No  | No |
| Combretastatin    | No | Yes | 1.564  | No | No  | No | 2.162 | 3.012 | Yes | No |
| Ellipticine       | No | Yes | 0.222  | No | No  | No | 2.076 | 2.104 | No  | No |
| Exemestane        | No | Yes | 0.543  | No | No  | No | 2.712 | 1.522 | No  | No |
| Formestane        | No | Yes | 0.832  | No | No  | No | 2.062 | 1.822 | No  | No |
| Homoharringtonine | No | Yes | 0.64   | No | No  | No | 2.214 | 1.918 | No  | No |
| Podophyllotoxin   | No | Yes | 1.543  | No | No  | No | 2.162 | 1.86  | Yes | No |
| Rohitukine        | No | Yes | 0.14   | No | No  | No | 2.205 | 2.284 | No  | No |
| Roscovitine       | No | Yes | 0.495  | No | Yes | No | 2.448 | 1.607 | Yes | No |
| Vindesine         | No | Yes | 1.174  | No | No  | No | 2.761 | 1.683 | Yes | No |
| Vinflunine        | No | Yes | 0.467  | No | No  | No | 2.414 | 2.729 | Yes | No |
| Amrubicin         | No | Yes | 0.365  | No | No  | No | 2.311 | 2.267 | No  | No |

### 3 Conclusions

4 Improving the survival rate of clinical candidates and reducing the drug attrition is governed by multi-  
5 factors, and thus, a holistic strategy that addresses key attrition factors (safety, ADME, and efficacy).  
6 Chemical space defined by physicochemical properties is vast, yet there are several design parameters that  
7 medicinal chemists can follow when designing druglike compounds (e.g., Lipinski's Rule of Five) and  
8 defining the parameters that increase the likelihood of identifying best in class molecules is of critical  
9 importance. Understanding the fundamental relationships between physicochemical properties and *in vitro*  
10 and *in vivo* results is primary need to prospectively design compounds with an overall desired profile. As  
11 part of our efforts to further build this understanding in the Anticancer drug development space, we  
12 undertook a thorough analysis of the physicochemical properties, ADME attributes, and safety end points  
13 for 24 natural product derived anticancer drugs and 27 natural product lead candidates. We examined a  
14 comparison of eight fundamental physicochemical properties associated with these two sets of  
15 compounds:  $\log P$ ,  $\log D$ ,  $M_w$ , TPSA, HBD, HBA,  $\log S$  and  $n_{\text{Rot}}$ . The anticancer drug space defined by these  
16 physicochemical properties is pretty broad, but our analysis identified the optimum ranges for each of  
17 these properties. The optimal property ranges (covering almost ~80 % or more of the anticancer drugs)  
18 were found to be  $200 < M_w \leq 800$  Da,  $1 < \log P \leq 5$ ,  $-6 \leq \log S \leq -1$ ,  $5 \leq \text{HBA} \leq 13$ ,  $1 \leq \text{HBD} \leq 5$ ,  $50 \leq \text{TPSA} \leq$   
19  $180 \text{ \AA}^2$ ,  $0 \leq n_{\text{Rot}} \leq 10$ ,  $\log D=2.8$ . Analysis of *in silico* generated ADME data reinforced that the majority of  
20 anticancer drugs (70 %) are low permeable (Caco2 of  $\log P_{\text{app}}$  (in  $10^{-6}$  cm/s)  $< 0.9$ ), and also all drugs are  
21 considered to be P-gp efflux substrates, and with low to moderate clearance rates.

22 On the other hand, our analysis showed that for anticancer drugs, there may be a need to optimize new  
23 compounds with further reduced  $M_w$ , HBA, and  $n_{\text{Rot}}$  to better match the corresponding properties in the  
24 marketed drug set. In addition, we have established meaningful correlations between the physicochemical  
25 properties specially HBA, HBD, and TPSA and ADME attributes of the molecules that might be generally  
26 applicable for the future anticancer drug development and optimization of the natural product derived  
27 anticancer leads/clinical candidates. Our study showed the meaningful correlations between  
28 physicochemical properties and toxicity profile of compounds.  $\log P$  and  $M_w$  are most critical  
29 physicochemical parameter and robust predictor of toxicity profile of anticancer leads/clinical candidates  
30 We showed by our analysis that early prediction of physicochemical properties, ADME attributes, and  
31 safety attributes through *in silico* tools are all important parameters to enable better lead candidate  
32 selections, saving considerable time and effort in the anticancer drug development.

33

34 **Acknowledgements:** Author is thankful to the director, Central Institute of Medicinal and Aromatic Plants  
35 (CIMAP-CSIR), Lucknow.

36

### 37 References

- 38 [1] A. Jemal, R. Siegel, J. Xu, E. Ward. *CA: A Cancer Journal for Clinicians* **60** (2010) 277-300.  
39 [2] B.B. Aggarwal, D. Danda, S. Gupta, P. Gehlot. *Biochemical Pharmacology* **78** (2009) 1083–1094.  
40 [3] S. Coseri, *Mini-Reviews in Medicinal Chemistry* **9** (2009) 560-571.  
41 [4] D.J. Newman, *Journal of Medicinal Chemistry* **51** (2008) 2589-2599.  
42 [5] D.J. Newman, G. M. Cragg, *Journal of Natural Products* **75** (2012) 311-335.  
43 [6] M. S. Kinch, A. Haynesworth, S.L. Kinch, D. Hoyer, *Drug Discovery Today* **19** (2014) 1033–1039.

- 44 [7] C. D. Bradley, O. Björn, G. Fabrizio, K. Jan, *Chemistry & Biology* **21** (2014) 1115 – 1142.
- 45 [8] N.A. Meanwell, *Chemical Research in Toxicology* **24** (2011) 1420–1456.
- 46 [9] P. Barton, R.J. Riley, *Drug Discovery Today* **21** (2016) 72-81.
- 47 [10] M.P. Gleeson, *Journal of Medicinal Chemistry* **51** (2008) 817–834.
- 48 [11] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, *Advanced Drug Delivery Reviews* **23** (1997) 3–
- 49 25.
- 50 [12] D.F. Veber, S.R. Johnson, H.Y. Cheng, B.R. Smith, K.W. Ward, K.D. Kopple, *Journal of Medicinal*
- 51 *Chemistry* **45** (2002) 2615-2623.
- 52 [13] A.K. Ghose, V.N. Viswanadhan, J.J. Wendoloski *Journal of Computational Chemistry* **1** (1999) 55-68.
- 53 [14] P. D. Leeson, *Advanced Drug Delivery Reviews* **101** (2016) 22-33.
- 54 [15] E.C.A Cornelis. Hop Attrition in the Pharmaceutical Industry: Reasons, Implications, and Pathways
- 55 Forward, Ed. A. Alex, C.J. Harris, D.A. Smith. (2016) John Wiley & Sons, Inc.
- 56 [16] D.C. Swinney, J. Anthony, *Nature Reviews Drug Discovery* **10** (2011) 507–519.
- 57 [17] K.K. Dholwani, A.K. Saluja, A.R. Gupta, D.R. Shah, *Indian Journal of Pharmacology* **40** (2008) 49–58.
- 58 [18] A. Bhanot, R. Sharma, M. N. Noolvi, *International Journal of Phytomedicine* **3** (2011) 9-26.
- 59 [19] T. Sander, J. Freyss, M.V. Korff, C. Rufener, *Journal of Chemical Information and Modeling*, **55** (2015)
- 60 460–473.
- 61 [20] M.D. Hanwell, D.E. Curtis, D.C. Lonie, T. Vandermeersch, E. Zurek, G.R. Hutchison, *Journal of*
- 62 *Cheminformatics* **4** (2012) 1-17.
- 63 [21] E.V.P. Douglas, T.L. Blundell, D.B. Ascher, *Journal of Medicinal Chemistry* **58** (2015) 4066–4072.
- 64 [22] M. J. Waring, *Bioorganic & Medicinal Chemistry Letters*, **19** (2009) 2844–2851
- 65 [23] D.A. Smith, L. Di, E.H. Kerns, *Nature Reviews Drug Discovery* **9** (2010) 929-939.
- 66 [24] D.A. Smith, H. Van de Waterbeemd, D.K. Walker, (2001) *Pharmacokinetics and Metabolism in Drug*
- 67 *Design*. Wiley–VCH, Weinheim, Germany.
- 68 [25] K. A. Houck, R.J. Kavlock, *Toxicology and Applied Pharmacology* **227** (2008) 163–178.
- 69 [26] V.P. Miller, D.M. Stresser, A. P. Blanchard, S. Turner, C. L. Crespi, *Annals of the New York Academy of*
- 70 *Sciences* **919** (2000) 26–32.
- 71 [27] M. Deacon, D. Singleton, N. Szalkai, R. Pasieczny, C. Peacock, D. Price, J. Boyd, H. Boyd, J.V. Steidl-
- 72 Nichols, C. Williams, *Journal of Pharmacological and Toxicological Methods* **55** (2007) 238–247.
- 73 [28] E.H. Kerns, L. Di, (2008). *Drug-Like Properties: Concepts, Structure Design and Methods: From ADME*
- 74 *to Toxicity Optimization*. (Amsterdam, Boston: Academic Press).
- 75 [29] C.M. Hosey, L.Z. Benet, *Molecular Pharmaceutics* **12** (2015) 1456-1466.
- 76 [30] F. Lovering, J. Bikker, C. Humblet, *Journal of Medicinal Chemistry* **52** (2009) 6752–6756.
- 77 [31] J.D. Hughes J. Blagg, D.A. Price, S. Bailey, G.A. DeCrescenzo, R.V. Devraj, E. Ellsworth, Y.M. Fobian,
- 78 M.E. Gibbs, R.W. Gilles, N. Greene, E. Huang, T. Krieger-Burke, J. Loesel, T. Wager, L. Whiteley, Y.
- 79 Zhang, *Bioorganic & Medicinal Chemistry Letters* **18** (2008) 4872–4875.
- 80 [32] D. Muthas, S. Boyera, C. Hasselgren, *Medicinal Chemical Communications* **4** (2013) 1058–1065.
- 81 [33] J.J. Sutherland, J.W. Raymond, J.L. Stevens, T.K. Baker, D.E. Watson, *Journal of Medicinal*
- 82 *Chemistry* **55** (2012) 6455–6466.
- 83
- 84