Drug-likeness and Lead-likeness: An Overview of Recent Studies

Zhengming (Jimmy) Chen

Computational, Combinatorial and Medicinal Chemistry, Purdue Pharma L. P., 6 Cedar Brook Drive, Cranbury, NJ 08512, zhengming.chen@pharma.com

Introduction

The distinction between drug-like and non drug-like molecules has been a hot research topic in recent years. The most well known early study in this field is the "Lipinski's rule of five" which was derived empirically from the analysis of the World Drug Index on the properties that maximize an oral drug candidate's probability of surviving clinical development: molecular weight (MW) < 500, number of hydrogen bond donors < 5, number of hydrogen bond acceptors < 10, and ClogP < 5. In order to be absorbed through the gut and enter the bloodstream, orally administered drugs must have certain physical properties, especially those described by the Lipinski Rule of 5. The rule of five is now widely used to filter out compounds likely to have poor pharmacokinetic properties early on in drug discovery. Lead-likeness (compounds' likelihood to be good lead candidates), as distinct from drug-likeness, is a new concept that is gaining acceptance in recent years. Lead-like molecules are generally smaller to allow for structural additions to enhance effectiveness during lead optimization, and is being incorporated into the library design and lead optimization processes. In this presentation, I am going to present an overview of recent studies on the topic of drug-likeness and lead-likeness. The presentation will provide a few intriguing insights into the influence of molecular properties on the likelihood of progression through the drug development process and the trends in modern drug discovery.

The Famous Lipinski "Rule of Five" [1]

- Molecular Weight <= 500 - Size
- # Hydrogen Bond Acceptors <=10 (Sum of N and O) — Pharmacophoric Properties
- # Hydrogen Bond Donors <= 5 (Sum of NH and OH) — Pharmacophoric Properties
- -2 < CLog P < 5. - Solubility
- # Rotatable Bonds <= 10 - Molecular Flexibility

"DRUG-LIKE BEHAVIOUR"

1: C. Lipinski et al, Adv. Drug. Del. Rev, 23, 3-25 (1997)

The New Lead-likeness Rule [2,3]

- Molecular Weight: 100–350 - Size
- Fewer Hydrogen Bond Acceptors (Sum of N and O) — Pharmacophoric Properties
- Fewer Aromatic Rings — Pharmacophoric Properties
- 1 < CLog P < 3. - Solubility
- Low to High Affinity for the target receptor

"LEAD-LIKE BEHAVIOUR"

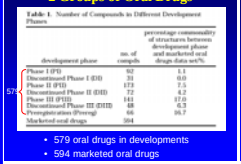
2: Teague, S.; Davis, A. M.; Leeson, P. D.; Oprea, T. Angew. Chem. Int. Ed. 1999, 38, 3743–3748
3: Hann, M. M.; Leach, A. R.; Harper, G. J. Chem. Inf. Comput. Sci. 2001, 41, 856–864

Drug-likeness: J. Med. Chem. 2003, 46, 1250-1256

Recently a comparison of physiochemical properties of marketed oral drugs with those in different development phase has been reported in the Journal of Medicinal Chemistry. It provides some intriguing insights into drug-likeness and the influence of these properties on the likelihood of progression through the drug development process.

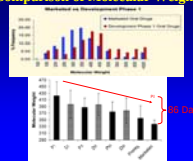
Wenlock and colleagues took 594 oral drugs marketed in the United States and 579 potential oral drugs from all phases of clinical development — both those still in trials and those for which trials had been discontinued — and calculated various physiochemical properties for each. Several trends emerged. Particularly notable was that the mean MW of orally administered drugs in development decreases on passing through each phase, and seems to converge towards the mean MW of the marketed drugs. Moreover, the mean MW of the compounds discontinued from a particular phase is greater than the mean MW of the compounds in the next phase. A similarly clear trend was apparent in the data for log P — a measure of lipophilicity — with the most lipophilic compounds being discontinued at each phase, consistent with the common finding that high lipophilicity frequently leads to compounds that are rapidly metabolized and that have low solubility and poor absorption.

2 Groups of Oral Drugs

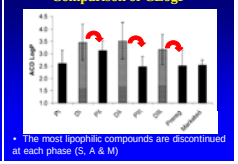


- 579 oral drugs in developments
- 594 marketed oral drugs

Comparison of Molecular Weight

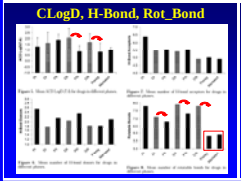


Comparison of CLogP

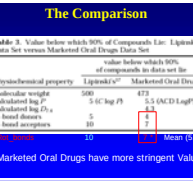


- The most lipophilic compounds are discontinued at each phase (S, A & M)

CLogD, H-Bond, Rot_Bond



The Comparison



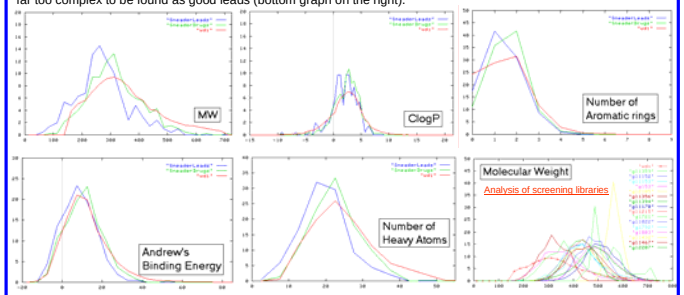
- Marketed Oral Drugs have more stringent Values

Key Findings in This Paper

- The mean MW of orally administered drugs in development decreases on passing through each phase, and seems to converge towards the mean MW of the marketed drugs
  - Implication: Larger size molecules are less favorable to cross lipid bilayers (reducing passive absorption)
  - Larger molecules are also more likely to contain toxic pharmacophores or rapidly metabolized moieties.
- The most lipophilic compounds being discontinued at each phase
  - Implication: high lipophilicity frequently leads to compounds that are rapidly metabolized and that have low solubility and poor absorption
- Compounds discontinued from a particular phase have a higher number of rotatable bonds than next phase
  - Implication: Compounds with lower molecular flexibility tend to have better oral bioavailability

Leads, Drugs and Non-drugs

Hann, et al. have taken data from W. Sneader's book "Drug Prototypes and Their Exploitation" and converted them into the Daylight Database and then profiled 480 drug case histories with ADEPT in the following plots. The blue line represents the Sneader's leads, green line represents the Sneader's drugs and the red line represents the WDI compounds (drugs and clinical candidates). These comparisons have supported the earlier findings that leads are simpler than drugs. An similar analysis of some screening libraries has shown that library compounds are often far too complex to be found as good leads (bottom graph on the right).



Jens Sadowski and Hugo Kubinyi have developed a scoring scheme for the rapid and automatic classification of molecules into drugs and non-drugs. The method is based on the extraction of knowledge from large databases of drugs and non-drugs (left graph below). It was set up by using atom type descriptors for encoding the molecular structures and by training a feed forward neural network for classifying the molecules. The method has been parameterized and validated by using large databases of drugs and non-drugs (169331 molecules from the Available Chemicals Directory (ACD), and 38416 molecules from the World Drug Index, WDI). This method included features in the molecular descriptors that either qualify or disqualify a molecule for being a drug and has classified 83% of the ACD and 77% of the WDI adequately. The red line represents the drug score of the 100 top-selling drugs in 1997.

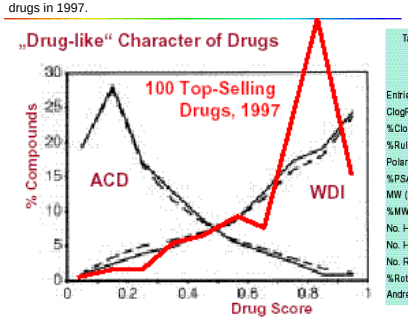


Table 1. Molecular Properties of "Launched" Compounds.

Property	Launched Compounds	Top Selling Compounds
Entries	882	138
ClogP	2.4	2.1
%ClogP > 5	13.5	8.7
%Rule-of-5 Violations	10.4	5.8
Polar Surface Area (Å²)	136.0	134.0
%PSA > 200 Å²	18.6	18.7
MW (AMU)	359.7	357.0
%MW > 500	11.2	10.1
No. H-Bond Acceptors	4.3	4.3
No. H-Bond Donors	2.0	2.0
No. Rotatable Bonds	6.9	6.7
%Rotatable Bonds > 10	16.2	14.5
Andrews' BE (kcal/mol)	11.7	11.3

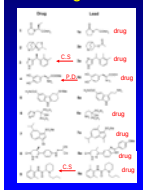
Blake, et al. have analyzed 882 lunched drugs which were derived from the MDDR-99.2 database after elimination of compounds considered to be diagnostics, topical agents, peptides, etc. For each of the compounds in the data sets, they computed a number of properties that have been shown to be important for characterizing the drug-likeness, such as Andrews' binding energy, polar surface area (PSA), rotatable bond counts, log P, molecular weight, and the number of H-bond donors and acceptors. Andrews' binding energy can be thought of as an empirical measure of molecular complexity. The properties described here are also more amenable to change via chemical synthesis. The calculations for each property are summarized in the Table on the top right. While average values for the given properties are useful, it is also important to consider how compounds fare when they possess out-of-range values. The percentage of compounds that fall outside of these cutoff values is also reported in the table.

A similar analysis of the top 200 best selling drugs based on total US prescriptions for 2001 (a data set of 138 compounds after same elimination) has also been done. Only eight of the top 200 selling small molecule drugs in 2001 violated two or more 'Rule of 5' parameters. Five of these compounds are known substrates for transporters, one is a pro-drug, and two require soft-gel formulations.

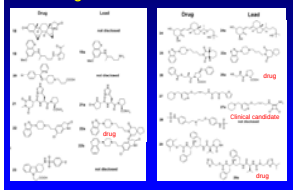
Leads *vs* Drugs: BMCL, 2002, 12, 1647-1650

Proudfoot has performed an analysis of drugs launched in 2000 and their corresponding lead structures. His study has demonstrated that the drug structures are very closely related to their leads although the leads are simpler in most cases. An analysis of the origins of these drugs also reveals that most of them were derived by modification of the known drug structures or from lead structures obtained from the scientific literature. High-throughput screening did not have a significant impact on the derivation of these drugs.

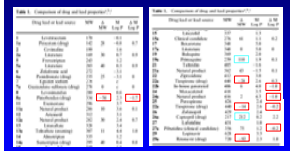
Drugs in 2000 & their leads



Drugs in 2000 & their leads



Comparison of Drugs & their leads



- Both M.W. and LogP increased from lead to drug

Key Findings in This Paper

- A majority of leads are drugs/clinical candidates or natural products, only one drug (#13) was it indicated that screening played a role in the discovery of the lead structure
  - Implication: High throughput screening did not have a significant impact on the discovery of these drugs
- Despite the varied objectives (improving potency, selectivity, chemical stability, PK, etc.) driving the optimization of these drugs, the drug and lead structures are generally remarkably similar (within 25% M.W. & 1 logP unit, 2 exceptions)
  - Implication #1: Structurally, most of the drugs are derived by quite limited modification of the lead
  - Implication #2: Successful, timely drug discovery campaigns require high quality lead structures and those lead structures may need to be much more drug like than is commonly accepted.

Implications in Lead Generation and Optimization

Despite continued and unprecedented levels of investment in high-throughput screening (HTS) and combinatorial chemistry technologies, lead discovery still remains a key bottleneck in today's drug discovery process. Several major pharmaceutical companies have acknowledged that they are only successful in identifying a high-quality lead for a druggable protein target in about one out of four attempts. The facts of that library compounds possess a significant increase in the mean MW compared with the marketed drugs and that HTS often pick up hits with high MW, lipophilicity and # of rotatable bonds could be a potential reason for the under delivery of the past investment.

The results in this drug-likeness and lead-likeness presentation could provide us with the following guidelines in our future design of screening library and selection of HTS hits for optimization:

1) Drug-like and lead-like character is more important than the synthetic accessibility in the design of screening library

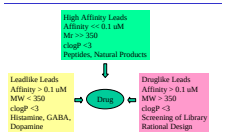
2) Screening libraries need to be more 'lead-like' — that is, have lower MW and lipophilicity than marketed drugs, and screening libraries should have a high degree of chemical diversity

3) Lead optimization libraries should have a high degree of similarity to cover the chemical space around a lead structure

4) A thoughtful strategy in lead optimization should include the simultaneous optimization of the pharmacological properties and molecular properties of the final compounds.

5) Pharmacophore based lead optimization approach should be used to reduce MW and lipophilicity of HTS hits.

Classification of Leads



Lead Optimization