

Examination of the Molecular Properties of Drugs

James Blake, Ph.D. - Senior Research Investigator, Array BioPharma

Dr. James F. Blake earned a Bachelors of Special Studies degree from Cornell College in 1984, majoring in chemistry, with minors in mathematics, and physics. His graduate study was carried out at Purdue University (Ph.D. 1990) with Professor William L. Jorgensen, investigating prototypical reactions via ab initio molecular orbital calculations and condensed phase Monte Carlo simulations. Dr. Blake spent the next 10 years at the Central Research Division of Pfizer, Inc. supporting CNS, lead generation, and attrition related discovery project areas. He has coauthored 40 publications and four patents.

Since September 2001, Dr. Blake has been working at Array supporting the lead generation and optimization areas.

In the late 1980's, the advent of random high-throughput screening (HTS) as a means of finding novel lead compounds introduced substantial changes in the way pharmaceutical companies view their compound collections. Since most in-house collections came primarily from either traditional synthetic techniques or targeted acquisitions, the size of these collections tended to be rather modest. The success of random HTS at finding novel leads relies on assaying large numbers, typically in the hundreds of thousands of compounds.¹ From an analysis of 56 HT screens carried out on a variety of enzyme, G-protein coupled receptor, ion channel, and receptor targets, Spencer² concluded that an average of one quality lead series is identified for every 120,000 compounds assayed. The link between deriving leads from HTS and the size of the compound collection screened led to a substantial increase in the number of compounds purchased from outside sources. Lipinski³ has suggested that this increased reliance on leads derived from HTS also led to the pursuit of compounds with less than optimal physicochemical properties, and hence to more difficulty in lead optimization and clinical development.

Experience has shown that improving the oral bioavailability of a compound already optimized for potency is the most time consuming, labor intensive, and variable part of the preclinical drug discovery process.⁴ These observations led to the creation of the 'Rule of 5'.⁵ Lipinski⁵ compared drug-like compounds (compounds with USAN or INN names extracted from the World Drug Index) with compounds not presumed to be drug-like (those found in the ACD) and determined that compounds with excessive logP, molecular weight (MW), and H-bond donors or acceptors were more likely to have solubility and or permeability problems that would lead to poor oral bioavailability. Since the publication of the 'Rule of 5,' there have been many additional studies aimed at distinguishing between drugs and non-drugs,⁶⁻⁷ characterizing the properties of drug-like databases,⁸⁻¹¹ and predicting oral bioavailability.¹²⁻¹⁴ From these exercises, one would hope to define the appropriate chemical space from which leads should be derived.

Proudfoot¹⁵ has performed an analysis of drugs launched in 2000 and their corresponding lead structures and demonstrated that leads are very similar to the final drugs, both in terms of structure and simple physical properties. For the 29 series of compounds in that study, the

leads and drugs were generally within 25% of the MW and one logP unit of each other. Many of the 'leads' in this study were actually marketed drugs. This result is consistent with an earlier and larger analysis by Oprea,¹⁶ which underscored the importance of starting with the best compounds to minimize clinical attrition rates. Higher-quality leads should ultimately give rise to clinical candidates with higher success rates. Since most, if not all, drugs bear a striking similarity to the leads from which they were derived, both in terms of structure and physicochemical properties, the question arises: which compounds should be added to the screening collection and what are appropriate physical properties?

For our study, drug and drug-like compounds were derived from the MDDR-99.2 database.¹⁷ The MDDR database contains compounds targeted and tested for potential therapeutic value that were compiled from various journals, meetings, and patent literature. Given the extremely complex nature of drug development, this database represents a "snapshot" of compounds in development. The MDDR database contains a total of 105,084 compounds. About 95,151 of these compounds fall into the class of "Biological Testing." For our purposes, we are only interested in small molecule compounds that are labeled as "Launched," indicating that they have gained FDA approval. We eliminated all compounds for which no therapeutic activity class was given, as well as compounds considered to be diagnostics, topical agents, peptides, or not targeted toward a specific disease state. This process reduced the MDDR database to a data set containing 882 compounds. A similar analysis of the top 200 selling drugs, based on total US prescriptions for 2001, resulted in a data set of 138 compounds.¹⁸

For each of the compounds in the data sets, we computed a number of properties that have been shown to be important for characterizing leads and drug-like molecules, such as Andrews' binding energy, polar surface area (PSA), rotatable bond counts, logP, 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.¹⁹ The relevance of these properties to the processes of solubility, permeability, and absorption has been reviewed.^{20,21} The properties described are also more amenable to change via chemical synthesis. The calculations for each property are summarized in Table 1.

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 (Å ²)	136.0	134.0
%PSA > 200 Å ²	18.6	16.7
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

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. For a 'Rule of 5' type analysis, this would include both ClogP > 5 and MW > 500. Analysis has also shown that a large PSA^{4,21} (greater than 150-200 Å²) or rotatable bonds¹⁴ beyond 10 can lead to dramatically decreased permeability and oral bioavailability. The percentage of compounds that fall outside of these cutoff values is also reported in Table 1. 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.

Implications for Combinatorial Library Design

How do the current results affect our thinking toward suitable properties for combinatorial library design and the chemical space of appropriate leads for optimization? Combinatorial libraries should be populated with compounds that have their physical properties distributed around the values in Table 1. Perhaps more important is ensuring that excessive numbers of compounds do not lie outside the property cut off ranges above, as these compounds will likely experience significant development issues. In designing combinatorial libraries, the chemist is often faced with manipulating many variables and having to select from a large number of possible inputs. A thoughtful strategy should include paying close attention to the properties of the final products, as this will have the greatest impact on the future success of any lead optimization program.

References

- Golebiowski, A.; Klopfenstein, S. R.; Portlock, D. E. Lead compounds discovered from libraries. *Curr. Opin. Chem. Biol.* **2001**, *5*, 273-284.
- Spencer, R.W. High-Throughput Screening of Historic Collections: Observations on File Size, Biological Targets, and File Diversity. *Biotechnol. Bioeng. (Comb. Chem.)* **1998**, *61*, 61-67.
- Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 235-249.
- Lipinski, C. A. Consideration of Drug Solubility in High Throughput Screening of Oral Drug Candidates. "Strategies for Optimizing Oral Drug Delivery: Scientific to Regulatory Approaches." **1999**, April 19-21, Kobe, Japan.
- Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **1997**, *23*, 3-25.
- Ajay; Walters, W. P.; Murcko, M. A. Can We Learn to Distinguish between "Drug-like" and "Nondrug-like" Molecules? *J. Med. Chem.* **1998**, *41*, 3314-3324.
- Sadowski, J.; Kubinyi H. A Scoring Scheme for Discriminating between Drugs and Nondrugs *J. Med. Chem.* **1998**, *41*, 3325-3329.
- Oprea, T. I. Property distribution of drug-related chemical databases. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 251-264.
- Muegge, I.; Heald, S. L.; Britelli, D. Simple Selection Criteria for Drug-like Chemical Matter. *J. Med. Chem.* **2001**, *44*, 1841-1846.
- Brüstle, M.; Beck, B.; Schindler, T.; King, W.; Mitchell, T.; Clark, T. Descriptors, Physical Properties, and Drug-Likeness. *J. Med. Chem.* **2002**, *45*, 3345-3355.
- Toung, B. A.; Pfahler, L. B.; Reynolds, C. H. Chemical Information Based Scaling of Molecular Descriptors: A Universal Chemical Scale for Library Design and Analysis *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 879-884.
- Wessel, M. D.; Jurs, P. C.; Tolan, J. W.; Muskal, S. M. Prediction of Human Intestinal Absorption of Drug Compounds from Molecular Structure. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 726-735.
- Yoshida, F.; Topliss, J. G. QSAR Model for Drug Human Bioavailability. *J. Med. Chem.* **2000**, *43*, 2575-2585.
- Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B.R.; Ward, K. W.; Kopple, K. D. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates *J. Med. Chem.* **2002**, *45*, 2615-2623.
- Proudfoot, J. R. Drugs, Leads, and Drug-Likeness: An Analysis of Some Recently Launched Drugs. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1647-1650.
- Oprea, T. I.; Davis, A. M.; Teague, S. J.; Leeson, P. D. Is There a Difference between Leads and Drugs? A Historical Perspective. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1308-1315.
- MDL Drug Data Report, MDL Information Systems Inc., San Leandro, CA, USA.
- Top selling compounds from <http://www.rxlist.com/top200.htm>
- Andrews, P. R.; Craik, D.J.; Martin, J. L. Functional Group Contributions to Drug-Receptor Interactions. *J. Med. Chem.* **1984**, *27*, 1648-1657.
- Blake, J. F. Chemoinformatics - predicting the physicochemical properties of 'drug-like' molecules. *Curr. Opin. Biotech.* **2000**, *11*, 104-107.
- Clark, D. E. and Pickett, S. D. Computational Methods for the Prediction of Drug-likeness. *Drug Discovery Today* **2000**, *5*, 49-58.