Drug Discovery Today: Technologies
Vol. 1, No. 4 2004

Editors-in-Chief
Kelvin Lam – Pfizer, Inc., USA
Henk Timmerman – Vrije Universiteit, The Netherlands

Lead profiling

Lead- and drug-like compounds: the rule-of-five revolution

Christopher A. Lipinski*

Pfizer Global R&D, Groton Labs, Eastern Point Road, MS 8200-36, Groton, CT 06340, USA

Citations in CAS SciFinder to the rule-of-five (RO5) publication will exceed 1000 by year-end 2004. Trends in the RO5 literature explosion that can be discerned are the further definitions of drug-like. This topic is explored in terms of drug-like physicochemical features, drug-like structural features, a comparison of drug-like and non-drug-like in drug discovery and a discussion of how drug-like features relate to clinical success. Physicochemical features of CNS drugs and features related to CNS blood–brain transporter affinity are briefly reviewed. Recent literature on features of non-oral drugs is reviewed and how features of lead-like compounds differ from those of drug-like compounds is discussed. Most recently, partly driven by NIH roadmap initiatives, considerations have arisen as to what tool-like means in the search for chemical tools to probe biology space. All these topics frame the scope of this short review/perspective.

Section Editors:
Han van de Waterbeemd, Christopher Kohl – Pfizer Global Research & Development, Sandwich Laboratories, PDM (Pharmacokinetics, Dynamics and Metabolism), Sandwich, Kent, UK CT13 9NJ

In the past, many clinical candidates failed during development. The reasons for failure are now much better understood. The author of this contribution, Chris Lipinski, was among the first to point out that drugs typically have physicochemical and structural properties within certain ranges. This review discusses the original rule-of-five concept and its variants, to be used in the design of orally active compounds. He also compares the concepts of drug-like, lead-like, and CNS-like compounds and drugs. It is important to consider differences better oral and non-oral drugs. Finally, the new idea of tool-like compounds is presented.

Introduction

Citations in CAS SciFinder to the original rule-of-five (RO5) publication [1] in 1997 and its reprint in 2001 will exceed 1000 by the end of 2004. Trends in the RO5 literature explosion that can be discerned include: (1) further definitions of drug-like; (2) definitions of lead-like and most recently; (3) considerations of what tool-like means in the search for chemical tools to probe biology space. These topics frame the scope of this short review.

Drug-like

The Comprehensive Medicinal Chemistry (CMC), Derwent Word Drug Index (WDI) and Modern Drug Data Report (MDDR) are among the more commonly used drug-like databases [2,3]. The meaning of ''drug-like'' is dependent on mode of administration. The original RO5 deals with orally active compounds and defines four simple physicochemical parameter ranges (MWT $\leq$ 500, log $P \leq 5$, H-bond donors ≤ 5, H-bond acceptors ≤ 10) associated with 90% of orally active drugs that have achieved phase II clinical status. These physicochemical parameters are associated with acceptable aqueous solubility and intestinal permeability and comprise the first steps in oral bioavailability. The RO5 was deliberately created to be a conservative predictor in an era where medicinal and combinatorial chemistry produced too many compounds with very poor physicochemical properties. The goal

E-mail address: (C.A. Lipinski) christopher.a.lipinski@pfizer.com.
* Retired.

 DOI: 10.1016/j.ddtec.2004.11.007

was to change chemistry behavior in the desired direction. If a compound fails the RO5 there is a high probability that oral activity problems will be encountered. However, passing the RO5 is no guarantee that a compound is drug-like. Moreover, the RO5 says nothing about specific chemistry structural features found in drugs or non-drugs.

Drug-like physicochemical properties

Rotatable bond count is now a widely used filter following the finding that greater than 10 rotatable bonds correlates with decreased rat oral bioavailability [4]. The mechanistic basis for the rotatable bond filter is unclear because the rotatable bond count does not correlate with *in vivo* clearance rate in the rat, but the filter is reasonable from an *in vitro* screening viewpoint because ligand affinity on average decreases 0.5 kcal for each two rotatable bonds [5]. An analysis of small drug-like molecules suggests a filter of $\log D > 0$ and <3 enhances the probability of good intestinal permeability [6].

Drug-like structural features

From the study of a database of commercially available drugs it is clear that the diversity of molecular framework (ring) shapes is extremely low. The shapes of half of the drugs in the database are described by the 32 most frequently occurring frameworks [7]. The diversity that side chains provide to drug molecules is also low because only 20 side chains account for over 70% of the side chains [8]. Defining drug-like by what exists in databases leads to the criticism that most of chemistry space will be undefined and that discovery opportunities in unexplored chemistry space will be limited. A solution is to populate chemistry space with non-drug-like markers akin to the way point in a GPS navigation system [9].

Comparison of drug-like and non-drug-like

Both simple and complex filters have a role in combinatorial library design. Simple properties, for example, privileged building blocks and counting of structural properties (e.g. number of H-bond parameters) to complex calculations (e.g. regression or neural network-based models) explain the relationship of structural features to ADME properties [10]. Drugs must contain adequate functionality to achieve acceptable receptor interactions. A single filter for underfunctionalization separates drug-like from non-drug-like compounds [11]. Using retrospective analyses of known drugs, including simple property counting schemes, machine learning methods, regression models, and clustering methods have all been employed to distinguish between drugs and non-drugs [12]. With relatively little computational effort, virtual libraries can be optimized with respect to diversity, high similarity with an already known screening hit or lead structure, or to improve drug-likeness and ADME properties [13]. The current trend is to smaller, high purity, information-rich libraries with reduced ADMET problems. At least part of this trend is due to the realization that large libraries based on a single combinatorial core are not very diverse. In contrast to the array character of combinatorial libraries, natural products are singleton sources of inspiration for scaffolds and reagents [14]. Privileged structures, such as benzodiazepines, are recurring structures active against targets unrelated by target family. They can be viewed as molecular filters selecting for desirable chemistry subject matter. As such they are rich singleton sources for screening libraries and have recently been reviewed [15,16]. Privileged reagents, akin to the concept of privileged structures, have been described using a retrosynthetic-like analysis of known drugs [17]. Exclusionary filters have been described that remove reactive chemical functionality on the basis of the premise that compounds having covalent chemistry possibilities have no place in drug discovery [18].

Clinical success

Property profiles of oral drugs are independent of the year in which the drug was approved to market and to some degree independent of target. As a compound progresses through clinical trials there is a steady change in properties, for example, molecular weight (MWT), log *P* and polar surface area (PSA) all decline with a MWT of about 340 found for marketed drugs [19,20]. The reason for this pattern is unclear because properties related to oral absorption would be expected to have reached a plateau by phase 2 and hence at that point selection pressure for properties related to oral absorption should have disappeared [21]. One possibility is that there is a selective failure of certain target classes in the clinical process and that propensity to failure is linked to simple properties, such as MWT. However, this argument, if correct, does not explain the issue of causality. For example, is it the case that certain target classes that tend to fail in the clinic also happen to have higher MWT ligands, or is it that something related to higher MWT *per se* leads to higher failure in certain target classes.

CNS drugs

Compounds are classified as central nervous system (CNS) active or inactive by one of two methods. Either the compound experimentally exhibits evidence of brain penetration or the compound is found in a CNS-active data set. Parameters related to CNS activity or inactivity are generally either (1) physicochemical properties or (2) properties related to CNS transporter affinity [most often the P-glycoprotein (PGP) efflux transporter]. A computational prediction cannot be better than the underlying experimental data set. The log–brain to blood–drug concentration ratio is almost universally the experimental measurement that is used for predicting CNS-active compounds. This convenient to measure parameter has been severely criticized with the prediction that if used ''*in silico* models of brain penetration will continue to be

of only limited benefit to industry'' [22]. Beyond the quality issue, the numerical paucity of brain permeation data is a current key issue in this area of research [23].

CNS drug physicochemical features

A PSA value of less than 60–70 tends to identify CNS-active compounds [24]. A very simple set of two rules predicts CNS activity: If N + O (the number of nitrogen and oxygen atoms) in a molecule is less than or equal to five, it has a high chance of entering the brain. The second rule predicts that if $\log P - (N + O)$ is positive then the compound is CNS-active [25].

CNS drug transporter affinity

The importance of transporter effects to CNS activity is emphasized by the estimate that about 15% of all genes selectively expressed at the blood–brain barrier (BBB) encode for transporter proteins, and that only about 50% of BBB transporters are currently known [26]. A scheme for separating CNS from non-CNS-active drugs in the WDI has enabled the discovery of simple parameters relating to passive BBB permeability and prediction PGP affinity [27]. PGP is a major barrier to entry of compounds into the CNS [28]. Appropriately determined PGP efflux ratios can be used as a measure of compound affinity to PGP. However, filters based on PGP efflux ratio from the Caco-2 colonic cell permeability cell culture assay do not correlate with *in vivo* rat brain penetration [29]. A collection of 1700 CNS and non-CNS drugs was used to model the passive diffusion component of BBB permeation and the physicochemical requirements of PGP (CNS-efflux) substrates. Very simple descriptors were sufficient to evaluate BBB permeation [30]. The considerable variety of *in silico* models used to predict blood–brain permeation have recently been reviewed [31].

Non-oral drugs

If a drug is low-MWT, does the method of delivery have to be oral? Can the rule of 5 be bypassed by delivering the drug by a non-oral route (e.g. pulmonary, intra nasal or dermal)? The answer depends very much on dose. If the total dose is 20 mg or less then alternative delivery routes begin to be feasible. However, a limitation is that approximately 10% of current clinical candidates have sufficient potency in the 0.1 mg/kg range to result in such a low dose [32]. Differences in property ranges between oral and injectable drugs have been summarized [33]. Oral drugs are lower in MWT and have fewer H-bond donors, acceptors and rotatable bonds. Pulmonary drugs tend to have higher PSA because pulmonary permeability is less sensitive to polar hydrogen-bonding functionality [34].

Lead-like drugs

The difference between drug-like and lead-like has been described [35]. There are two general meanings of lead-like. In one lead-like definition, compounds have reduced property range dimensions compared to the drug. In another definition, lead-like discovery refers to the screening of small MWT libraries with detection of weak affinities in the high micromolar to millimolar range. These low MWT libraries are often referred to as fragment libraries. Small fragment screening can be by NMR [36–38] or by X-ray [39,40] or in theory by any method capable of detecting weak interactions. A rule of three has been coined for these small molecule fragment screening libraries; MWT < 300; $\log P < 3$; H-bond donors and acceptors <3 and rotatable bonds <3 [41]. The experimental compared with the theoretically achievable diversity is much higher in a low-MWT fragment-based library than with a conventional higher MWT library [42]. Leads are less complex in most parameters than drugs, which is understandable in that medicinal chemistry optimization almost invariably increases MWT and log *P* [43]. However, there is a strong structural resemblance between starting lead and drug [44]. This implies that a quality lead as opposed to a flawed lead is more likely to result in a real drug [45].

Tool-like

The issue of what is required for a chemical tool that is capable of interrogating biology function has only recently arisen and hence what follows is the author's perspective rather than a review. The context is that of the NIH roadmap small molecule repository initiative and the initiatives in chemical genetics. A chemical tool can be used as a starting point in drug discovery – the successful tool in essence becomes the lead. A tool can also be useful in target validation even if the chemical features of the tool do not lend themselves to the chemical features of a drug. Finally, a chemical tool can be used to probe biological function without consideration of whether the biological target has any biomedical utility in a drug discovery sense. Regardless of the definition, the tool must not contain structural features that would compromise its use in interrogating protein function or elucidating pathway function. As much selectivity as possible in the compound should be a goal. This means avoiding compounds with covalent chemistry possibilities or those compounds that could, on the time scale of a tool experiment, non-selectively perturb protein function via covalent chemistry. Pharmaceutical industry filters simply state whether a structural feature is acceptable for oral drug activity (the dominant industry business plan). Useful tools require a staged set of ''tool-like'' filters in which those structural features most likely to invalidate a compound as a tool are prioritized. Those filters that directly code for covalent chemistry should be prioritized (and compounds with these features should be avoided). Filters that code for covalent chemistry due to metabolic activation or that code for longer term *in vivo* toxicity may not be so important. These types of features may not be terribly relevant in a purely

biochemical assay, or a short duration cell culture experiment. Covalent chemistry functionality must be taken in the context of structural complexity. An epoxide in an otherwise complex natural product is a lot less objectionable than an epoxide in a much lower MWT simple synthetic.

At one extreme the chemistry in the tool itself is drug-like: the chemical structure is good enough for a tool but not good enough for a drug. There may be deficiencies in some or all of those attributes required for a drug but the chemistry itself is drug-like. The main advantage here is that there is a clear path to a drug. Another advantage is that drug-like means avoiding the pitfalls of compounds with covalent chemistry liability. A disadvantage is that the breadth of commercially available chemistry space is decreased on the order of 50–80%. Another disadvantage is cost. Frequently, drug-like compounds are more expensive than non-drug-like compounds.

At the other extreme the chemistry in the tool is not drug-like. Chemistry ''flaws'' are present but they still allow the chemical to be used as a tool. For example, the presence of a moiety associated with toxicity may be present provided that the unwanted toxicity does not present on the time scale of the tool experiment. An advantage of this extreme is that commercially available chemistry space is larger. Another substantial advantage is that the interrogation of biology is unhindered by drug discovery considerations. If there are no chemistry limits imposed in relation to drug discovery, in theory any protein or RNA or DNA target can be questioned without any consideration of whether the target has any current or future relevance to human health. A disadvantage is that the path is unclear if drug discovery is an eventual goal. Another disadvantage is the lack of clarity as to whether or not chemical features will defeat the tool utility. A useful tool requires selectivity to interrogate biology. In drug discovery, despite stringent drug-like criteria, lack of selectivity resulting from target effects is distressingly often encountered. Lack of selectivity in a tool with relaxed chemistry criteria is a very real possibility.

Concluding remarks

The issue of tool-like qualities is likely to expand as the NIH roadmap small molecule repository initiative is implemented. Combinatorial compounds will probably be screened against targets that have been avoided by the pharmaceutical industry. Within the ''envelope'' of chemical genetics, proteins of unknown function will be exposed to perturbation by small organic molecules. All these initiatives are to a considerable extent ''terra incognita'' to industry, government and academia. The industry experience in the chemistry of drug-like and lead-like compounds is only partially applicable in the new arena of tool-like compounds. Historically successful explorers have been pragmatic entrepreneurs who share and use all the available information and who innovate when there is no precedent. By analogy, I suspect that the endeavors on enabling chemical tools will similarly benefit from information sharing and innovation among industry, government and academia.

References

1 Lipinski, C.A. *et al.* (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 23, 3–25

2 Oprea, T.I. (2000) Property distribution of drug-related chemical databases. *J. Comput. Aided Mol. Des.* 14, 251–264

3 Bradley, M.P. (2002) An overview of the diversity represented in commercially-available databases. *Mol. Divers.* 5, 175–183

4 Veber, D.F. *et al.* (2002) Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* 45, 2615–2623

5 Andrews, P.R. *et al.* (1984) Functional group contributions to drug–receptor interactions. *J. Med. Chem.* 27, 1648–1657

6 Fichert, T. *et al.* (2003) A structure–permeability study of small drug-like molecules. *Bioorg. Med. Chem. Lett.* 13, 719–722

7 Bemis, G.W. and Murcko, M.A. (1996) The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* 39, 2887–2893

8 Bemis, G.W. and Murcko, M.A. (1999) Properties of known drugs. 2. Side chains. *J. Med. Chem.* 42, 5095–5099

9 Oprea, T.I. *et al.* (2002) Pharmacokinetically based mapping device for chemical space navigation. *J. Comb. Chem.* 4, 258–266

10 Matter, H. *et al.* (2001) Computational approaches towards the rational design of drug-like compound libraries. *Comb. Chem. High Throughput Screen.* 4, 453–475

11 Muegge, I. *et al.* (2001) Simple selection criteria for drug-like chemical matter. *J. Med. Chem.* 44, 1841–1846

12 Muegge, I. (2003) Selection criteria for drug-like compounds. *Med. Res. Rev.* 23, 302–321

13 Brunne, R.M. *et al.* (2002) Virtual compound libraries and molecular modelling. In *Handbook of Combinatorial Chemistry*, (Vol. 2) *Handbook of Combinatorial Chemistry* (Vol. 2) (Nicolaou, K.C., Hanko, R., Hartwig, W. eds), pp. 761–783

14 Rose, S. and Stevens, A. (2003) Computational design strategies for combinatorial libraries. *Curr. Opin. Chem. Biol.* 7, 331–339

15 Patchett, A.A. and Nargund, R.P. (2000) Privileged structures – an update. *Ann. Rep. Med. Chem.* 35, 289–298 (Chapter 26)

16 DeSimone, R.W. *et al.* (2004) Privileged structures: applications in drug discovery. *Comb. Chem. High Throughput Screen.* 7, 473–493

17 Tounge, B.A. and Reynolds, C.H. (2004) Defining privileged reagents using subsimilarity comparison. *J. Chem. Inf. Comput. Sci.* 44, 1810–1815

18 Rishton, G.M. (1997) Reactive compounds and *in vitro* false positives in HTS. *Drug Discov. Today* 2, 382–384

19 Wenlock, M.C. *et al.* (2003) A comparison of physiochemical property profiles of development and marketed oral drugs. *J. Med. Chem.* 46, 1250–1256

20 Blake, J.F. (2003) Examination of the computed molecular properties of compounds selected for clinical development. *BioTechniques* (Suppl.), 16–20

21 Askenash, J. (2004) The rule of five revisited. *Preclinica* 2, 92

22 Pardridge, W.M. (2004) Log(BB), PS products and *in silico* models of drug brain penetration. *Drug Discov. Today* 9, 392–393

23 Clark, D.E. (2003) *In silico* prediction of blood–brain barrier permeation. *Drug Discov. Today* 8, 927–933

24 Kelder, J. *et al.* (1999) Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs. *Pharm. Res.* 16, 1514–1519

25 Norinder, U. and Haeberlein, M. (2002) Computational approaches to the prediction of the blood–brain distribution. *Adv. Drug Deliv. Rev.* 54, 291–313

26 Pardridge, W.M. (2003) Blood–brain barrier genomics and the use of endogenous transporters to cause drug penetration into the brain. *Curr. Opin. Drug Discov. Dev.* 6, 683–691

27 Adenot, M. *et al.* (2004) Blood–brain barrier permeation models: discriminating between potential CNS and non-CNS drugs including P-glycoprotein substrates. *J. Chem. Inf. Comput. Sci.* 44, 239–248

28 Doan, K.M.M. *et al.* (2002) Passive permeability and P-glycoprotein-mediated efflux differentiate central nervous system (CNS) and non-CNS marketed drugs. *J. Pharmacol. Exp. Ther.* 303, 1029–1037

29 Faassen, F. *et al.* (2003) Caco-2 permeability, P-glycoprotein transport ratios and brain penetration of heterocyclic drugs. *Int. J. Pharm.* 263, 113–122

30 Adenot, M. and Lahana, R. (2004) Blood–brain barrier permeation models: discriminating between potential CNS and non-CNS drugs including P-glycoprotein substrates. *J. Chem. Inf. Comput. Sci.* 44, 239–248

31 Ecker, G.F. and Noe, C.R. (2004) *In silico* prediction models for blood–brain barrier permeation. *Curr. Med. Chem.* 11, 1617–1628

32 Horspool, K.R. and Lipinski, C.A. (2003) Advancing new drug delivery concepts to gain the lead. *Drug Deliv. Technol.* 3, 34–46

33 Vieth, M. *et al.* (2004) Characteristic physical properties and structural fragments of marketed oral drugs. *J. Med. Chem.* 47, 224–232

34 Tronde, A. *et al.* (2003) Pulmonary absorption rate and bioavailability of drugs *in vivo* in rats: structure–absorption relationships and physicochemical profiling of inhaled drugs. *J. Pharm. Sci.* 92, 1216–1233

35 Oprea, T.I. *et al.* (2001) Is there a difference between leads and drugs? A historical perspective *J. Chem. Inf. Comput. Sci.* 41, 1308–1315

36 Jacoby, E. *et al.* (2003) Design of small molecule libraries for NMR screening and other applications in drug discovery. *Curr. Top. Med. Chem.* 3, 11–23

37 Lepre, C.A. *et al.* (2002) Applications of SHAPES screening in drug discovery. *Comb. Chem. High Throughput Screen.* 5, 583–590

38 Lepre, C.A. (2003) Strategies for NMR screening and library design. *Methods Princ. Med. Chem.* 16, 391–415 (BioNMR in Drug Research)

39 Carr, R. and Jhoti, H. (2002) Structure-based screening of low-affinity compounds. *Drug Discov. Today* 7, 522–527

40 Davis, A.M. *et al.* (2003) Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angew. Chem. Int. Ed.* 42, 2718–2736

41 Congreve, M. *et al.* (2003) A 'rule of three' for fragment-based lead discovery? *Drug Discov. Today* 8, 876–877

42 Hann, M.H. and Oprea, T.I. (2004) Pursuing the leadlikeness concept in pharmaceutical research. *Curr. Opin. Chem. Biol.* 8, 255–263

43 Oprea, T.I. (2002) Current trends in lead discovery: are we looking for the appropriate properties? *Mol. Divers.* 5, 199–208

44 Proudfoot, J.R. (2002) Drugs, leads, and drug-likeness: an analysis of some recently launched drugs. *Bioorg. Med. Chem. Lett.* 12, 1647–1650

45 Lipinski, C.A. (2003) Compound properties and drug quality. In *Practice of Medicinal Chemistry* 2nd edn (Wermuth, C.G. ed.), pp. 341–349