ELSEVIER

# A practical view of 'druggability'

Thomas H Keller, Arkadius Pichota and Zheng Yin

The introduction of Lipinski's 'Rule of Five' has initiated a profound shift in the thinking paradigm of medicinal chemists. Understanding the difference between biologically active small molecules and drugs became a priority in the drug discovery process, and the importance of addressing pharmacokinetic properties early during lead optimization is a clear result. These concepts of 'drug-likeness' and 'druggability' have been extended to proteins and genes for target identification and selection. How should these concepts be integrated practically into the drug discovery process? This review summarizes the recent advances in the field and examines the usefulness of 'the rules of the game' in practice from a medicinal chemist's standpoint.

**Addresses**
Novartis Institute for Tropical Diseases (NITD), 10 Biopolis Road, #05-01 Chromos, 138670 Singapore

Corresponding author: Keller, Thomas H (thomas.keller@novartis.com)

**Current Opinion in Chemical Biology** 2006, **10**:357–361

This review comes from a themed issue on
Next-generation therapeutics
Edited by Clifton E Barry III and Alex Matter

Available online 30th June 2006

1367-5931/$ – see front matter


**DOI** 10.1016/j.cbpa.2006.06.014

## Introduction

During the 1990s, the pharmaceutical industry noticed that too many compounds were terminated in clinical development because of unsatisfactory pharmacokinetics (PK) [1•]. It became clear that medicinal chemists needed to address this parameter during lead optimization and therefore tools were needed to assess the relationship between structure and PK properties.

This search for an understanding of what is responsible for compound attrition has led to the development of criteria which are characteristic for compounds that successfully pass through the development process [2••]. Such compounds have been called 'druggable' or 'drug-like' [3,4] (Box 1).

An extension of this work to protein targets that can bind such 'drug-like' compounds and therefore are thought to be amenable to modulation by compounds with oral bioavailability has led to the terms 'druggable protein' [3] and 'druggable genome' [5••] (Box 2).

Many authors have discussed these topics but important questions remain unanswered. How useful are these concepts in the daily life of a drug discovery scientist? And, what are the recent advances that make these theoretical concepts really useful in practice?

This review addresses these questions. First, we summarize the literature and clarify the useful aspects of the compound property debate with special emphasis on the past three years (2003–2006). We then examine the recent work on protein druggability from the same period. For clarity, we use the term 'druggable' only for proteins (targets) and apply the original [6], more appropriate term 'drug-like' for compounds.

## Drug-like compounds

Lipinski analyzed, in his seminal publication [6], the attrition problems of the pharmaceutical industry in the 1980s and 1990s and came to the surprising conclusion that a simple set of physicochemical parameter ranges, the 'Rule of Five' (RO5), was associated with 90% of orally active drugs that achieved phase II status [2••]. The stated goal of these rules was to guide chemists in the design and selection of compounds with appropriate physicochemical properties in order to reduce attrition during clinical development [2••].

This analysis by the Pfizer group attracted considerable attention and has been cited over 1000 times [2••]. Recently, similar examinations have come from scientists at GSK [7], Boehringer Ingelheim [8], Astra Zeneca [9], Bayer [10] and Lilly [11]. Although all of these studies analyzed slightly different sets of marketed drugs and/or compounds in clinical development, the conclusions are very similar: drug-like compounds exhibit physicochemical properties that are important for successful drug development (Box 1). For example, Wenlock and coworkers [9] nicely demonstrated the higher attrition rate of large, hydrophobic, flexible, hydrogen-bond-acceptor rich compounds in clinical development.

There is little question that the concept of drug-likeness is widely accepted among medicinal chemists. The RO5 and its extensions (Box 1) have raised the awareness that addressing PK properties early in the drug discovery process is of vital importance. Together with the implementation of *in vitro* physicochemical profiling assays this has significantly reduced the attrition rate due to adverse PK and poor bioavailability [1•].

Nevertheless, it is important to emphasize the limitations of these rules:

**Box 1** Drug-like compounds

Rule of 5: [6]
- MW $\leq$500
- ClogP $\leq$5
- H-bond donors $\leq$5
- H-bond acceptors (sum of N and O atoms) $\leq$10

Remarks: No more than one violation; not applicable for substrates of transporters and natural products

Extensions: [7]
- Polar surface area $\leq$140 $A^2$ or Sum of H-bond donors and acceptors $\leq$12
- Rotatable bonds $\leq$10

1. RO5 applies only to compounds that are delivered by the oral route.
2. RO5 applies only to compounds that are absorbed by passive mechanisms [6].
3. There are important exceptions (see below).
4. RO5 compliant compounds are not automatically good drugs [2••].

It is also important to realize that drug-likeness is only a useful concept in connection with biological activity [12]. Drug-like compounds must contain enough functionality to interact in a meaningful way with a protein [10]. This requirement eliminates simple compounds (like hexane or benzene) that at first glance adhere to RO5.

## Exceptions to the RO5

One practically important exception was described by the GSK group [7]. They demonstrated that compounds with molecular weight (MW) >500 but with reduced molecular flexibility and constrained polar surface area may also show good oral bioavailability [2••].

Natural products are another important exception to the RO5. This is disconcerting as fungi, bacteria and plants have provided a number of very successful oral drugs [13]. Clardy and Walsh have highlighted the structural characteristics that make these compounds special. They are of high complexity and rich in stereogenic centers, and what is especially striking compared with synthetic drugs, they rarely contain nitrogen [13]. It is not clear why complex natural product-based drugs violate property-based rules, but it is unlikely that this is due to structural parameters (structural diversity, chiral centers, atomic composition [14]). It is much more likely that over the millennia of evolution these compounds have been optimized by nature to take advantage of active transport or have developed special conformational features that are beneficial for passive transport.

**Box 2** Druggability

*Druggable genome*: Genes that encode disease related proteins that can be modulated by drug-like molecules [5••].

*Druggable protein*: Proteins that can bind drug-like compounds with binding affinity below 10 $\mu$M (it is inferred that the compound must be able to functionally modulate the protein).

*Pharmaceutically tractable genome*: Genes that encode proteins which can be targeted by small molecular weight compounds, antibodies and recombinant therapeutic proteins for pharmaceutical use [19].

## Lead-like compounds

Analyses of drug discovery projects have shown that chemists tend to increase molecular weight and lipophilicity during lead optimization [15]. This has led to the thinking that RO5-compliant drug candidates would be easier to attain if lead finding libraries would only contain small compounds (Box 1) [16]. Such libraries have been called 'lead-like'.

It is debatable whether such collections are necessary in the current HTS-driven environment [17•], especially with the increasing awareness of drug-like properties in the medicinal chemistry community. Nevertheless, such libraries will be useful for emerging lead identification technologies such as fragment-based screening by NMR [18] or X-ray crystallography [19].

## Druggability of proteins

In 2002, Hopkins and Groom introduced the concept of the 'druggable genome' [5••]. Their purpose was to identify the limited set of molecular targets for which commercially viable, oral compounds can be developed. Because such targets are expected to bind RO5-compliant compounds, they analyzed databases and used computational methods to identify all proteins belonging to families which have at least one member that has successfully been targeted by drug-like molecules. Assuming that druggability is shared among protein families and taking into account that, by necessity, a drug target needs to have the potential to be disease-modifying, they estimated that the human genome encodes 600–1500 targets for small-molecule drugs [5••].

The findings of this pioneering publication have stimulated considerable discussion in the pharmaceutical industry as they imply that the number of protein targets for commercially viable compounds is considerably smaller than originally thought. It was pointed out that such an analysis on the gene level is too simplistic because alternative splicing and post-translational modifications lead to a considerably larger druggable proteome [17•], and additionally biologicals have become an important class of therapeutics. Therefore the size of the Pharmaceutically Tractable Genome, a term that was suggested for genes that can be targeted by small molecular weight compounds, antibodies and recombinant therapeutic proteins [20] is considerably larger. However, the key point in Hopkins and Groom's paper, that the target pool which can be addressed with oral drugs is limited and relatively small, has remained correct [21,22].

The weakest point of a genome level analysis is that only qualitative druggability estimates can be derived, whereas in practice more detailed information about individual protein targets is desired. For example, researchers working on a protease may not be satisfied by knowing that the protease class is generally druggable but may rather want to know whether their specific target can be inhibited by drug-like compounds.

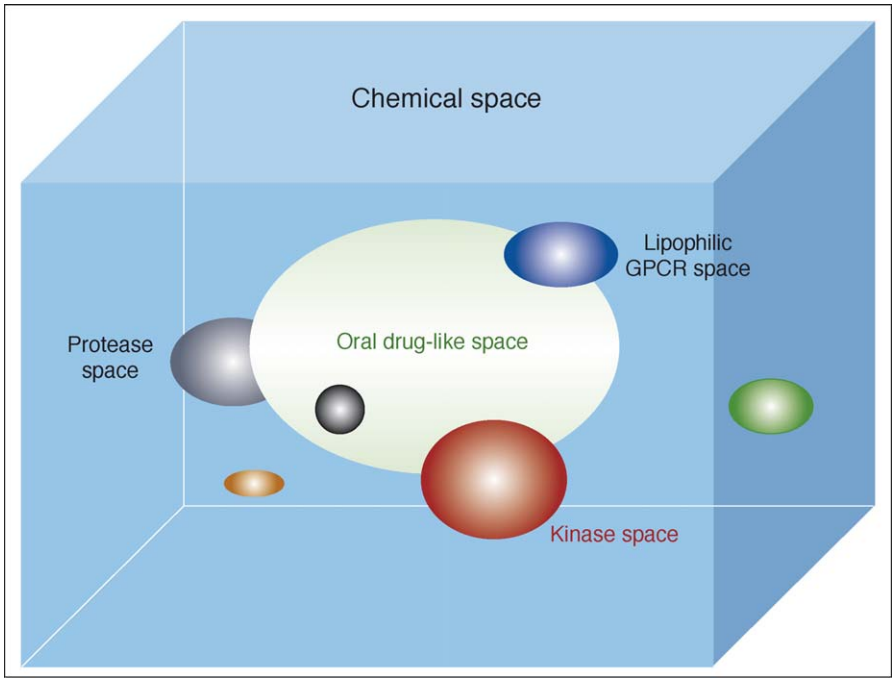
This is best explained using the cartoon representation of chemical space introduced by Lipinski and Hopkins (Figure 1) [23•] where compounds are mapped onto coordinates of chemical descriptors of physicochemical or topological properties. For example, it is known that active site inhibitors for protease families cluster together in a discrete region of chemical space [23•]. The intersection of this cluster with the descriptor space for drug-like compounds would contain in this example the protease inhibitors that have the potential to be developed as oral drugs [23•]. Such an overlap would allow the conclusion that the protease family is druggable.

The most challenging part of a protease inhibitor optimization is changing a peptidic lead compound into an orally bioavailable peptidomimetic. This step has proven to be relatively straightforward for some proteases (e.g. thrombin) [24], whereas for others it may be an almost impossible job due to the amino acid content of the lead and the shallowness of the binding pocket [25]. The latter case would probably constitute an exception in the generally druggable protease family (Figure 2). Such subtleties create uncertainties about the reliability of druggability arguments for strategic decision making, especially in cases of well validated targets with borderline druggability.

A first step towards a more reliable way to assess the druggability of individual proteins is the identification of binding sites for drug-like molecules [26•]. Kellenberger and co-workers have created an annotated database of ligand binding sites extracted from several experimental structure databases [27]. Such data can be used to derive rules or training sets for the computational identification of binding pockets. Many algorithms [28] are available for this purpose and in general they have been successfully applied in identifying true ligand binding pockets on the surface of proteins [26•] (see also Update).

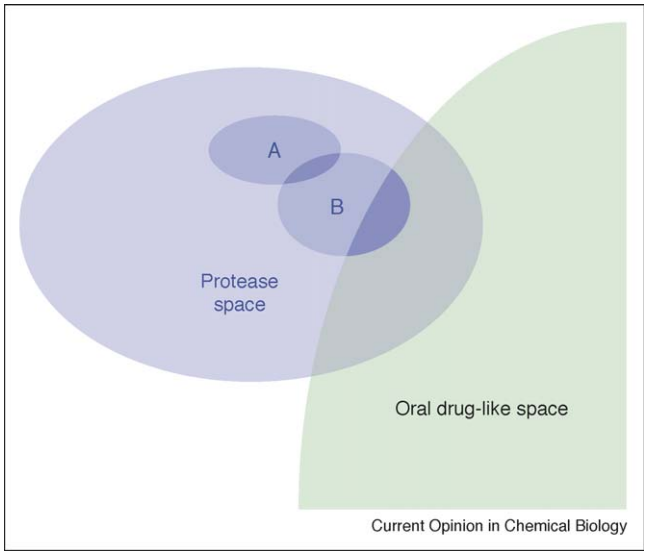
The second step of quantitatively assessing the druggability of the identified pockets is more challenging. An obvious approach is to screen a large library of drug-like compounds (pharmaceutical compound collection or a chemical genetics library [29]) and assess the resulting hits. Unfortunately, this approach has three significant drawbacks: it is very expensive, applied rather late in the drug discovery process and it produces a large number of

**Figure 1**




Cartoon representation of the relationship between the continuum of chemical space (light blue) and the discrete areas occupied by compounds with specific affinity for certain protein classes. The independent intersection of compounds with drug-like properties is shown in green. Reprinted with permission from [23•]. Copyright 2004, Macmillan Publishers Ltd.

**Figure 2**




A close-up of the cartoon in Figure 1 that shows the chemical space occupied by active site serine protease inhibitors. Since this bubble intersects with the oral drug-like space, we would come to the conclusion that the serine protease family is druggable. The situation is, however, more complex. Some serine proteases will be druggable (e.g. B) whereas others are not (e.g. A).

false positives and promiscuous hits which complicate the analysis [26•,30].

Another method has recently been developed by researchers from Abbott [26•,31••]. Using 2-D heteronuclear-NMR they studied the interactions of 10 000 lead-like or fragment-like compounds with protein surfaces. This approach has the advantage that it samples a large fraction of chemical space (even though the size of the library is small) and yields more reliable data than conventional high-throughput screening. Furthermore, such an NMR-based analysis of druggability could be performed with limited resources and relatively early in the drug discovery process. Most importantly, an analysis of the NMR data has led to the development of 'druggability-indices' that can be used for the computational assessment of proteins with known structure [26•,31••,32].

Such quantitative assessments of druggability would find wide application if it were not for the limited availability of experimental 3-D structures [33]. Despite the progress that has been made in structural biology, especially with the structural genomics approaches, only a small fraction of all proteins have been experimentally characterized [33]. As a consequence, most structure-based assessments of target druggability still need to be performed with homology models [33]. Unfortunately the predictive quality of homology models and therefore their usefulness is rather uncertain since often no closely related protein with known 3-D structure is available [34].

## Conclusion

The RO5 and its extensions (Box 1) have been useful tools to generate awareness about the importance of PK parameters for development. In addition, this concept has led to the realization that there may be whole families of proteins for which it is either extremely challenging or impossible to design compounds with good oral bioavailability.

The available evidence suggests that qualitative druggability arguments are useful strategic tools; however, more accurate, quantitative assessments are needed, especially for proteins with borderline druggability. Recent developments suggest that NMR-based screening could deliver such information. In our view, this is a very attractive approach that can be used in the early stages of drug discovery and provides a solid basis for computational druggability estimations [26•,31••].

It has been suggested that the inability of the pharmaceutical industry to solve druggability problems is due to limitations in current medicinal chemistry approaches [35]. As outlined above, druggability is defined by molecular properties and therefore it is unlikely that advances in synthesis, profiling or innovative design will solve these problems [36]. The much-quoted advances in this area (e.g., finding inhibitors for protein–protein interactions) [35,36] rarely address the real problems: PK and especially oral bioavailability. It is often possible to find inhibitors of proteins with questionable druggability [35–37], but this is only the first step. The real druggability challenge arrives when these 'leads' have to be turned into orally bioavailable, pharmaceutically useful drug candidates. It has never been challenging to make inhibitors of proteins with questionable druggability, but rather the challenge has been to make orally bioavailable, pharmaceutically useful inhibitors that successfully advance through clinical development. It is likely that oral drugs for the modulation of such proteins will continue to come from the natural-product pool. In addition, future advances in drug delivery may offer solutions to address problems of druggability [38].

## Update

A recent review summarizes computational methods to identify protein binding pockets for small drug-like compounds. Classical geometric and energy-based computational methods are discussed, with particular focus on two powerful technologies: computational solvent mapping and grand canonical Monte Carlo simulations [39].

## References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. • Kola I, Landis J: **Can the pharmaceutical industry reduce attrition rates?** *Nat Rev Drug Discov* 2004, **3**:711-715.

An excellent review of industry attrition rates over the years. The authors discuss possible causes and remedies. Interestingly, the impact of druggability on these attrition rates is not discussed.

2. •• Lipinski CA: **Lead- and drug-like compounds: the rule-of-five revolution**. *Drug Discov Today: Technologies* 2004, **1**:337-341.
A recent summary of Lipinski's druggability work. The analysis is easily accessible and should give scientist a nice entry to the subject.

3. Sugiyama Y: **Druggability: selecting optimized drug candidates**. *Drug Discov Today* 2005, **10**:1577-1579.

4. Sirois S, Hatzakis G, Wei D, Du Q, Chou KC: **Assessment of chemical libraries for their druggability**. *Comput Biol Chem* 2005, **29**:55-67.

5. •• Hopkins AL, Groom CR: **The druggable genome**. *Nat Rev Drug Discov* 2002, **1**:727-730.
This paper defined and used the term ''protein druggability'' for the first time. It is easy to read and provides an essential basis to understand the various aspects of protein druggability.

6. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ: **Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings**. *Adv Drug Deliv Rev* 2001, **46**:3-26.

7. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD: **Molecular properties that influence the oral bioavailability or drugs**. *J Med Chem* 2002, **45**:2615-2623.

8. Fichert T, Yazdanian M, Proudfoot JR: **A structure-permeability study of small drug-like molecules**. *Bioorg Med Chem Lett* 2003, **13**:719-722.

9. Wenlock MC, Austin RP, Barton P, Davis AM, Leeson PD: **A comparison of physicochemical property profiles of development and marketed oral drugs**. *J Med Chem* 2003, **46**:1250-1256.

10. Muegge I: **Selection criteria for drug-like compounds**. *Med Res Rev* 2003, **23**:302-321.

11. Vieth M, Siegel MG, Higgs RE, Watson IA, Robertson DA, Savin KA, Durst GL, Hipskind PA: **Characteristic physical properties and structural fragments of marketed drugs**. *J Med Chem* 2004, **47**:224-232.

12. Lajiness M, Vieth M, Erickson J: **Molecular properties that influence oral drug-like behavior**. *Curr Opin Drug Discov Devel* 2004, **7**:470-477.

13. Clardy J, Walsh C: **Lessons from natural molecules**. *Nature* 2004, **432**:829-837.

14. Feher M, Schmidt JM: **Property distributions: differences between drugs, natural products and molecules from combinatorial chemistry**. *J Chem Inf Comput Sci* 2003, **43**:218-227.

15. Oprea TI: **Current trend in lead discovery: are we looking for the appropriate properties?** *Mol Divers* 2002, **5**:199-208.

16. Rishton GM: **Nonleadlikeness and leadlikeness in biochemical screening**. *Drug Discov Today* 2003, **8**:86-96.

17. • Kubinyi H: **Drug research: myths, hype and reality**. *Nat Rev Drug Discov* 2003, **2**:665-668.
A well-argued opinion piece about the realities of drug discovery research. It provides an interesting counterpoint to the Hopkins and Groom [5••] arguments about protein druggability.

18. Schuffenhauer A, Ruedisser S, Marzinzik AL, Jahnke W, Blommers M, Selzer P, Jacoby E: **Library design for fragment based screening**. *Curr Top Med Chem* 2005, **5**:751-762.

19. Carr RAE, Congreve M, Murray CW, Rees DC: **Fragment-based lead discovery: leads by design**. *Drug Discov Today* 2005, **10**:987-992.

20. Gould Rothberg BE, Peña CEA, Rothberg JM: **A systems biology approach to target identification and validation for human chronic disease drug discovery**. In *Modern Biopharmaceuticals*, **1**. Edited by Knäblein J. Wiley-VCH Verlag; 2005:99-125.

21. Russ AP, Lampel S: **The druggable genome: an update**. *Drug Discov Today* 2005, **10**:1607-1610.

22. Orth AP, Batalov S, Perrone M, Chanda SK: **The promise of genomics to identify novel therapeutic targets**. *Expert Opin Ther Targets* 2004, **8**:587-596.

23. • Lipinski C, Hopkins A: **Navigating chemical space for biology and medicine**. *Nature* 2004, **432**:855-861.
Interesting examination of chemical space and the chances of finding small molecules to modulate biological targets. The authors discuss the strategic differences between the use of compounds as biological tools and compound as drugs.

24. Hopkins AL, Groom CR: **Target analysis: a priory assessment of druggability**. *Ernst Schering Res Found Workshop* 2003, **42**:11-17.

25. Erbel P, Schiering N, D'Arcy A, Renatus M, Kroemer M, Lim SP, Yin Z, Keller TH, Vasudevan SG, Hommel U: **Structural basis for the activation of flaviviral NS3 proteases from dengue and west Nile virus**. *Nat Struct Mol Biol* 2006.

26. • Hajduk PJ, Huth JR, Tse C: **Predicting protein druggability**. *Drug Discov Today* 2005, **10**:1675-1682.
This review discusses the determination of druggability-indices by NMR and their application for the computational assessment of proteins with experimental 3D structures. It is easier to read than the rather original, technical publication by the same authors [30].

27. Kellenberger E, Muller P, Schalon C, Bret G, Foata N, Rognan D: **Sc-PDB: an annotated database of druggable binding sites from the protein data bank**. *J Chem Inf Model* 2006, **46**:717-727.

28. An J, Totrov M, Abagyan R: **Comprehensive identification of druggable protein ligand binding sites**. *Genome Inform Ser Workshop Genome Inform* 2004, **15**:31-41.

29. Darvas F, Dorman G, Puskas LG, Bucsai A, Urge L: **Chemical genomics for fast and integrated target identification and lead optimization**. *Med Chem Res* 2004, **13**:643-659.

30. Lipinski CA: **Filtering in drug discovery**. *Ann Rep Comp Chem* 2005, **1**:155-168.

31. •• Hajduk PJ, Huth JR, Fesik SW: **Druggability indices for protein targets derived from NMR-based screening data**. *J Med Chem* 2005, **48**:2518-2525.
Excellent, rather technical, publication about the experimental determination of druggability by NMR. The authors develop druggability indices that can be used for the *de novo* assessment of proteins.

32. Brown SP, Hajduk PJ: **Effect of conformational dynamics on predicted protein druggability**. *Chem Med Chem* 2006, **1**:70-72.

33. Hillisch A, Pineda LF, Hilgenfeld R: **Utility of homology models in the drug discovery process**. *Drug Discov Today* 2004, **9**:659-669.

34. Kairys V, Fernandes MX, Gilson MK: **Screening drug-like compounds by docking to homology models: a systematic study**. *J Chem Inf Model* 2006, **46**:365-379.

35. Haberman AB: **Strategies to move beyond target validation**. *Genet Eng News* 2005, **25**:36.

36. Betz UAK: **How many genomic targets can a portfolio afford?** *Drug Discov Today* 2005, **10**:1057-1063.

37. Arkin MR, Wells JA: **Small-molecule inhibitors of protein-protein interactions: progressing towards a dream**. *Nat Rev Drug Discov* 2004, **3**:301-317.

38. Horspool KR, Lipinski CA: **Advancing new drug delivery concepts to gain the lead**. *Drug Delivery Tech* 2003, **3**:34-46.

39. Vajda S, Guarnieri F: **Characterization of protein-ligand interaction sites using experimental and computational methods**. *Curr Opin Drug Discovery Dev* 2006, **9**:354-362.