



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

CHAPTER 11

Filtering in Drug Discovery

Christopher A. Lipinski

Pfizer Global Research and Development, Groton, CT 06340, USA

Contents

1. Drug-likeness	155
2. Drug-likeness and chemistry quality	157
3. Positive desirable chemistry filters	158
4. Lead-likeness	159
5. Oral drug activity	159
6. CNS drugs	160
7. Intestinal permeability	161
8. Aqueous solubility	162
9. Drug metabolism	162
10. Promiscuous compounds	162
11. Agrochemicals	163
References	163

Recent advances in filtering in drug discovery are reviewed. Filtering is used in a global broad sense to include exclusionary as well as inclusionary criteria and encompasses the following topics related to the discovery of drugs: (1) computational definitions of drug-likeness; (2) positive filters for drug-like activity; (3) lead-likeness as a concept; (4) oral activity filters; (5) CNS drug filters; (6) intestinal permeability filters; (7) drug metabolism parameter filters; (8) promiscuous compound filters; and (9) agrochemical filters. This review does not cover what could be termed as local parameters such as pharmacophore models, docking and scoring, etc.

1. DRUG-LIKENESS

To this author there seems to be much more agreement as to what is drug-like than there is as to what is diverse. Defining what is drug-like and non-drug-like requires some type of reference point. The property distributions of commercially available databases have been examined. The available chemicals directory (ACD) seems to be the most common non-drug-like database. The pesticide manual has been used as an alternate standard for non-drug-likeness because it

is composed primarily of compounds designed to cause fatality to the primary organism. 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 [1,2]. In addition, the 10,000 or so Phase II compounds are used to define drug-like compounds [3]. A compound's drug-like index has been calculated based upon the knowledge derived from known drugs selected from the CMC database [4]. The property distributions in combinatorial compounds compared to drugs or natural products largely reflect combinatorial chemistry synthesis constraints such that there are fewer chiral centers and complex ring systems [5]. The distribution of ring systems across multiple databases has been described [6] and a program was written and tested on the MDDR database [7] to identify candidate chemical ring replacements (bioisosteres). 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 [8]. The diversity that side chains provide to drug molecules is quite low since only 20 side chains account for over 70% of the side chains [9]. 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 [10].

Multiple filters (properties) may be incorporated into a definition of drug-likeness and this leads to trade-offs among compound properties in compounds intended for screening [11]. Optimization of compound properties may require some type of multi-parameter optimization scheme in library design [12]. Fingerprint algorithms can be used to guide diversity [13]. Filters also need to be employed in the chemistry synthesis planning process so that good quality compounds are made [14]. Differences in property ranges between oral and injectable drugs have been summarized [15]. Oral drugs are lower in MWT and have fewer H-bond donors, acceptors and rotatable bonds. 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. Polar surface area (PSA) in one definition is the solvent accessible surface covered by oxygen, nitrogen and the hydrogens attached to oxygen and nitrogen. As a compound progresses through clinical trials there is a steady change in properties, e.g., MWT, Log *P* and PSA all decline with a MWT of about 340 found for marketed drugs [16,17]. The reason for this pattern is unclear since properties related to oral absorption would be expected to have reached a plateau by Phase II and hence selection pressure for properties related to oral absorption should have disappeared by then [18]. Pulmonary drugs tend to have higher PSA because pulmonary permeability is less sensitive to polar

hydrogen-bonding functionality [19]. Anatomically this makes sense since lungs are a closed compartment and any accumulating fluid and compounds in terminal alveoli must be cleared. Discrimination between antibacterial and non-antibacterial activity has been achieved based on 3D molecular descriptors. The overall classification rate was around 90% on a data set of 661 compounds using 2–3 variables selected from log P , charged-weighted negative surface area, positive surface area of heavy atoms and maximum donor delocalizability. Three-dimensional geometry variations had little impact on the discriminatory performance [20].

2. DRUG-LIKENESS AND CHEMISTRY QUALITY

Descriptors for drug-like are most effective if they have physical meaning so as to facilitate chemists designing in drug-likeness [21]. Drug-likeness in the design of combinatorial libraries [22,23] involves the use of rule-based filters like the rule of 5 [24], the use of exclusionary filters to remove undesired chemistry functionality [25] and the capture of privileged structure information, e.g., from natural product collections [26] or from retro synthetic analysis of collections of bioactive molecules [27]. Natural product structural features are particularly well represented in the cancer chemotherapy and infectious disease areas [28]. Exclusionary filters have been described that remove reactive chemical functionality based on the premise that compounds having covalent chemistry possibilities have no place in drug discovery [29]. Filters are also necessary to remove cross reactivity in pooled compounds [30]. Pooling is a procedure in which single well-characterized compounds are deliberately mixed to speed screening. Components of the mixture must neither contain structural features causing assay false positives nor must they contain common substructural elements that would confuse the deconvolution of activities of the individual components. The magnitude of the number of poor quality screening compounds is emphasized by the report that only 37% of 1.6 million unique commercially available compounds are drug-like [31]. A very similar result was found in a virtual screen for SARS-CoV protease against commercially available and academic compounds. Of the 0.07% virtual hits against 3.6 million compounds, 47% failed three or more of 13 druggability criteria [32]. The criteria were based on physical, chemical and structural properties. Providing high-quality chemistry subject matter is now supported under the NIH molecular library small molecule repository initiative which aims to collect one million drug-like molecules from commercial, industry and government sources [33]. Emphasizing the point that drugs must contain adequate functionality to achieve acceptable receptor interactions, a single filter separates drug-like from non-drug-like compounds based on the observation that non-drugs are often under-functionalized [34].

3. POSITIVE DESIRABLE CHEMISTRY FILTERS

Privileged structures, e.g., 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 sources for screening libraries and have recently been reviewed [35]. Privileged structure features have been employed in the combinatorial design of GPCR libraries [36], in the combinatorial synthesis of privileged bicyclic structures [37] and in the combinatorial synthesis of cyclic peptides [38]. Homology modeling suggests a parallelism between common privileged GPCR ligand features and complementary deeply buried protein features in class A GPCRs [39]. Grouping by target family is also another method helping focus on particular target-directed privileged structures [40]. The idea is that structurally similar target family members will bind structurally similar small molecule ligands [41]. NMR screening helps identify privileged protein binding elements albeit of smaller size [42]. Although not strictly speaking a privileged structure, privileged structural elements such as the hydroxamate moiety found in many metalloprotease inhibitors can be identified [43]. Discernment of privileged structures has historically largely been a data mining exercise. However, very similar recurring structural motifs, so-called 'molecular anchors' have been described based on structure-based ligand binding considerations [44,45]. Rigid small molecule ligands (the molecular anchors) are incapable of hydrophobic collapse and a single non-collapsible ligand conformer binds at a protein cavity site which is also often incapable of hydrophobic collapse. This concept explains the frequent occurrence of non-collapsible spiro structures in privileged structures/molecular anchors. Chemistry design principles directed to the very difficult goal of small molecule interference with protein-protein interactions *via* an allosteric interaction have been described [46]. An intriguing aspect is the hypothesis that chemistry emphasis should be placed on compound cores capable of interacting with relatively fixed protein hinge regions rather than on elaboration of lipophilic side chains attached to the core. The thermodynamic penalty attendant to ligand binding to a non-lowest energy protein conformer suggests that screening should allow for slow binding with adequate assay equilibration time. An implication is that for this type of target it is better to make larger numbers of smaller libraries than fewer numbers of large libraries. This trend to smaller libraries is now well documented [47]. Taken to its extreme this approach takes the typical dense chemistry space coverage of the traditional combinatorial library (target-oriented synthesis) towards the direction of the diversity-oriented synthesis approach to chemistry lead generation which populates diverse single molecules broadly through chemistry space [48]. This direction is of course in the direction of less efficient, more difficult chemistry. The focus on biological information content richness suggests natural products as combinatorial library starting points [49]. Chemical content richness is found in

compounds produced by multicomponent reactions (MCR) which are chemical transformations in which as many as four components form a new compound in a single chemistry reaction step. The Ugi reaction of a carboxylic acid, amine, aldehyde and isonitrile is a classic example. In theory while offering an efficient approach to synthesis of diverse compounds, MCR in high-throughput mode currently suffers from significant chemistry limitations [50].

4. LEAD-LIKENESS

The difference between drug-like and lead-like has been described [51]. Leads are less complex in most parameters than drugs, which is understandable in that medicinal chemistry optimization almost invariably increases MWT and Log *P* [52]. However, the structural resemblance between a starting lead and a drug is marked [53]. The implication is that a quality lead as opposed to a flawed lead is far likely to lead to a real drug [54]. Lead-like discovery also refers to the screening of small molecule libraries with detection of weak affinities in the high micromolar to millimolar range. The process usually by itself does not lead to an acceptable chemistry starting point. Something else has to be added after the primary screen. Generally, multiple small molecules do not bind to non-adjacent target sites [55], so the screening is that of small MWT singletons. However, binding of two components to the same receptor site is possible as attested by the discovery of sub-nanomolar ligands in what is termed click chemistry [56]. In this process an acetylene and azide terminus from two receptor site independently bound molecules cyclize to a single compound with the two components linked *via* a 1,2,3 triazole ring. Filtering in the context of lead-like small molecule screening implies control of the properties of drug starting points that eventually result from this process. A rule of three [57] has been coined for small molecule fragment screening libraries; MWT < 300; Log *P* < 3; H-bond donors and acceptors < 3 and rotatable bonds < 3. Small fragment screening can be by NMR [58–60], by X-ray [61,62], or in theory by any method capable of detecting weak interactions.

5. ORAL DRUG ACTIVITY

The topic of filtering in human therapeutic drug discovery has received numerous frequent reviews [23,63–65] as well as criticism if fundamental medicinal chemistry principles are neglected [66]. The ‘rule of 5’ describes four simple parameters associated with improved prospects for oral activity. Poor solubility or poor permeability are more likely if there are > 5 H-bond donors (expressed as sum of OH and NH); > 10 H-bond acceptors (expressed as sum of O + N); MWT > 500 and Log *P* > 5. There are only four rules. The 5 in rule of 5 arises

from the frequent appearance of a 5 in the cutoff parameters. Compound classes such as natural products, infectious disease drugs, etc. where transporter affinity is prevalent are exceptions [24]. Rotatable bond count is now a widely used filter following the finding that greater than 10 rotatable bonds correlates with decreased rat oral bioavailability [67]. The mechanistic basis for the rotatable bond filter is unclear since 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 since ligand affinity on average decreases 0.5 kcal for each two rotatable bonds [68]. Compounds indexed in medicinal chemistry journals show the recent trend towards poor properties. Over 50% of medicinal chemistry compounds with activities above 1 nM have $MW > 425$, $\text{Log } P > 4.25$ and $\text{Log } Sw < -4.75$, indicating that these compounds are larger, more hydrophobic and less soluble when compared to time-tested quality leads [52]. The concept of the importance of compound properties (e.g., rule of 5 compliance) beyond potency is widely accepted [69]. although there are notable occasional exceptions where an orally bioavailable compound is found that lies well outside the rule of 5 limits [70]. Can the rule of 5 be bypassed by delivering drug by a non-oral route, e.g., pulmonary, intra-nasal or dermal? The answer depends very much on the dose. If the total dose is 20 mg or less then alternative delivery routes begin to be feasible. However, a limitation is that only about 10% of current clinical candidates have sufficient potency in the 0.1 mg/kg range to result in such a low dose and finding such very potent compounds seems to be mostly a matter of luck [71]. Beyond chemistry-based features, oral drugs can also be defined by their biological target. It is striking that the 100 best selling (mostly oral) drugs are ligands for proteins encoded by only a very small subset of genes and that a very considerable portion of the targets for orally active drugs may have already been discovered [72]. The term 'druggable genome' has been coined to describe the severe restriction that chemistry considerations related to oral activity superimpose on possible biology target space [73].

6. CNS DRUGS

A scheme for separating CNS from non-CNS active drugs in the WDI allowed discovery of simple parameters relating to passive blood brain barrier (BBB) permeability and prediction of p-glycoprotein (PGP) affinity [74]. The PGP transporter is a major barrier to the entry of compounds to the CNS [75]. Appropriately determined PGP efflux ratios can be used as a measure of compound affinity to PGP. However, the value of filters based on PGP efflux ratios from the commonly used high-throughput mode Caco-2 colonic cell permeability cell culture assay have been questioned as efflux ratios do not correlate with *in vivo* rat brain penetration [76]. A PSA value of less than 60–70 Å² tends to identify

CNS active compounds [77]. 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 5, 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 [78]. More complex commercially available software programs have been compared as to their ability to predict CNS log BBB ratio [79]. Experimental and theoretical reasons support the belief that surface tension measurements can be predictors for blood brain permeability [80]. Predictors for absorption, distribution, metabolism and excretion (ADME) currently appear most useful in global models. Limitations in local models likely reflect a lack of quality experimental data sets [81] and user dissatisfaction may result from unrealistic expectations given the magnitude of experimental ADME errors [82]. An additional limitation to schemes for separating CNS from non-CNS compounds is the complexity of the BBB. Compounds with affinity to transporters are exceptions to physicochemically based filters like the rule of 5. This is a problem for the CNS since it is estimated that about 15% of all genes selectively expressed at the BBB encode for transporter proteins and that only about 50% of BBB transporters are currently known [83].

7. INTESTINAL PERMEABILITY

PSA in rather simple models is a commonly used parameter to predict intestinal permeability [84]. Its rule-based calculation (TPSA) is very fast and does not require 3D structure [85]. A better prediction of intestinal permeability has been reported when PSA is partitioned into smaller molecularly based components [86]. Using molecular surface properties compounds selected from the World Health Organization's (WHO) list of essential drugs could be classified with 87% accuracy as to permeability and solubility using a six bin scheme similar to that in the FDA biopharmaceutical classification system [87]. Pharmacokinetic parameters including permeability can also be generated for filtering or ligand affinity prediction through the Volsurf software [88]. An analysis of small drug-like molecules suggests a filter of $\log D > 0$ and < 3 enhances the probability of good permeability [89]. A collection of 222 commercially available drugs was used to determine the exclusion criteria that differentiate poorly absorbed drugs from well-absorbed drugs. Similar to the rule of 5, $MWT < 500$ and $\log P < 5$ were associated with better absorbed compounds. Exceptions to the MWT criteria were compounds with a sugar moiety, high atomic weight and large cyclic structure [90] suggesting the involvement of absorptive biological transporter systems. Based on the intestinal absorption of 158 drug and drug-like compounds in rats there is a significant relationship between rat intestinal absorption, and by extrapolation human absorption, to drug hydrogen-bond acidity and basicity [91].

8. AQUEOUS SOLUBILITY

Poor aqueous solubility is a wide spread problem in combinatorial libraries as opposed to poor intestinal permeability which is much less of a problem. About one-half of poor solubility is due to large size/lipophilicity. Log $P > 5$ identifies 75% of these compounds. The other 50% of poor aqueous solubility is due to crystal packing considerations for which there is no computational filter [92]. Melting point is an experimental indicator of crystal packing. Aqueous solubility decreases about $10 \times$ for each 100°C rise in melting point and so melting point, if available, is a valuable parameter in solubility prediction [93]. Progress toward a computational melting point is suggested by the 63% success in qualitative ranking of compounds into low-medium and high-solubility bins. Descriptors for hydrophilicity, polarity, partial atom charge and molecular rigidity were found to be positively correlated with melting point whereas non-polar atoms and high flexibility within the molecule were negatively correlated [94].

9. DRUG METABOLISM

Volume of distribution (VD) is a key pharmacokinetic parameter. A low VD of less than 1 l/kg identifies drugs residing in the plasma compartment. A VD greater than 1 l/kg identifies compounds accessing tissue compartments outside the plasma compartment, e.g., many CNS drugs have VD values in the tens or higher. A recently developed computational approach to predict VD for neutral and basic drugs works as well as the *in vivo* experimental measurement provided that accurate experimental compound log D and pK_a are available. Predictivity is retained if computed log D and pK_a are used but accuracy declines somewhat [95,96]. Approximately 80% of drugs are oxidized by the cytochrome P450 (CYP) family of enzymes; hence a decision tree for CYP substrate affinity is important. This has been described in that characteristics of CYP substrates, such as lipophilicity, MWT and hydrogen-bonding potential, govern selectivity towards individual CYPs [97].

10. PROMISCUOUS COMPOUNDS

Compounds with a marked propensity to bind to multiple targets, so-called nuisance compounds, are of little value in drug discovery. Such compounds can be experimentally identified by their binding to fetal calf serum [98]. It has long been known that compounds could be identified as reproducible actives in HTS screens that could not be optimized in chemistry. Such compounds often appear active in multiple screens that have no biological relationship. An analysis of such

promiscuous compounds from HTS hits led to the conclusion that colloidal aggregates in the 50–1000 nm size range were responsible. The apparent HTS screen activity was due to a biophysical effect rather than due to a normal ligand receptor affinity and hence the hits were unoptimizable in chemistry [99]. This promiscuous aggregation effect was found among 8 of 15 kinase inhibitors widely used in biology screening [100] emphasizing the importance of exclusionary filters to prevent wasting of biology research time by testing compounds with flawed properties. The aggregation phenomenon has been found among known drugs, albeit only when tested at high non-physiological concentrations and a predictive model was developed [101].

11. AGROCHEMICALS

Filtering has also been applied to agrochemicals. Compared to drugs intended for human use, agrochemicals tend to have fewer hydrogen-bond donors [102]. For agrochemical screening computationally intensive surface area parameters offered no advantage over the rule of 5 [103]. Analogous to drug-likeness, agrochem-likeness for large compound collections has been explored using support vector machines (SVM). In this study SVM performed better than neural networks [104].

REFERENCES

- [1] T. I. Oprea, Property distribution of drug-related chemical databases, *J. Comput.-Aided Mol. Des.*, 2000, **14**, 251–264.
- [2] M. P. Bradley, An overview of the diversity represented in commercially-available databases, *Mol. Divers.*, 2002, **5**, 175–183.
- [3] C. A. Lipinski, Drug-like properties and the causes of poor solubility and poor permeability, *J. Pharmacol. Toxicol. Methods*, 2001, **44**, 235–249.
- [4] J. Xu and J. Stevenson, Drug-like index: a new approach to measure drug-like compounds and their diversity, *J. Chem. Inf. Comput. Sci.*, 2000, **40**, 1177–1187.
- [5] M. Feher and J. M. Schmidt, Property distributions: differences between drugs natural products, and molecules from combinatorial chemistry, *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 218–227.
- [6] X. Q. Lewell, A. C. Jones, C. L. Bruce, G. Harper, M. M. Jones, I. M. McLay and J. Bradshaw, Drug rings database with web interface. A tool for identifying alternative chemical rings in lead discovery programs, *J. Med. Chem.*, 2003, **46**, 3257–3274.
- [7] R. P. Sheridan, The most common chemical replacements in drug-like compounds, *J. Chem. Inf. Comput. Sci.*, 2002, **42**, 103–108.
- [8] G. W. Bemis and M. A. Murcko, The properties of known drugs. 1. Molecular frameworks, *J. Med. Chem.*, 1996, **39**, 2887–2893.
- [9] G. W. Bemis and M. A. Murcko, Properties of known drugs. 2. Side chains, *J. Med. Chem.*, 1999, **42**, 5095–5099.
- [10] T. I. Oprea, I. Zamora and A.-L. Ungell, Pharmacokinetically based mapping device for chemical space navigation, *J. Comb. Chem.*, 2002, **4**, 258–266.

- [11] T. Wright, V. J. Gillet, D. V. S. Green and S. D. Pickett, Optimizing the size and configuration of combinatorial libraries, *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 381–390.
- [12] D. K. Agrafiotis, Multiobjective optimization of combinatorial libraries, *J. Comput.-Aided Mol. Des.*, 2002, **16**, 335–356.
- [13] F. L. Stahura, L. Xue, J. W. Godden and J. Bajorath, Methods for compound selection focused on hits and application in drug discovery, *J. Mol. Graph. Model.*, 2002, **20**, 439–446.
- [14] T. I. Oprea, J. Gottfries, V. Sherbukhin, P. Svensson and T. C. Kuhler, Chemical information management in drug discovery: optimizing the computational and combinatorial chemistry interfaces, *J. Mol. Graph. Model.*, 2000, **18**, 512–524.
- [15] M. Vieth, M. G. Siegel, R. E. Higgs, I. A. Watson, D. H. Robertson, K. A. Savin, G. L. Durst and P. A. Hippskind, Characteristic physical properties and structural fragments of marketed oral drugs, *J. Med. Chem.*, 2004, **47**, 224–232.
- [16] M. C. Wenlock, R. P. Austin, P. Barton, A. M. Davis and P. D. Leeson, A comparison of physicochemical property profiles of development and marketed oral drugs, *J. Med. Chem.*, 2003, **46**, 1250–1256.
- [17] J. F. Blake, Examination of the computed molecular properties of compounds selected for clinical development, *BioTechniques*, 2003, **Suppl.**, 16–20.
- [18] J. Askenash, The rule of five revisited, *Preclinica*, 2004, **2**, 92.
- [19] A. Tronde, B. Norden, H. Marchner, A.-K. Wendel, H. Lennernaes and U. H. Bengtsson, Pulmonary absorption rate and bioavailability of drugs in vivo in rats: structure-absorption relationships and physicochemical profiling of inhaled drugs, *J. Pharm. Sci.*, 2003, **92**, 1216–1233.
- [20] A. O. Aptula, R. Kuehne, R. Ebert, M. T. D. Cronin, T. I. Netzeva and G. Schueuermann, Modeling discrimination between antibacterial and non-antibacterial activity based on 3D molecular descriptors, *QSAR Comb. Sci.*, 2003, **22**, 113–128.
- [21] T. Mitchell and G. A. Showell, Design strategies for building drug-like chemical libraries, *Curr. Opin. Drug Discov. Dev.*, 2001, **4**, 314–318.
- [22] A. Polinsky and G. B. Shaw, High-speed chemistry libraries: assessment of drug-likeness. In *Practice of Medicinal Chemistry* (ed. C. G. Wermuth), 2nd edition, Elsevier, London, UK, 2003, pp. 147–157.
- [23] H. Matter, K.-H. Baringhaus, T. Naumann, T. Klabunde and B. Pirard, Computational approaches towards the rational design of drug-like compound libraries, *Comb. Chem. High Throughput Screening*, 2001, **4**, 453–475.
- [24] C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Adv. Drug Deliv. Rev.*, 1997, **23**, 3–25.
- [25] W. P. Walters, Ajay and M. A. Murcko, Recognizing molecules with drug-like properties, *Curr. Opin. Chem. Biol.*, 1999, **3**, 384–387.
- [26] G. Muller and H. Giera, Protein secondary structure templates derived from bioactive natural products, *J. Comput.-Aided Mol. Des.*, 1998, **12**, 1–6.
- [27] X. Q. Lewell, D. Judd, S. Watson and M. Hann, RECAP – Retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry, *J. Chem. Inf. Comput. Sci.*, 1998, **38**, 511–522.
- [28] D. J. Newman, G. M. Cragg and K. M. Snader, Natural products as sources of new drugs over the period 1981–2002, *J. Nat. Prod.*, 2003, **66**, 1022–1037.
- [29] G. M. Rishon, Reactive compounds and in vitro false positives in HTS, *Drug Discov. Today*, 1997, **2**, 382–384.
- [30] M. Hann, B. Hudson, X. Lewell, R. Lively, L. Miller and N. Ramsden, Strategic pooling of compounds for high-throughput screening, *J. Chem. Inf. Comput. Sci.*, 1999, **39**, 897–902.

- [31] N. Baurin, R. Baker, C. Richardson, I. Chen, N. Foloppe, A. Potter, A. Jordan, S. Roughley, M. Parratt, P. Greaney, D. Morley and R. E. Hubbard, Drug-like annotation and duplicate analysis of a 23-supplier chemical database totaling 2.7 million compounds, *J. Chem. Inf. Comput. Sci.*, 2004, **44**, 643–651.
- [32] S. Sirois, D.-Q. Wei, Q. Du and K.-C. Chou, Virtual screening for SARS-CoV protease based on KZ7088 pharmacophore points, *J. Chem. Inf. Comput. Sci.*, 2004, **44**, 1111–1122.
- [33] <http://nihroadmap.nih.gov/molecularlibraries/index.asp>.
- [34] I. Muegge, S. L. Heald and D. Brittelli, Simple selection criteria for drug-like chemical matter, *J. Med. Chem.*, 2001, **44**, 1841–1846.
- [35] A. A. Patchett and R. P. Nargund, Chapter 26. Privileged structures – an update, *Annu. Rep. Med. Chem.*, 2000, **35**, 289–298.
- [36] T. Guo and D. W. Hobbs, Privileged structure-based combinatorial libraries targeting G protein-coupled receptors, *Assay Drug Dev. Technol.*, 2003, **1**, 579–592.
- [37] D. A. Horton, G. T. Bourne and M. L. Smythe, The combinatorial synthesis of bicyclic privileged structures or privileged substructures, *Chem. Rev.*, 2003, **103**, 893–930.
- [38] D. A. Horton, G. T. Bourne and M. L. Smythe, Exploring privileged structures: the combinatorial synthesis of cyclic peptides, *Mol. Divers.*, 2002, **5**, 289–304.
- [39] K. Bondensgaard, M. Ankersen, H. Thogersen, B. S. Hansen, B. S. Wulff and R. P. Bywater, Recognition of privileged structures by G-protein coupled receptors, *J. Med. Chem.*, 2004, **47**, 888–899.
- [40] B. R. Roberts, Target information in lead discovery, *Targets*, 2003, **2**, 14–18.
- [41] A. Schuffenhauer, P. Floersheim, P. Acklin and E. Jacoby, Similarity metrics for ligands reflecting the similarity of the target proteins, *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 391–405.
- [42] P. J. Hajduk, M. Bures, J. Praestgaard and S. W. Fesik, Privileged molecules for protein binding identified from NMR-based screening, *J. Med. Chem.*, 2000, **43**, 3443–3447.
- [43] G. Muller, Medicinal chemistry of target family-directed masterkeys, *Drug Discov. Today*, 2003, **8**, 681–691.
- [44] P. A. Rejto and G. M. Verkhrivker, Unraveling principles of lead discovery: from unfrustrated energy landscapes to novel molecular anchors, *Proc. Natl Acad. Sci. USA*, 1996, **93**, 8945–8950.
- [45] P. A. Rejto and G. M. Verkhrivker, Molecular anchors with large stability gaps ensure linear binding free energy relationships for hydrophobic substituents. In *Pacific Symposium on Biocomputing'98* (ed. R. B. Altman), Maui, Hawaii, January 4–9, 1998, pp. 362–373.
- [46] S. J. Teague, Implications of protein flexibility for drug discovery, *Nat. Rev. Drug Discov.*, 2003, **2**, 527–541.
- [47] R. Goodnow, Small molecule lead generation processes for drug discovery, *Drug Future*, 2002, **27**, 1165–1180.
- [48] M. D. Burke and S. L. Schreiber, A planning strategy for diversity-oriented synthesis, *Angew. Chem. Int. Ed.*, 2004, **43**, 46–58.
- [49] R. Breinbauer, I. R. Vetter and H. Waldmann, From protein domains to drug candidates: natural products as guiding principles in the design and synthesis of compound libraries, *Angew. Chem. Int. Ed.*, 2002, **41**, 2878–2890.
- [50] C. Hulme and V. Gore, Multi-component reactions: emerging chemistry in drug discovery from xylocain to crivivan, *Curr. Med. Chem.*, 2003, **10**, 51–80.
- [51] T. I. Oprea, A. M. Davis, S. J. Teague and P. D. Leeson, Is there a difference between leads and drugs? A historical perspective, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 1308–1315.
- [52] T. I. Oprea, Current trends in lead discovery: are we looking for the appropriate properties?, *Mol. Divers.*, 2002, **5**, 199–208.

- [53] J. R. Proudfoot, Drugs, leads, and drug-likeness: an analysis of some recently launched drugs, *Bioorg. Med. Chem. Lett.*, 2002, **12**, 1647–1650.
- [54] C. A. Lipinski, Compound properties and drug quality. In *Practice of Medicinal Chemistry* (ed. C. G. Wermuth), 2nd edition, 2003, pp. 341–349.
- [55] C. W. Murray and M. L. Verdonk, The consequences of translational and rotational entropy lost by small molecules on binding to proteins, *J. Comput.-Aided Mol. Des.*, 2002, **16**, 741–753.
- [56] H. C. Kolb and K. B. Sharpless, The growing impact of click chemistry on drug discovery, *Drug Discov. Today*, 2003, **8**, 1128–1137.
- [57] M. Congreve, R. Carr, C. Murray and H. Jhoti, A 'rule of three' for fragment-based lead discovery?, *Drug Discov. Today*, 2003, **8**, 876–877.
- [58] E. Jacoby, J. Davies and M. J. J. Blommers, Design of small molecule libraries for NMR screening and other applications in drug discovery, *Curr. Top. Med. Chem.*, 2003, **3**, 11–23.
- [59] C. A. Lepre, J. Peng, J. Fejzo, N. Abdul-Manan, J. Pocas, M. Jacobs, X. Xie and J. M. Moore, Applications of SHAPES screening in drug discovery, *Comb. Chem. High Throughput Screen*, 2002, **5**, 583–590.
- [60] C. A. Lepre, *Strategies for NMR Screening and Library Design. Methods and Principles in Medicinal Chemistry*, BioNMR in Drug Research, 2003, Vol. 16, pp. 391–415.
- [61] R. Carr and H. Jhoti, Structure-based screening of low-affinity compounds, *Drug Discov. Today*, 2002, **7**, 522–527.
- [62] A. M. Davis, S. J. Teague and G. J. Kleywegt, Application and limitations of X-ray crystallographic data in structure-based ligand and drug design, *Angew. Chem. Int. Ed.*, 2003, **42**, 2718–2736.
- [63] D. Fattori, Molecular recognition: the fragment approach in lead generation, *Drug Discov. Today*, 2004, **9**, 229–238.
- [64] P. S. Charifson and W. P. Walters, Filtering databases and chemical libraries, *Mol. Divers.*, 2002, **5**, 185–197.
- [65] V. J. Gillet and P. Willett, *Computational Methods for the Analysis of Molecular Diversity. Pharmacochemistry Library*, Trends in Drug Research III, 2002, Vol. 32, pp. 125–133.
- [66] H. Kubinyi, Opinion: drug research: myths, hype and reality, *Nat. Rev. Drug Discov.*, 2003, **2**, 665–668.
- [67] D. F. Veber, S. R. Johnson, H.-Y. Cheng, B. R. Smith, K. W. Ward and K. D. Kopple, Molecular properties that influence the oral bioavailability of drug candidates, *J. Med. Chem.*, 2002, **45**, 2615–2623.
- [68] P. R. Andrews, D. J. Craik and J. L. Martin, Functional group contributions to drug-receptor interactions, *J. Med. Chem.*, 1984, **27**, 1648–1657.
- [69] T. I. Oprea, Virtual screening in lead discovery: a viewpoint, *Molecules*, 2002, **7**, 51–62.
- [70] P. Sauerberg, P. S. Bury, J. P. Mogensen, H.-J. Deussen, I. Pettersson, J. Fleckner, J. Nehlin, K. S. Frederiksen, T. Albrektsen, N. Din, L. A. Svensson, L. Ynddal, E. M. Wulff and L. Jeppesen, Large dimeric ligands with favorable pharmacokinetic properties and peroxisome proliferator-activated receptor agonist activity in vitro and in vivo, *J. Med. Chem.*, 2003, **46**, 4883–4894.
- [71] K. R. Horspool and C. A. Lipinski, Advancing new drug delivery concepts to gain the lead, *Drug Deliv. Technol.*, 2003, **3**, 34–46.
- [72] B. P. Zambrowicz and A. T. Sands, Knockouts model the 100 best-selling drugs – will they model the next 100?, *Nat. Rev. Drug Discov.*, 2003, **2**, 38–51.
- [73] A. L. Hopkins and C. R. Groom, Opinion: the druggable genome, *Nat. Rev. Drug Discov.*, 2002, **1**, 727–730.
- [74] M. Adenot, R. S. Lahana and F. Nimes, Blood–brain barrier permeation models: discriminating between potential CNS and non-CNS drugs including P-glycoprotein substrates, *J. Chem. Inf. Comput. Sci.*, 2004, **44**, 239–248.

- [75] K. M. M. Doan, J. E. Humphreys, L. O. Webster, S. A. Wring, L. J. Shampine, C. J. Serabjit-Singh, K. K. Adkison and J. W. Polli, Passive permeability and P-glycoprotein-mediated efflux differentiate central nervous system (CNS) and non-CNS marketed drugs, *J. Pharmacol. Exp. Ther.*, 2002, **303**, 1029–1037.
- [76] F. Faassen, G. Vogel, H. Spanings and H. Vromans, Caco-2 permeability, P-glycoprotein transport ratios and brain penetration of heterocyclic drugs, *Int. J. Pharm.*, 2003, **263**, 113–122.
- [77] J. Kelder, P. D. J. Grootenhuys, D. M. Bayada, L. P. C. Delbressine and J.-P. Ploemen, Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs, *Pharm. Res.*, 1999, **16**, 1514–1519.
- [78] U. Norinder and M. Haerberlein, Computational approaches to the prediction of the blood–brain distribution, *Adv. Drug Deliv. Rev.*, 2002, **54**, 291–313.
- [79] M. Lobell, L. Molnar and G. M. Keseru, Recent advances in the prediction of blood–brain partitioning from molecular structure, *J. Pharm. Sci.*, 2003, **92**, 360–370.
- [80] P. Suomalainen, C. Johans, T. Soederlund and P. K. J. Kinnunen, Surface activity profiling of drugs applied to the prediction of blood–brain barrier permeability, *J. Med. Chem.*, 2004, **47**, 1783–1788.
- [81] F. Lombardo, E. Gifford and M. Y. Shalaeva, In silico ADME prediction: data, models, facts and myths, *Mini-Rev. Med. Chem.*, 2003, **3**, 861–875.
- [82] T. R. Stouch, J. R. Kenyon, S. R. Johnson, X.-Q. Chen, A. Doweyko and Y. Li, In silico ADME/Tox: why models fail, *J. Comput.-Aided Mol. Des.*, 2003, **17**, 83–92.
- [83] W. M. Pardridge, Blood–brain barrier genomics and the use of endogenous transporters to cause drug penetration into the brain, *Curr. Opin. Drug Discov. Dev.*, 2003, **6**, 683–691.
- [84] P. Stenberg, U. Norinder, K. Luthman and P. Artursson, Experimental and computational screening models for the prediction of intestinal drug absorption, *J. Med. Chem.*, 2001, **44**, 1927–1937.
- [85] P. Ertl, B. Rohde and P. Selzer, Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties, *J. Med. Chem.*, 2000, **43**, 3714–3717.
- [86] P. Artursson and C. A. S. Bergstroem, *Intestinal Absorption: the Role of Polar Surface Area. Methods and Principles in Medicinal Chemistry*, Drug Bioavailability, 2003, Vol. 18, pp. 341–357.
- [87] C. A. S. Bergstroem, M. Strafford, L. Lazorova, Al. Avdeef, K. Luthman and P. Artursson, Absorption classification of oral drugs based on molecular surface properties, *J. Med. Chem.*, 2003, **46**, 558–570.
- [88] I. Zamora, T. Oprea, G. Cruciani, M. Pastor and A.-L. Ungell, Surface descriptors for protein–ligand affinity prediction, *J. Med. Chem.*, 2003, **46**, 25–33.
- [89] T. Fichert, M. Yazdanian and J. R. Proudfoot, A structure-permeability study of small drug-like molecules, *Bioorg. Med. Chem. Lett.*, 2003, **13**, 719–722.
- [90] T. Sakaeda, N. Okamura, S. Nagata, T. Yagami, M. Horinouchi, K. Okumura, F. Yamashita and M. Hashida, Molecular and pharmacokinetic properties of 222 commercially available oral drugs in humans, *Biol. Pharm. Bull.*, 2001, **24**, 935–940.
- [91] Y. H. Zhao, M. H. Abraham, A. Hersey and C. N. Luscombe, Quantitative relationship between rat intestinal absorption and Abraham descriptors, *Eur. J. Med. Chem.*, 2003, **38**, 939–947.
- [92] C. A. Lipinski, Integration of physicochemical property considerations into the design of combinatorial libraries, *Pharm. News*, 2002, **9**, 195–202.
- [93] Y. Ran and S. H. Yalkowsky, Prediction of drug solubility by the general solubility equation (GSE), *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 354–357.
- [94] C. A. S. Bergstroem, U. Norinder, K. Luthman and P. Artursson, Molecular descriptors influencing melting point and their role in classification of solid drugs, *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 1177–1185.

- [95] F. Lombardo, R. S. Obach, M. Y. Shalaeva and F. Gao, Prediction of human volume of distribution values for neutral and basic drugs. 2. Extended data set and leave-class-out statistics, *J. Med. Chem.*, 2004, **47**, 1242–1250.
- [96] F. Lombardo, R. S. Obach, M. Y. Shalaeva and F. Gao, Prediction of volume of distribution values in humans for neutral and basic drugs using physicochemical measurements and plasma protein binding data, *J. Med. Chem.*, 2002, **45**, 2867–2876.
- [97] D. F. V. Lewis and M. Dickins, Substrate SARs in human P450s, *Drug Discov. Today*, 2002, **7**, 918–925.
- [98] K. M. Comess, M. J. Voorbach, M. L. Coen, H. Tang, L. Gao, X. Cheng, M. E. Schurdak, B. A. Beutel and D. J. Burns, Affinity-based high-throughput screening of orphan targets: practical solutions for removing promiscuous binders, Abstracts of Papers, 224th ACS National Meeting, Boston, MA, United States, August 18–22, 2002, COMP-220.
- [99] S. L. McGovern, B. T. Helfand, B. Feng and B. K. Shoichet, A specific mechanism of nonspecific inhibition, *J. Med. Chem.*, 2003, **46**, 4265–4272.
- [100] S. L. McGovern and B. K. Shoichet, Kinase inhibitors: not just for kinases anymore, *J. Med. Chem.*, 2003, **46**, 1478–1483.
- [101] J. Seidler, S. L. McGovern, T. N. Doman and B. K. Shoichet, Identification and prediction of promiscuous aggregating inhibitors among known drugs, *J. Med. Chem.*, 2003, **46**, 4477–4486.
- [102] C. M. Tice, Selecting the right compounds for screening: does Lipinski's rule of 5 for pharmaceuticals apply to agrochemicals?, *Pest Manage. Sci.*, 2001, **57**, 3–16.
- [103] C. M. Tice, Selecting the right compounds for screening: use of surface-area parameters, *Pest Manage. Sci.*, 2002, **58**, 219–233.
- [104] V. V. Zernov, K. V. Balakin, A. A. Ivaschenko, N. P. Savchuk and I. V. Pletnev, Drug discovery using support vector machines. The case studies of drug-likeness, agrochemical-likeness, and enzyme inhibition predictions, *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 2048–2056.