ELSEVIER

Improving the hit-to-lead process: data-driven assessment of drug-like and lead-like screening hits

Tobias Wunberg, Martin Hendrix, Alexander Hillisch, Mario Lobell, Heinrich Meier, Carsten Schmeck, Hanno Wild and Berthold Hinzen

Bayer HealthCare, Pharma Research & Development, Medicinal Chemistry, D-42096 Wuppertal, Germany

Drug-like and lead-like hits derived from HTS campaigns provide good starting points for lead optimization. However, too strong emphasis on potency as hit-selection parameter might hamper the success of such projects. A detailed absorption, distribution, metabolism, excretion and toxicology (ADME–Tox) profiling is needed to help identify hits with a minimum number of (known) liabilities. This is particularly true for drug-like hits. Herein, we describe how to break down large numbers of screening hits and we provide a comprehensive overview of the strengths and weaknesses for each structural class. The overall profile (e.g. ligand efficiency, selectivity and ADME–Tox) is the distinctive feature that will define the priority for follow-up.

Drugs have often been discovered starting from endogenous ligands or natural products, followed by chemical modifications to improve *in vivo* efficacy. With the breakthroughs in biochemistry in the early 1980s, this successful approach was further supplemented by *in vitro* assays with increased throughput. In the 1990s, this also paved the way for a new research paradigm that relied to a significant degree on HTS of large compound libraries to identify novel lead structures.

With the advent of genomics, an increasing number of novel and clinically less-validated targets identified from the human genome project have been screened in recent years. Despite these fundamentally positive developments of new therapeutic target opportunities, the number of new drug launches is decreasing and the overall costs for the development of a new drug are rising steadily [1]. Consequently, the pharmaceutical industry has recently come under significant pressure to improve its R&D process.

Together with the degree of target validation, the processing of HTS hit lists (i.e. selection, prioritization and chemical follow-up of hits) is a crucial phase in drug discovery to reduce attrition and therefore improve the overall R&D productivity. Here, we will discuss our experience with this process and a selection cascade for the identification of the most suitable screening hits will be outlined and will be illustrated by case studies. We will focus only on HTS-derived hit lists for drugs that are intended for oral administration. Other considerations need to be applied for hits derived from other techniques, such as fragment-based screening [2], or projects with the requirement of a parenteral formulation.

Corresponding author: Hinzen, B. (Berthold.Hinzen@BayerHealthcare.com).

Hits, leads and drugs

Hits are defined as nonreactive compounds [3] of verified structure and purity, with a confirmed minimal *in vitro* potency (<20 μM) derived from HTS. By contrast, leads are compounds that, in addition to their potency, show some degree of selectivity for the given target, along with confirmed specific binding, emerging SAR, biophysical and absorption, distribution, metabolism, excretion and toxicity (ADME–Tox) properties, which are promising for their eventual optimization into small-molecule drugs. Furthermore, when considering a set of different leads, their chemical accessibility (e.g. number of synthetic steps) and ease in the synthesis of their analogs might play an important role for further selection, especially for those leads that are not fully drug-like in their properties and thus need to undergo extensive modifications.

In recent years, the definition of a lead and especially of lead-likeness in physicochemical terms has drawn considerable attention. This debate was initiated by the pioneering work of Lipinski [4] and Teague and co-workers [5]. Triggered by the fact that hits

1359-6446/06/$ – see front matter PII: S1359-6446(05)03700-1

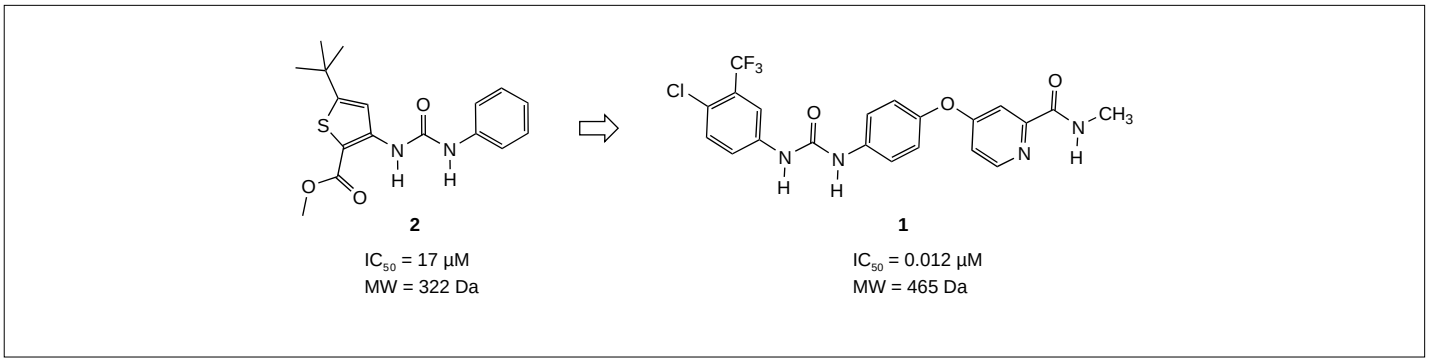


FIGURE 1

Discovery of sorafenib (1) from a lead-like low-potency screening hit.

BOX 1

Rule of five and early definition of lead-likeness.

Rule of five (RO5):
(i) MW ≤ 500 Da;
(ii) Lipophilicity (log *P*) ≤ 5;
(iii) Sum of H-bond donors ≤ 5;
(iv) Sum of N- and O-acceptors ≤ 10.

Lead-likeness:
(i) MW < 350 Da;
(ii) Log *P* < 3;
(iii) Affinity ~ 0.1 μM.

identified via HTS and combinatorial chemistry could not fulfill the initial expectations of a more-rapid identification of promising investigational new drugs (INDs) and based on the observation that compounds violating defined limits in molecular weight (MW), number of H-bond donors and acceptors and lipophilicity often have poor oral absorption, Lipinski published the 'rule of five' (RO5), which defines the desired properties for oral absorption (Box 1).

According to Teague *et al.* [5], lead optimization generally results in increased molecular weight and lipophilicity. Consequently, drug discovery should start with lead-like structures not only fulfilling the RO5 but also with even lower molecular weights and lipophilicity. This observation resulted in the concept of lead-likeness (Box 1). These criteria have been used for the design of discovery libraries and for hit selection from HTS campaigns.

Since the introduction of these concepts (Box 1), more detailed analysis followed, which compared drugs and their corresponding leads. These investigations resulted in a more-profound understanding of the lead optimization process on a molecular basis [e.g. MW, rotable bonds and polar surface area (PSA)]. In parallel, additional aspects, such as toxicity, synthetic accessibility or pharmacokinetic properties, have been considered, which have not only led to a refined definition of drug-likeness [6] but also impacted the criteria of lead-likeness [7,8].

Potency alone is a false predictor for hit selection

It is now widely accepted that proper selection of screening hits is of central importance for the success of lead optimization efforts [9] and that these hits will ideally exhibit lead-like properties (i.e. possess a well-balanced mixture of potency, selectivity and other important features). Such features would include favorable physicochemical properties, absence of toxophores and amenability to the rapid synthesis of chemical analogues [10]. More recently, the concept of ligand efficiency has been introduced for the analysis of screening hit lists [11]; this parameter uses the experimental binding energy per atom (or Andrews' binding energy) to normalize molecular weight and potency. In this scenario, it would be preferable to follow up on hits with the highest ligand efficiency, rather than those with the greatest potencies *per se*. Compounds carefully selected using this paradigm will represent the most promising starting points for successful lead optimization [12].

This new paradigm is clearly not in line with the traditional lead identification process, which had potency as the most important driving force. Typically, early *in vivo* proof-of-principle studies have been performed with the most potent screening hits. First pharmacological effects *in vivo* have often been obtained after intravenous or intraperitoneal administration (even if oral administration is part of the targeted drug profile) using formulations not suitable for later studies or clinical development ('protype drugs' and 'chemical validation tools'). Rather than using them solely for target validation, these promising early biological data gave such compounds sufficient support for further chemical resourcing, despite the fact that the overall profile (e.g. ADME–Tox) had hardly been investigated and might have been far from optimal. Optimization of ADME–Tox properties is considered to be much more difficult than optimizing potency and/or selectivity [13]. Therefore, this procedure led to apparent high success rate in lead identification and high failure rate in the lead optimization phase [14].

In summary, it is of central importance that the hit selection process is based not only on potency, but also on all other parameters that will subsequently strongly influence the successful outcome of the lead optimization phase. Ideally, this process will be data-driven as much as possible.

The importance of drug-like hits

Collection of ADME–Tox data will be tedious for the screening of hit lists comprising a few thousand compounds. Therefore, it might be tempting to reduce this number by simply applying filters based on the definition of lead-likeness. A successful example of the metamorphosis of a lead-like hit into a candidate for clinical development is the kinase inhibitor sorafenib (**1**), currently in Phase III clinical development (Figure 1) [15]. This compound was derived

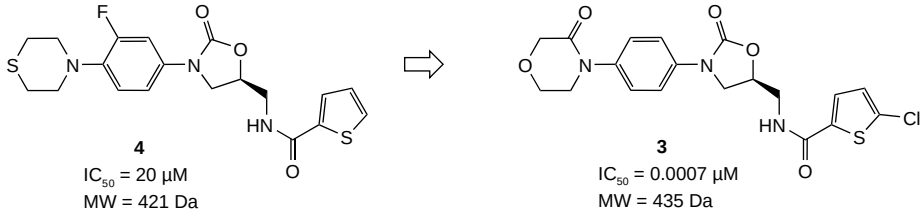


FIGURE 2

Discovery of BAY 59-7939 (3) from a lead-like low-potency screening hit.

TABLE 1

Examples of drug-like HTS hits that allowed the identification of candidates for clinical development[a]

No.	Hit potency (μM)	Hit MW (Da)	Hit log *P*	Candidate MW (Da)	Candidate potency (μM)	MW difference	Change in Potency (fold)	Class target	Original indication	New indication
1	0.1	512	2.4	573	0.001	+61	100	DNA processing		Anti-infectives
2	1.0	500	3.1	400	0.2	−100	50	Protease	Cardio-vascular	COPD
3	0.05	465	5.0	519	0.004	+54	12	Nuclear receptor	Dyslipedemia	Dyslipedemia
4	0.8	458	4.5	479	0.08	+21	10	Enzyme	Cancer	Obesity
5	1.0	414	1.6	395	0.002	−19	500	Kinase	Cardiovascular	Asthma

[a]All of the values for potency were derived from a biochemical assay (IC_{50}) with the exception of example number 1, which includes activities derived from a cellular assay (EC_{50}).
Abbreviation: CODP, chronic obstructive pulmonary disease

from the commercially available screening hit **2**, which clearly exhibits lead-like properties (Figure 1). Here, medicinal chemistry efforts resulted in significant changes of the core structure and the molecular weight (difference in MW = 143 Da).

Despite this story of success, medicinal chemists often encounter a different situation when dealing with screening hit lists. At Bayer, and most likely at other pharmaceutical companies as well, screening pools are composed, to a large extend, of compounds originating from earlier lead optimization programs and from combinatorial chemistry efforts (large, random libraries including compounds with high MW and lipophilicity but also focused libraries synthesized in lead optimization programs). Therefore, hits are routinely identified that possess drug-like rather than lead-like properties in terms of MW and lipophilicity. Sometimes these more drug-like than lead-like compounds will also have higher potency, but even if this is not the case, these hits might show their own advantages as starting points for medicinal chemistry programs.

A striking example of the importance of using screening hits with more drug-like properties as starting points for medicinal chemistry campaigns is the discovery of BAY 59-7939 (**3**), a factor Xa inhibitor, currently in Phase II clinical trials (Figure 2). Screening hit **4** was identified as a low-potency inhibitor of factor Xa. This hit clearly showed drug-like properties. Despite its low potency, optimization by medicinal chemistry efforts ultimately led to **3** (Figure 2). Notably, a total of only four minor modifications, such as removal or attachment of a halide, caused an almost 30,000-fold increase in activity, although the MW was increased by only 14 Da [16].

Additional examples of drug-like hits further underline the usefulness of these compounds as leads for drug discovery (Table 1). Candidates for clinical development were derived from drug-like hits for a variety of targets in different indications. In some cases, medicinal chemistry efforts were even able to achieve a reduction of molecular weight and still increase potency by at least one order of magnitude.

Among the five hits shown in Table 1, four compounds originated from previous lead optimization programs. Only screening hit number **1** was synthesized by combinatorial chemistry as a chemistry-driven library not addressing a specific target class. Selectivity versus the original target was addressed quite easily for two reasons: (i) difference in SAR for the two targets; (ii) compounds lacked or were only weakly active versus the original target. The latter because they might represent synthetic intermediates or they are compounds from, for example, a combinatorial library designed to expand the SAR.

These examples clearly demonstrate that drug-like hits and lead-like hits are complementary, even if they show only moderate-to-low affinity for the target. Therefore, the following conclusions can be drawn:

(i) It does not matter *per se* if a hit or lead is lead-like or drug-like. We need to know the overall profile to judge its suitability for further chemical optimization.

(ii) Discarding drug-like hits or leads (even low-potency ones) would be a waste of precious chemical matter and intellectual property.

(iii) A data-driven process addressing ADME–Tox issues is needed to identify among the drug-like and the lead-like hits those compounds with a minimum number of liabilities. This process needs to be able to deal with hit lists of >1000 compounds.

(iv) Not all liabilities should be equally weighted, some are easier to address by chemical means than others (e.g. potency and selectivity versus ADME–Tox).

TABLE 2

Cascade for hit characterization and prioritization

	Assay	Throughput ADME–Tox data per HTS assay	Comment
Phase 1	*In silico* profiling	unlimited	Solubility, molecular size, lipophilicity, PSA, rotable bonds, structural alerts
Phase 2	CYP3A4 inhibition	~100 compounds	Single point (% inhibition at 10 μM) using human microsomes and midazolam
	Physicochemical profile	~100 compounds	Solubility in buffer (pH = 6.5), membrane affinity, protein binding
	Microsomal stability	~100 compounds	Stability after incubation in human liver microsomes
	hERG activity	~100 compounds	*In vitro* assay as a filter for functional assay
Phase 3	Exposure after oral administration	~10 compounds	Cassette dosing might be applied

TABLE 3

***In silico* ADME–Tox traffic light colors and corresponding property values[a]**

Calculated solubility TL (mg/l)[b]	Molecular volume TL (MW_{corr}[c])	Lipophilicity TL (Log *P*)[d]	PSA (Å^2) TL[e]	Rotable bonds TL[f]	Metabolic clearance TL[g]	hERG inhibition TL	CYP inhibition TL	Undesirable groups TL[h]
≥50	≤400	≤3	≤120	≤7	No alert	No alert	No alert	No alert
10–50	400–500	3–5	120–140	8–10			Weak alert	Weak alert
<10	>500	>5	>140	≥11	Alert	Alert	Strong alert	Strong alert

[a]We have also developed and validated a four-level *in silico* classification system. For the undesirable groups TL, a finer grading and inclusion of a black label to highlight particularly undesirable (e.g. highly reactive) features has been found useful.

[b]Solubility: buffer solubility at pH 6.5; the prediction algorithm has been developed by stepwise multiple linear regression analysis on a training set of 4806 Bayer in-house compounds (values determined from DMSO stock solutions by precipitation).

[c]MW_{corr}: molecular weight corrected for the occurrence of halogen atoms to achieve proportionality to molecular volume.

[d]Lipophilicity: expressed as log *P* values (log of the octanol:water partition coefficient), which are calculated with the CLOGP program (BioByte Corp., version 4.2, fragment database version 22) as implemented in SYBYL (TRIPOS).

[e]PSA: polar surface area; calculated from the structure in Ref. [20]; in our implementation only nitrogen and oxygen atoms and their attached hydrogen atoms are regarded as polar, whereas sulfur atoms are regarded as nonpolar.

[g]Metabolic clearance: stability in human microsomes; these three traffic lights and the undesirable groups TL are used as alerts, indicating an increased risk with regard to fast metabolic degradation, hERG channel blockade, CYP inhibition or undesirable chemical properties (e.g. chemical reactivity, mutagenicity and propensity to the formation of reactive metabolites). The underlying algorithms for these four TLs are based on a combination of substructure recognition and calculated physicochemical properties and have been developed from literature and in-house data and experiences.

[f]Rotable bonds: the number of rotable bonds can be easily calculated from the structure and is a reflection of the conformational flexibility of a molecule, which in turn influences the oral bioavailability of drug candidates [21].

[h]Undesirable groups: the underlying algorithm of this traffic light is based on substructure recognition and has been developed with in-house data and personal experiences by our medicinal chemists.

Implementation of a data-driven hit-to-lead process

The term hit-to-lead describes the process of filtering and assessing hits (derived from HTS) and the limited chemical modifications of these hits to establish an initial SAR. Data other than potency and selectivity of a given hit or cluster of hits are collected to allow for a data-driven decision on which class should be advanced into the lead optimization phase [17]. The goal is to identify compounds with a minimum number of known liabilities that will provide the optimal starting point for further activities. Because lead optimization is a much more costly and resource-intensive process than hit evaluation, the generation of data that will lead to an early decision to terminate activities on a given series of compounds is of particular importance. Moreover, the knowledge of the individual liabilities of selected compounds will provide a guideline for the lead optimization activities, addressing crucial issues early on.

Hit selection and prioritization can be based on two categories, namely biological profile (potency, selectivity and specificity) and ADME–Tox and physicochemical properties. The first category is beyond the scope of this review. In the following, we will describe in detail our hit selection process with regard to the second category (Table 2).

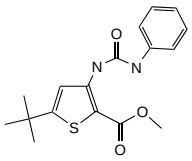
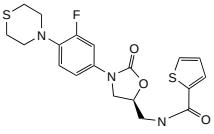
Because a hit set from an HTS campaign often comprises several hundred to several thousand compounds, experimental assessment of the ADME–Tox and physicochemical profile of the entire hit set represents a considerable challenge. To rationally break down these large numbers, the first characterization phase (phase 1) comprises an *in silico* assessment via an *in silico* ADME–Tox scoring system. Most of the scores can be conveniently labeled by a 'traffic light' (TL) color code (Table 3). The color coding indicates a risk assessment and facilitates data analysis and visualization. All generated *in silico* data are considered purely as alerts, no hits will be discarded based on *in silico* information only.

To exemplify the *in silico* ADME–Tox traffic lights, the screening hits described in Figure 1 and 2 have been assessed accordingly (Table 4). Potential strengths and weaknesses of the compounds are visualized in a straightforward manner.

In the second phase of the cascade (phase 2), experimental *in vitro* characterization of selected screening hits is initiated. Structural clustering is performed to aid compound selection (typically structural clustering is based on CATS or Unity Fingerprints). Following positive *in silico* assessment, singletons and cluster representatives are selected for purity assessment by LC–MS. For large

TABLE 4

***In silico* ADME–Tox traffic lights for kinase inhibitor and factor Xa inhibitor screening hits from Figure 1 and 2**

Structure	Calculated solubility TL (mg/l)	Molecular volume TL (MW_{corr})	Lipophilicity TL (Log *P*)	PSA (Å^2) TL	Rotable bonds TL	Metabolic clearance TL	hERG inhibition TL	CYP inhibition TL	Undesirable groups TL
	<10	332.4	5.1	67.4	5	No alert	No alert	No alert	Non-fused thiophene
	10–50	407.7	2.6	61.9	5	No alert	No alert	No alert	Non-fused thiophene

Abbreviations: MW_{corr}, molecular weight corrected for the occurrence of halogen atoms; log *P*, log of the octanol:water partition coefficient.

clusters, several representatives are selected, based on maximum structural diversity and *in silico* assessment. High to medium throughput assays are used to characterize the representative screening hits with respect to human ether-a-go-go related gene (hERG) activity, cytochrome P450 (CYP) 3A4 inhibition and stability in human liver microsomes. To obtain reasonable data from the experimental physicochemical profiling and from pharmacokinetic studies, a purity of >80% is mandatory. The available throughput of these assays is in the range of ~100 compounds per assay. Collection of experimental data helps to validate the *in silico* assessment and provides a reliable picture of the hits.

In the final phase (phase 3), the ADME–Tox and physicochemical profile is complemented for examples of the most promising clusters and singletons from phase 2 by an *in vivo* pharmacokinetic profile, as determined after oral administration. From these experiments, area-under-curve (AUC) and C_{max} are determined as key parameters reflecting oral exposure.

After these filters have been passed, the profile of all remaining hits is assessed with regard to the chances to solve the identified crucial issues ('path forward') in a subsequent lead optimization program. Depending on the availability of suitable analogs, this might require limited chemistry resources per hit. In general, the following criteria are applied before a lead series is moved forward into a lead optimization program: known SAR (multiples of actives and inactives), reasonable biochemical and/or functional activity (depending on target and indication), selectivity for the target, demonstrated *in vivo* efficacy, preliminary intellectual property (IP) assessment.

Discussion

Processing through the described cascade will generate a data package for each selected cluster and singleton. These data, in combination with each compound's biological profile, need to be assessed by experienced medicinal chemists supported by efficient tools for data analysis and visualization [18,19]. The prioritization of the hits for chemical follow-up is clearly not a deterministic process. Typically, from a successful screen 3–5 structural classes remain worthwhile for initiating the hit-to-lead optimization process. Other structural classes are, in most cases, discarded because of liabilities identified for the entire cluster (resulting in very limited chances to overcome this issue within the SAR) or because of multiple severe liabilities within the same class or compound (again rendering the corresponding multi-parameter optimization too difficult).

The most promising compounds in this context are those with a favorable ligand:efficiency ratio, as well as a minimum number of liabilities and (ideally) liabilities that are considered to be addressable by medicinal chemistry in a straightforward fashion. Generally, they cover the drug-like and the lead-like chemical space. Truly lead-like molecules remaining at this stage of the hit-to-lead process are attractive starting points for discovery programs because they usually offer large room for optimization. However, this advantage of lead-like hits is typically counterbalanced by a weak biological activity and/or selectivity at comparable ligand efficiencies. Furthermore, truly lead-like leads are more likely to fall under the claims of competitor patents or are highly generic and the chemical and structural direction of their further optimization is unclear, thus rendering the eventual IP situation less predictable. Finally, if no structural information on the interaction with the target is available, more time and resources will be required to explore the mostly undefined chemical space around the lead. Thus, lead-like hits might add additional obstacles to the lead optimization process, potentially making it a longer route to the final development candidate.

By contrast, drug-like leads have some advantages compared with lead-like leads, despite the fact that they might offer less room for chemical optimization. Because these compounds originate from earlier lead optimization projects, the chemical class has already been proven to possess some degree of biological specificity, thereby avoiding the issue of false positives caused by an unspecific interaction with the biological readout system. Furthermore, as a result of their more-complex structure, their IP situation can be better defined. Finally, the chemical space around the leads is often well evaluated from an earlier medicinal chemistry program. Therefore, clear SAR information with respect to selectivity, potency and pharmacokinetic properties can be obtained from already existing analogs and thus offering a clear path to overcome identified liabilities very early in the optimization process. This renders the entire endeavor less time consuming and more

predictive. Even if a lead cannot be optimized, the necessary information for termination will be available sooner than for a more lead-like hit.

In conclusion, drug-like and lead-like hits derived from HTS campaigns provide good starting points for lead optimization. Independent of the category under which a hit might fall, a detailed hit profiling should be performed based on experimental data, prior to the lead optimization phase. This allows the identification of hits with a minimum number of (known) liabilities. This type of hit will provide the best starting point for lead optimization. We have implemented a process to break down large numbers of screening hits that provides a comprehensive overview of the strengths and weaknesses of each structural class. The overall profile is the distinctive feature that will define the priority for follow-up. Along these lines, non-lead-like hits can be very promising leads for drug discovery and should not be discarded on the basis of a too stringent application of the various rules for lead-likeness.

Acknowledgements

We would like to thank all our colleagues from the hit-to-lead groups and our screening departments with whom we had many hours of fruitful discussions about specific aspects of the hit-to-lead process. Their input certainly has helped to shape the described cascade and to test it under 'real life' conditions. We would also like to thank all our colleagues who have contributed with their experience to set up the traffic light systems. Finally, we would like to thank J. R. Proudfoot for providing us with electronic copies of the chemical structures used in his publication.

References

1 Drews, J. (2003) Strategic trends in the drug industry. *Drug Discov. Today* 8, 411–420
2 Rees, D.C. *et al.* (2004) Fragment-based lead discovery. *Nat. Rev. Drug Discov.* 3, 660–672
3 Rishton, G.M. (2003) Nonleadlikeness and leadlikeness in biochemical screening. *Drug Discov. Today* 8, 86–96
4 Lipinski, C.A. *et al.* (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery settings. *Adv. Drug Del. Rev.* 23, 3–25
5 Teague, S.J. *et al.* (1999) The design of leadlike combinatorial libraries. *Angew. Chem. Int. Ed. Engl.* 38, 3743–3748
6 Lajiness, M.S. *et al.* (2004) Molecular properties that influence oral drug-like behavior. *Curr. Opin. Drug Discov. Devel.* 7, 470–477
7 Proudfoot, J.R. (2002) Drugs, leads, and drug-likeness: an analysis of some recently launched drugs. *Bioorg. Med. Chem. Lett.* 12, 1647–1650
8 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
9 Milne, G.M. (2003) Pharmaceutical productivity – the imperative for new paradigms. *Annu. Rep. Med. Chem.* 38, 383–396
10 Gribbon, P. and Sewing, A. (2005) High-throughput drug discovery: what can we expect from HTS? *Drug Discov. Today* 10, 17–22
11 Hopkins, A.L. *et al.* (2004) Ligand efficiency: a useful metric for lead selection. *Drug Discov. Today* 9, 430–431
12 Bleicher, K.H. *et al.* (2003) Hit and lead generation: beyond high throughput screening. *Nat. Rev. Drug Discov.* 2, 369–378
13 MacCoss, M. and Baillie, T.A. (2004) Organic chemistry in drug discovery. *Science* 303, 1810–1813
14 Alanine, A. *et al.* (2003) Lead generation – enhancing the success of drug discovery by investing the hit to lead process. *Comb. Chem. High Throughput Screen.* 6, 51–66
15 Lowinger, T. *et al.* (2002) Design and discovery of small molecules targeting Raf-1 kinase. *Curr. Pharm. Des.* 8, 2269–2278
16 Roehrig, S. *et al.* (2005) Discovery of the novel antithrombotic agent 5-Chloro-N-({(5S)-2-oxo-3-[4-(3-oxomorpholin-4-yl)phenyl]-1,3-oxazolindin-5-yl}methyl)thiophene-2-carboxamide (BAY 59-7939): an oral direct factor Xa inhibitor. *J. Med. Chem.* 48, 5900–5908
17 Caldwell, G.W. *et al.* (2001) The new pre-clinical paradigm: compound optimization in early and late phase drug discovery. *Curr. Top. Med. Chem.* 1, 353–366
18 Mittendorf, J. (2004) Integrated approaches to informatics: Bayer healthcare pharmacophore informatics platform. *Hit-to-Lead*, 25–26 November, London, UK
19 Shen J. *et al.* (2004) Decision support systems in drug discovery. *Mini-Rev. Med. Chem.* 4, 1019–1028
20 Ertl, P. *et al.* (2000) 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.* 43, 3714–3717
21 Veber, D.F. *et al.* (2002) Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *J. Med. Chem.* 45, 2615–2623