DOI: 10.1002/anie.201105840

Synthetic Chemistry

Lead-Oriented Synthesis: A New Opportunity for Synthetic Chemistry

*Alan Nadin, Channa Hattotuwagama, and Ian Churcher**

drug discovery · lipophilicity · medicinal chemistry · molecular diversity · synthetic methods

The pharmaceutical industry remains solely reliant on synthetic chemistry methodology to prepare compounds for small-molecule drug discovery programmes. The importance of the physicochemical properties of these molecules in determining their success in drug development is now well understood but we present here data suggesting that much synthetic methodology is unintentionally predisposed to producing molecules with poorer drug-like properties. This bias may have ramifications to the early hit- and lead-finding phases of the drug discovery process when larger numbers of compounds from array techniques are prepared. To address this issue we describe for the first time the concept of lead-oriented synthesis and the opportunity for its adoption to increase the range and quality of molecules used to develop new medicines.

1. Introduction

The expansion of synthetic methodology in recent years has greatly facilitated the preparation of molecules that would once have been considered an insurmountable synthetic challenge. In turn, the drug discovery industry, where large numbers of molecules are prepared and tested as potential new medicines, is one of the principal end-users and beneficiaries of this enlarged toolkit. Developing a drug suitable to treat patients effectively remains an enormous and well-documented[1,2] challenge. Our ability to address the issues which determine drug success or failure (e.g. safety, efficacy, pharmacokinetics and metabolism, speed to market etc.) is critically reliant on synthetic chemistry methodology to make the right molecules quickly and predictably. The impact of organic synthesis and the desired properties[3,4] of molecules at key phases of the drug discovery process is shown schematically in Figure 1. Initial starting points (termed "hits" or "leads") are often found by screening of compounds from chemical libraries which are usually reliant on robust chemical methodology for their production. Once initial hits have been identified, the "lead optimization" process aims to improve their drug-like profile through the synthesis of many designed analogs by using a range of chemistries. With the need to deliver medicines to patients quickly, time available for expansion of substrate scope or reaction optimization is usually limited and inevitably robust, predictable chemistries find the most utility. Typically many hundreds or even thousands of compounds are synthesized in the search for a clinical candidate drug. The ultimate choice of molecule is critical as it cannot be changed in development without large amounts of additional work. Once a single clinical candidate molecule has been identified, scale-up and process chemistry then devises often extremely elegant and efficient syntheses to deliver the final compound on multi-kg scales.

Many studies[3,4] now show the significant effect of the physicochemical properties of drug molecules on their likelihood of success in development. *For maximum benefit, synthetic methodologies should be well suited to the delivery of molecules with these properties—but are current methodologies well positioned to do this?*

2. Properties of Successful Hits and Drugs

The link between physicochemical descriptors and drug-like properties was identified some time ago[5,6] with, for example, the Lipinski description that orally available drug molecules usually have $\mathrm{mw} < 500$ and $\mathrm{Log}P < 5$ (together with limits on hydrogen bonding groups) often being used as a mnemonic to describe small-molecule drug-like space (it is often overlooked that the Lipinski rules provide a guide to

---

[*] Dr. A. Nadin, Dr. C. Hattotuwagama, Dr. I. Churcher
GlaxoSmithKline Medicines Research Centre
Gunnels Wood Road, Stevenage, SG1 2NY (UK)
E-mail: ian.2.churcher@gsk.com

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201105840.

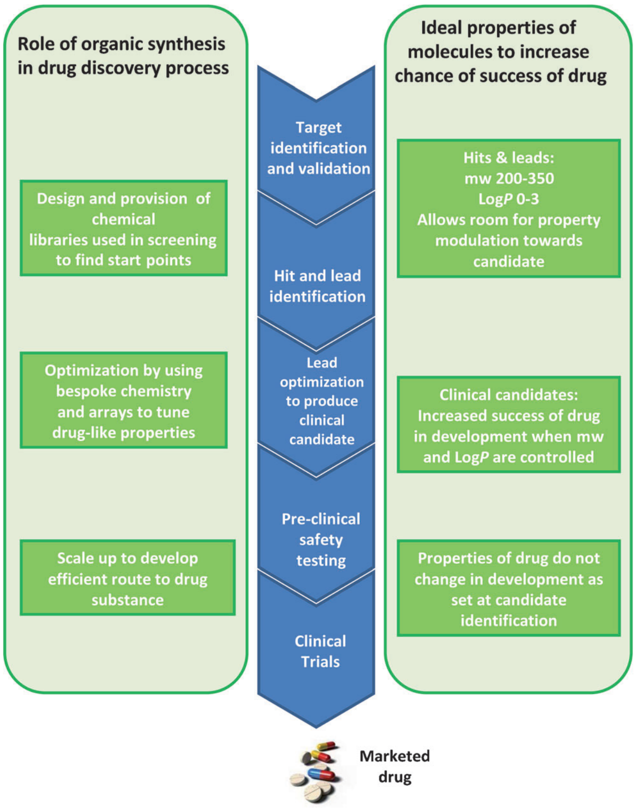

Figure 1. The role of organic synthesis in the drug discovery process. mw = molecular weight in Da; Log*P* = Log(partition coefficient in water/*n*-octanol).

minimizing absorption issues *only*). More recent studies[7–9] have shown, however, that compounds close to the Lipinski limits actually have a lower probability of success in development with significantly lower mw and LogP values favored in successful medicines. These observations have led to several revised guidelines suggesting, for example, that molecules with $\mathrm{Log}P < 3$ and polar surface area[10] > 75 Å^2 will show greater safety in pre-clinical studies[9] or that molecules with $\mathrm{mw} < 400$ and $\mathrm{Log}P < 4$ appear to be more successful in a range of assays of drug-like character.[7] The risks associated with excessively lipophilic molecules and a clear correlation between lipophilicity and number of off-target biological activities has also been noted.[3]

Whilst the chance of success in development of a drug may depend on many properties of the molecule, the detrimental effect of excessive hydrophobicity/lipophilicity (i.e. high Log*P*) is probably the most critical factor.[3,11–13] Unlike molecular weight however, it is one that is not easily estimated by inspection or calculated without specific tools (there are several free, publicly available tools to calculate Log*P*).[14] Log*P* is formally the log of the ratio of concentrations of a solute between immiscible phases, often water and octanol. It can more easily be considered as the relative ease with which a molecule will dissolve in an organic solvent relative to water with higher values representing less polar, more lipophilic molecules. This hydrophobic behavior can result in poor aqueous solubility[15–17] but also favors the binding of drug molecules to either their protein targets (i.e. a relatively more hydrophobic environment relative to bulk solution) and/or the apolar environment of cellular membranes, often the location of many biological targets which may mediate toxic effects. The strategy of increasing lipophilicity to gain better binding to a drug target is thus often used to increase apparent drug potency but almost certainly also results in a greater probability of binding to other, undesired drug targets. This promiscuous binding can lead to undesired biological effects, which may translate to toxicity and side effects contributing to the high failure rate of these types of compounds. Whilst making molecules too lipophilic is a more common issue leading to increased risk, making molecules highly polar (e.g. negative Log*P*) may also introduce issues as these polar molecules may have difficulty crossing cellular membranes unless specific transporting mechanisms are operating.[18–20] Some targets in, for example, anti-bacterial research, however, are thought to require more polar molecules to efficiently reach their sites of action.

Controlling the lipophilicity of drug substances is thus of central importance to the drug discovery process and LogP values in the range 1–3 are now thought to give the best balance of properties for most oral drugs.

To produce drug molecules in the historically lower-risk areas of property space (i.e. $\mathrm{Log}P \approx 1$–3 with molecular weight < 400) the properties of molecular starting points (hits or leads) needed to deliver these drug candidates can be identified. Regrettably, it is extremely rare that a molecule found as an initial hit from a biological screen will be suitable to be the final drug candidate; usually a large amount of lead optimization will be required. This process aims to improve many properties of the potential drug including 1) potency at the drug target; 2) selectivity over unwanted biological activities and toxicities; and 3) drug metabolism and pharmacokinetic properties. Historic precedent has shown that this lead optimization process is usually accompanied with an increase in molecular weight and lipophilicity of the lead[21–23] as medicinal chemists add complexity and size to the molecule to search for additional molecular interactions to achieve the goals set out above. In order to allow for this flexibility and property inflation during the lead optimization process and still deliver a final drug molecule with preferred properties, an area of physicochemical space known as "lead-like" space can be defined[21,24] and is shown schematically in Figure 2. In general, lead-like molecules will usually have well-controlled properties i.e. Log*P* values in the range -1 to $+3$ and molecular weight in the range 200 to 350 g mol^{-1} (i.e. 14 to 26 non-hydrogen atoms). Though drug molecules dosed through non-oral routes may in some cases be tolerant of wider ranges of physical properties, lead-like molecules still represent the best starting points to allow maximum flexibility through optimization. It is noteworthy that many early combinatorial chemistry libraries produced larger molecules which require a lead optimization vector opposite to that usually seen in order to reach optimal drug-like space.

There are of course many successful drugs with properties outside of those described above but in general these molecules rely on specific transport mechanisms for their pharmacology or else have unique biological action able to

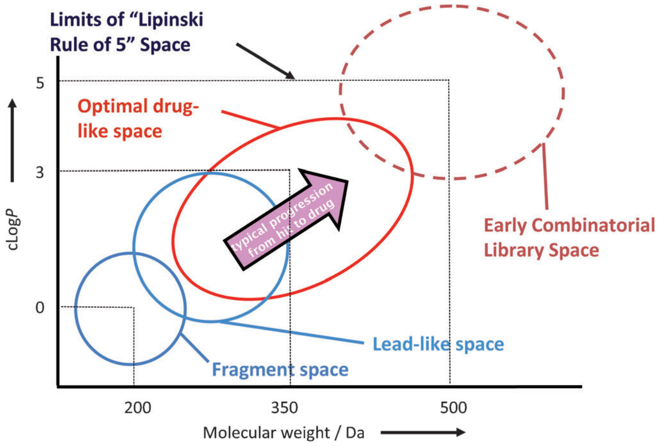

Figure 2. From recent analyses, optimal oral drug-like space can be defined in broad terms shown by the central, red oval. As optimization tends to progress by addition of complexity and lipophilicity (arrow), starting points should be in lead-like (or fragment-like) areas of property space.

overcome other issues. Exploiting molecules outside drug-like space is often a risky and niche approach but can nonetheless be successful when underlying pharmacology and mechanism of action are well understood.

2.1. Size and Complexity

An additional advantage of looking at smaller, less complex molecules for hits is that these molecules sample chemical space more efficiently[22,25,26] and have a greater chance of fitting a given drug binding site (albeit potentially with a weaker binding constant). It has been estimated that for every extra heavy atom added to an organic molecule, the number of biologically relevant potential structures increase by around a factor of 10[27–29] thus there are approximately 10^7 more molecules with mw $\approx$ 400 relative to those with mw $\approx$ 300, so by screening sets of lower molecular weight compounds a relatively greater proportion of accessible chemical space can be sampled thereby increasing the chance of finding hit molecules. The field of fragment-based drug design[30,31] takes this concept further using structure-based design and sensitive biophysical techniques to identify and optimize small-molecule fragments usually in the molecular weight range 100–250 Da. By analogy to the Lipinski Rule of 5, fragments are often described using a modified Rule of 3.[32]

2.2. Shape

In addition to molecular weight and lipophilicity considerations, a final parameter which has shown increased relevance to drug discovery recently[33,34] is the degree of three-dimensional shape and aromatic character of the molecule. A clear detrimental effect of increasing the number of aromatic rings in a molecule on its aqueous solubility has been noted[15,34] along with a further observation that as drug molecules progress through development, it is the more highly aromatic ones which preferentially suffer problems in development and ultimately fail.[33] Though the degree of aromaticity may also be related to the Log$P$ of molecules, there remains a clear signal that excessive levels of aromaticity are historically associated with undesirable outcomes and properties leading us to suggest a limit of a maximum of three aromatic rings on lead-like molecules and an aspiration that lead-like molecules should contain a lower proportion of sp^2 carbon atoms than typical historically. Given the preponderance of sp^2–sp^2 palladium-catalyzed cross-coupling methodologies which have made routine the synthesis of aromatic, aryl–aryl systems, one must wonder if this is an example of synthetic methodology which by means of its success and robustness has unwittingly negatively impacted the drug discovery process by facilitating the preparation of less drug-like molecules.

2.3. Sub-Structural Considerations

A further factor critically important in determining the "attractiveness" of a molecule as a starting point for drug discovery is the presence or absence of particular functional groups or sub-structures which are known to present issues to the drug discovery process. Many functional groups are obviously undesirable due to chemical stability issues (e.g. acid chlorides, most organometallics) but the large amounts of collective experience of medicinal chemists has also identified large numbers of groups which have the potential to bring problems to the drug discovery process such as potential for toxicity, unwanted interactions with biological systems or general instability. Summarizing these observations, in general, *a molecule is perceived as carrying risks if it suffers from some form of reactivity* which can be manifest in several ways:

1) Lack of chemical stability—the molecule decomposes on storage or in solution and its integrity cannot be guaranteed making interpreting biological data difficult and presenting significant quality control issues for production and clinical supply. Stability within biological systems is also often considered, e.g., esters are frequently cleaved enzymatically in vivo though metabolic instability is often something that can be addressed during the course of lead optimization.
2) Electrophilicity—electrophilic molecules often show a propensity to hydrolyse or react with bio-nucleophiles (e.g. DNA, cysteine, serine or lysine residues in proteins) leading to degradation of the drug and irreversible modification of biomacromolecules which may lead to unwanted immune responses amongst other issues. Some successful drugs are designed to react irreversibly[35] though only a minority of drug design efforts opt to use this mechanism.
3) Potential for redox chemistry—if a molecule is able to easily change oxidation state, the potential for unwanted interference in mitochondrial coupling systems exists. In addition, many bioassays utilize redox systems and redox-active small molecules may interfere and give confounding biological responses in these assays.[36]

Molecules which contain sub-structures suggestive of one or more of the issues outlined above represent less attractive starting points for drug discovery as they carry a real or theoretical risk which must be addressed in addition to all the other usual challenges encountered in turning a hit molecule into a drug. Because of this additional hurdle, these molecules are usually avoided by medicinal chemists: Given that the number of potential drug starting points is practically infinite, there should be no need to consider chemically reactive molecules.

Several attempts have been made to capture lists of these reactivity-based filters: A major public domain list of undesirable functional groups is that compiled by Shoichet, Simeonov et al.[37] and used at the NIH Chemical Genomics Center.[38] A set of thiol-reactive functional groups has been identified through application of NMR techniques by Hajduk et al.[39] whilst an additional set of chemotypes that often interfere in high-throughput assays has been identified by Baell et al.[40] For the removal of molecules containing undesired functional groups, GSK uses a comprehensive set of >300 structural filters which we term the GSK B filters. These filters have been defined and evolved over a number of years based on the collective experience of a wide range of medicinal chemists. Despite the detailed nature of these filters, all stem from the three concepts of undesired reactivity outlined above. The GSK B filters broadly mirror the themes seen in the NIH set and some of these filters we use in addition to the NIH sets are described in the Supporting Information (Table S1). For the analyses below, we focus mainly on the GSK B functional group filters (and give selected results from application of the NIH set).

2.4. Lead-Likeness

Bringing all these concepts together, we can define some simple properties which broadly describe lead-like space (Table 1). Whilst these criteria define molecules which give optimal quality starting points for a wide range of drug discovery programmes, molecules with properties outside these ranges still offer significant value to drug discovery. Indeed molecules with specific, non-lead-like or drug-like properties are sometimes needed to interact with particular types of biological target (e.g. protein–protein interactions). Despite this, GSK's experience on looking for hits against novel biological targets has shown that lead-like molecules generally act as a rich source of optimizable molecules.

Lead-likeness guide	Preferred Values
Lipophilicity guide	$-1 \leq \mathrm{cLog}P \leq 3$
Molecular size guide	$14 \leq$ heavy atoms ≤ 26 (mw=200-350 Da)
Undesired sub-structure filters	• Remove molecules containing chemically reactive, electrophilic or redox active groups. • Favor molecules with lower degree of aromatic character and/or more 3D shape.

In proposing these guides for lead-like space, we acknowledge the hazards of suggesting specific property ranges to define utility though we contend the guides here provide a useful description of lead-like space. *Molecules with properties that fall significantly outside lead-like space are much less likely to have high value as starting points for a wide range of drug discovery optimization efforts.*

Given the critical role of synthetic methodology in facilitating the synthesis of starting points for drug discovery we sought to assess how current and emerging synthetic methodology is suited to delivering the lead-like molecules drug discovery efforts optimally desire.

3. Analysis of Existing Commercial Compound Collections

A common source of novel starting points for drug discovery programmes used by academic and industrial medicinal chemistry groups is commercial compound vendors who provide a vast selection of small molecules for purchase in quantities suitable for biological screening. Some 4.9 million unique compounds from these vendors (Table S2, Supporting Information) were analyzed for lead-likeness. Our analysis (Figure 3) indicates that only a very small proportion (2.6%) of these commercial offerings are within lead-like space, as defined above, with the majority of compounds failing one or more of the filters relating to molecular weight, lipophilicity or undesirable substructures suggesting that commercial libraries, whilst potentially more drug-like, are not generally lead-like. A similar study by Shivanyuk and coworkers[41] on 7.9 million commercially available compounds suggested that 16% were lead-like—the larger proportion arising from their more generous physicochemical definition of lead-like space and more relaxed set of functional group filters.

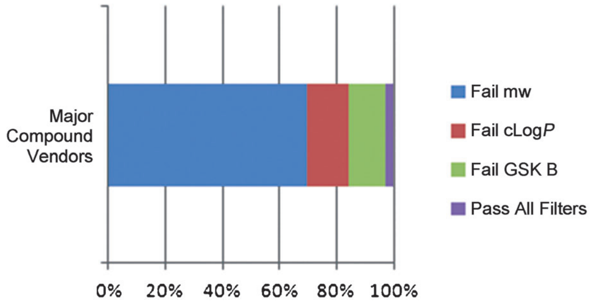

Figure 3. Analysis of lead-likeness of commercially available compounds from selected vendors ($n \approx 4.9$ million compounds) using GSK lead-likeness guides. The length of each colored band represents the effect of the successive application of various filters. When applied independently, each filter removes between 65–83% of the compounds. Approximately 69.5% of the data set fell outside of lead-like space due to molecular weight (mainly mw $\geq$ 360; blue); of those compounds remaining, around half had inappropriate Log$P$ (mainly Log$P > 3$, red bar) and the majority of the rest contained undesired functional groups (fail GSK B filters; green band), leaving 2.6% (ca. 125 000 compounds) (purple band) of the compounds that passed all the filters. See also Tables S3a,b in the Supporting Information.

In order for compound vendors to prepare and cost-effectively supply thousands of compounds in multi-mg amounts they rely heavily on robust, predictable methodology for compound production. To sample the range of existing and emerging methodology on which compound vendors and all practitioners of medicinal chemistry rely, we sought to analyze representative ranges of contemporary literature to identify the prevalence of methodologies able to facilitate the population of lead-like space.

4. Analysis of Synthetic Methodologies

In an attempt to measure the extent to which modern synthetic chemistry practice accesses lead-like space, we undertook three types of literature survey (Figure 4) across different subsets of synthetic chemistry literature published in 2009. In the first, we analyzed all of the reaction products reported in the 2009 volume of *Journal of Organic Chemistry* (issues 1–24). From some 1495 papers, ca. 32 700 compounds were indexed with "preparation" in CAPlus (via SciFinder[42] [metals, unusual isotopes, and commercially available molecules removed]). From these, a total of 13 194 compounds fell within our 200–360 molecular weight range of which 5267 compounds satisfied lead-like Log*P* limits (i.e. $-1 \leq \mathrm{Log}P \leq 3$). These compounds were subsequently passed through a variety of structural filters to remove reactive or otherwise undesirable functional groups. The NIH filters removed 28 % of compounds; the Baell filters 2 %; and GSK B 86.9 %. Using the GSK B filters as a measure of how experienced medicinal chemists would view the molecules, from $>32\,000$ reaction products reported in *Journal of Organic Chemistry* in 2009, only around 692 (2.0 %) (purple band, Figure 4) would be viewed as lead-like. Additional filters such as targeting more three-dimensionally shaped molecules and avoiding over-population of structurally similar molecules have not been applied but would reduce these figures further.

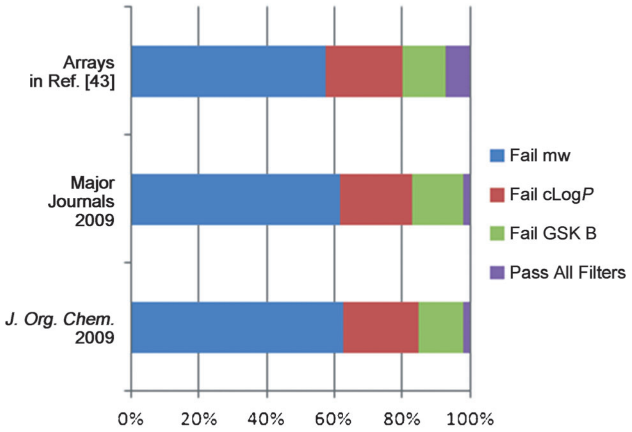

Figure 4. Analysis of lead-likeness in synthetic organic chemistry literature 2009. The effect on successive application of lead-like filters (filters as in Figure 3) on numbers of compounds from three sources of data are shown (see text for details). In all cases, a low proportion (2–7 %) of described compounds are lead-like. See also Table S4 in the Supporting Information.

Our second analysis surveyed a wider range of journals likely to contain significant amounts of new synthetic methodology (*Organic Letters*, *European Journal of Organic Chemistry*, *Journal of Organic Chemistry*, *Tetrahedron Letters*, *Tetrahedron Synthesis*, and *Synlett*), focusing on references with the words "new" and "synthesis" (or related terms) in the title/abstract. This analysis identified the preparation of 13 454 compounds which were filtered as above to give 249 (1.8 %) lead-like molecules.

Our final method of literature analysis made use of the annual survey of chemical library synthesis by Dolle and colleagues[43] in the *Journal of Combinatorial Chemistry* which provides a valuable snapshot of synthetic chemistry as it is currently applied to array and library construction (i.e. the application of robust chemistries). We analyzed the 2009 survey for lead-likeness in the following way. The 196 libraries not designed against specific biological targets (i.e. Tables 7–10 of ref. [43]) were visually inspected for lead-likeness. A total of 87 arrays with high scaffold molecular weight and/or undesirable chemotype were readily discarded, as it was clear the array would be unable to meet our definition of lead-likeness. The references for the remaining 109 libraries were inspected and the products for each (as indexed in SciFinder)[42] extracted and analyzed for lead-likeness as described above. Over half the arrays (61) still occupied non-lead-like space entirely (see Table S5, Supporting Information). However, from a total of 4926 compounds made in these 109 arrays, 353 (7.2 %) compounds met our lead-likeness definition. Despite this relatively low proportion of lead-like compounds, 9/109 arrays gave products (shown in Figure 5) of which $>30\,\%$ were lead-like, suggesting methodologies to produce the desired compounds robustly can indeed be identified. The main reasons for non-lead-likeness were an excess of lipophilicity, and the presence of overtly

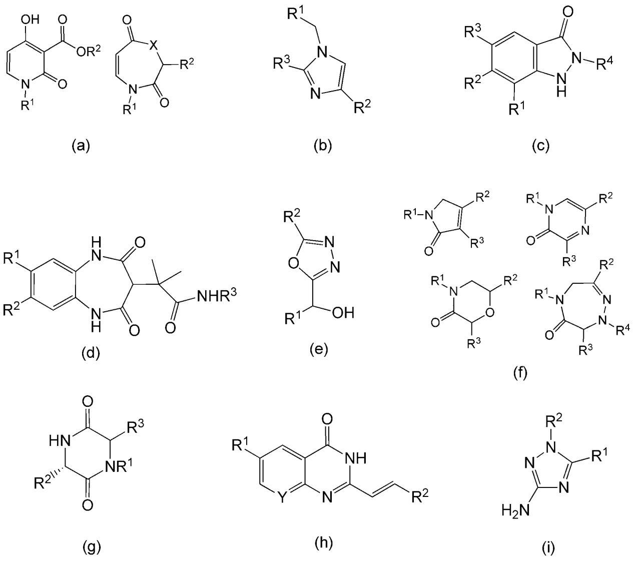

Figure 5. Some libraries[44–52] highlighted from ref. [43] with higher lead-like character. (X = N, O; Y = N, CH; R^1–R^4 = various substituents).

electrophilic centers (often Michael acceptors) in the final molecules. In some instances, a change in the choice of building blocks or capping groups to include smaller, more polar ones would lead to a more lead-like array. However, as we show below, synthetic chemistry in array format tends to discriminate against the successful synthesis of these compounds—it is quite possible that many of the arrays were designed to produce larger numbers of lead-like compounds, but have drifted out of it owing to synthetic attrition.

We acknowledge that the foregoing literature analyses have potential limitations but without subjective inspection of every published paper, the output of contemporary methodology development work is very difficult to summarize. Whilst we do not advocate that the synthesis of lead-like molecules should be the only or even main driver of synthetic methodology development, it is clear the physicochemical properties of the majority of molecules being prepared today are far removed from those of greatest value to the identification of new medicines suggesting that chemistries able to populate lead-like space are both limited and difficult to develop.

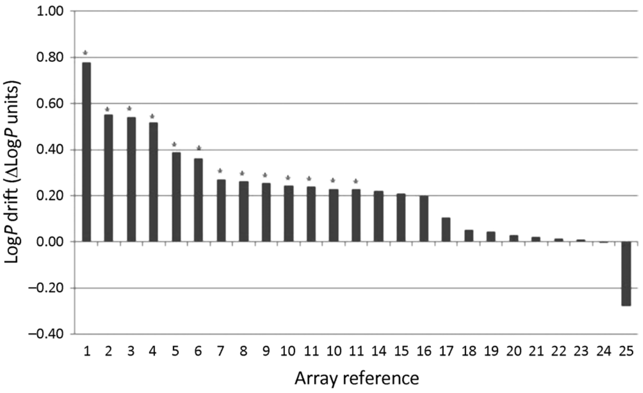

Figure 6. Mean Log*P* drift (ΔLog*P* units) for a range of 25 arrays. Mean Log*P* drift is the difference of the mean Log*P* of a completed array and the mean Log*P* of the designed array (* denotes $z < 0.05$). A positive Log*P* drift indicates that the set of compounds actually synthesized in an array were on average more lipophilic than those designed.

5. LogP Drift in Array Synthesis

To investigate further the potential for synthetic methodology to favor less lead-like molecules, we analyzed a diverse set of 25 arrays synthesized by GSK in 2010 where we knew both the array reactions which successfully delivered desired products as well as those which failed. Arrays are usually designed to have products with a wide spread of structural features and properties, but if array success is significantly below 100 % (a common occurrence) there is a scope for the Log*P*/mw profile of the completed array to deviate significantly from that of the originally designed array.

The 25 arrays studied used many of the robust chemistries most typically found in medicinal chemistry[53] and arrays,[54] such as reductive aminations, acylations, Pd-mediated cross-couplings and S_NAr reactions. The average designed array size was around 160 compounds (range 20–426), from which an average of 53 % of products were successfully isolated in pure form. Analysis of the Log*P* and molecular weight profiles of the completed and designed arrays (Table S6, Supporting Information) showed the mean mw of successful products did not change significantly from those of the designed arrays. Significant effects were however seen with Log*P*, the exact property medicinal chemists now seek to control so closely. For 23/25 of the arrays, the mean and median Log*P* of the completed compounds was higher than those designed (Figure 6), with the increase for 13/25 of the arrays being statistically significant ($z < 0.05$, one-sided z-test; see Table S7 in the Supporting Information for further data).

We refer to this clear trend of preferential success of higher Log*P* products from an array as "Log*P* drift". Although the magnitude of the mean Log*P* drift (ca. 0.22 log units, from 1.20 to 1.42) is not large, it is persistent and a sensitive marker of an array's outcome. The maximal Log*P* drift depends on the percentage of an array that is completed and the Log*P* distribution profile of the designed array. For an array covering a large range of Log*P* values, the mean drift of a partially completed array has the opportunity to be larger than one covering a smaller range of values. We calculated the Log*P* drift for each array as a percentage of the maximal theoretical drift based on the number of compounds successfully synthesized (i.e. a value of +100 % would mean that only the most lipophilic products were isolated whilst −100 % would suggest that only the least lipophilic products were seen). The mean of these figures is around +33 % (or +48 % for those arrays with $z < 0.05$, Table S8 and Figure S1, Supporting Information), representing a consistent and meaningful shift in the profile of an array towards a more lipophilic state relative to that designed. In other words, *the more polar products in an array tended to systematically fail more often in synthesis.*

This analysis seeks to identify trends in array outcomes only and several caveats prevent further, detailed conclusions being drawn. For example, some arrays receive a greater amount of pre-production validation and, once in production, a greater degree of manual intervention to complete the synthesis of more troublesome members of an array—effects which may both act to modulate any inherent Log*P* drift. Array methodologies carried out under differing conditions (solvents, reagents, etc.) may be differentially susceptible to Log*P* drift though the data set is too small to begin to identify such observations reliably. Whilst this effect is seen clearly here in arrays, one must speculate if the same factors are at work in determining the relative probability of success in discrete, non-array reactions which may disfavor the synthesis of molecules with preferred lead-like or drug-like properties and unwittingly favor synthesis of other, less attractive molecules.

Put another way, Log*P* drift is a quantification of the anecdotal observation that many methodologies are relatively intolerant to a wide range of functional groups in substrates. As introduction of functional groups usually increases polarity, it follows that this functional group intolerance leads to a polarity intolerance and selective production of less polar molecules. There are a number of likely causes of Log*P* drift, such as insolubility, poor extraction into organic solvents, coordination to metal catalysts, functional group incompatability and problems in chromatographic purification.

Our quantitative identification of Log*P* drift may offer an explanation as to why much modern synthetic methodology is developed in physicochemical space far removed from lead-like but the fact remains that applying such methodology to the synthesis of more polar molecules is often difficult or even impossible. This gap in the armory of methodology suggests that there is a great opportunity for practitioners of new synthetic methodology to develop and apply new approaches which will have immense impact on the drug discovery process.

6. Lead-Oriented Synthesis

Over the last two decades, advances in chemical methodology have identified novel ways of efficiently producing large numbers of novel molecules: Early combinatorial synthesis libraries relied on diverse capping groups for their overall diversity[55] but often, in their efforts to increase library size by the addition of multiple points of diversity, lie outside of lead-like space on grounds of size and lipophilicity. More recently, diversity-oriented synthesis (DOS), first described by Schreiber and co-workers[56] achieves diversity by elegant use of cascades producing large numbers of molecular scaffolds using a small set of transformations. Many early DOS libraries were again large in size[57] to allow for the high degree of diversification sought and have found applications in chemical biology and the discovery of chemical probes. It is too early to say whether any of these DOS approaches will yield successful medicines in the clinic. Though the lead-likeness of DOS libraries is frequently low they are instead often likened more to natural products which themselves have been a rich source of medicines[58] which may in some cases operate outside of Lipinski Rule of 5 space.

Thus, whilst existing approaches can efficiently produce diverse drug-like molecules, our ability to produce large numbers of smaller, attractive molecules remains limited prompting us to consider the concept of *lead-oriented synthesis* (LOS). In contrast to target-oriented synthesis, which targets just one compound; diversity-oriented synthesis, which targets scaffold diversity mainly in drug-like space; and combinatorial chemistry, which targets large numbers of compounds, lead-oriented synthesis must be able to deliver molecules with *specific molecular properties* with utility in the drug discovery and optimization process. As we have described above, lead-oriented syntheses need to pay particular attention to the physicochemical and functional group properties of the target molecules while also maintaining the synthetic efficiency to allow their cost effective utilization.

The challenges associated with designing successful LOS approaches are significant as witnessed by the low number of publications ideally suited to the preparation of lead-like molecules. Important factors which must be tackled in the development of new LOS sequences are summarized in Table 2.

Lead-oriented syntheses should:
Be able to produce a wide range of lead-like chemical structures
Be efficient and use cheap reagents and conditions to allow wide utilization in array formats
Not be susceptible to excessive Log*P* drift
Be tolerant to a wide range of polar functional groups
Produce molecules without multiple residual electrophilic or otherwise reactive centers

Tolerance towards polar substituents is perhaps the most important factor to address with many existing arrayable chemistries suffering in this regard (i.e. showing a positive Log*P* drift) making increased functional group compatibility a key goal. Despite this significant challenge some groups are now tackling this actively.[59] Unprotected polar functionality is often (though not always)[60] poorly compatible with many reagents owing to reasons of reactivity, insolubility in non-polar solvents, or coordination to a catalyst.

The drug discovery process seeks to identify molecules able to efficiently interact with biological systems and frequently these interactions are through polar or hydrogen-bonding interactions. These polar interactions tend to rely on functional groups such as weakly acidic OH and NH bonds or Lewis base/hydrogen bond acceptors which are typically the kinds of groups which also interact with chemical reagent systems leading to poor tolerance under reaction conditions. For lead-oriented syntheses to be of greatest value, they should be able to mediate novel transformations in the presence of such biologically relevant functional groups as for example, heterocycles (e.g. pyrazole, imidazole, pyridine, pyridone, etc.), acidic NH groups (e.g. amides, sulfonamides), other Lewis bases (e.g. nitriles, sulfones, etc.) or Brønsted bases (e.g. amines). Attractive lead-oriented syntheses should also ideally construct these biologically relevant sub-structures in a manner able to produce a diverse set of products with lead-like properties.

Advances in the range of chemistries which can be carried out in polar solvents (e.g. water) may help this, and some preliminary data (I. Churcher, unpublished observations) suggests that Log*P* drift may be lessened or even reversed for reactions carried out in largely aqueous media. With an emphasis on more polar reaction products we must also not overlook the importance of product isolation and purification which itself can be a contributing factor to Log*P* drift. Many traditional purification techniques ranging from normal and reverse phase chromatography through to simple organic/aqueous work-up procedures often encounter issues with excessively polar or water-soluble molecules which may prevent their efficient isolation.

An additional factor to consider is that potentially chemically reactive functional groups such as alkynes, esters, Michael acceptors, and nitro groups often remain as vestigial “scars” on molecules left over from a need to increase reactivity in a key bond-forming step. Whilst these reactive groups can in turn often be used to introduce further functionality or diversification, the presence of multiple such groups can necessitate extensive transformation to deliver attractive molecules devoid of undesired structural features, a process which reduces the attractiveness of the sequences. Similarly protecting groups (acetals, carbamates, silyl ethers) often remain in reaction products (note though the rise in traceless techniques)[61,62] requiring extra chemical steps for their removal which may severely limit the practical application in an array format. Lead-oriented syntheses which do not leave excessive residual undesired reactive centers or superfluous protecting groups are therefore of particular value.

7. Summary and Outlook

Medicinal chemistry relies heavily on robust synthetic methodologies to prepare libraries of potential starting points and to modify these molecules in the search for drug candidates, but there is a risk that much of our established methodology is not ideally suited to preparing molecules with optimal properties. There is evidence to suggest that current array chemistry relying on precedented synthetic methodology has an unintentional bias towards the synthesis of less drug-like products leading to a preponderance of these molecules. Contemporary methodology is not affecting this bias but by adopting the principles of lead-oriented synthesis, novel methodologies that address these issues can be identified and fully exploited.

Over recent years, the drug discovery community has better defined the kinds of molecules associated with increased clinical success but has perhaps not effectively communicated these observations to the key partners who so influence the drug discovery process, namely those groups working on novel synthetic methodologies. For drug discovery to increase its productivity, closer links must be forged between the synthetic chemistry and drug discovery communities to help better define the contemporary challenges and identify the best ways of tackling them. Whilst it would of course be inappropriate to suggest that all new methodology should be aimed at making molecules of direct interest to drug discovery, the application of new methodology to the preparation of bioactive molecules ultimately aimed at improving human health remains one of the major applications of organic synthesis in both academic and industrial laboratories. Given the increasing importance of the translational nature of much basic research, the drug discovery community can greatly help this by being clear on the kinds of molecules which will have the biggest impact on the search for new medicines.

As well as the property guides we have discussed, for maximum utility, methodologies must be robust, reproducible and possess a good substrate scope so that they may be used with confidence in the often time-constrained environment of drug discovery and/or be commercializable for preparing hundreds or thousands of attractive molecules. Without this robustness and predictability, methodology will have limited application to drug discovery.

The concepts of lead-oriented synthesis we describe here represent significant challenges. To develop new LOS methodologies will not be easy as finding methodology suitable to make large numbers of small, polar molecules is much more difficult than to make heavy, non-polar ones but this is a challenge the collective creativity of modern synthetic chemists is well placed to address. In discussions with a number of scientists at the forefront of the search for novel methodologies, there has been a willingness to recognize and tackle the challenges of LOS and a realization of the benefits this strategy can deliver for all.

The challenge now is to make the concept of LOS sustainable and impactful: many readers may be sceptical of the introduction of yet another concept to organic chemistry that describes a seemingly familiar phenomenon. We believe, however, there will be a step change increase in the utility and application of new methodologies which embrace the concepts of LOS. There is an opportunity for entrepreneurial synthetic organic chemists to profit from LOS, safe in the knowledge that the compounds they make will have greater application and a much higher utility in delivering progress in biomedical research in academic and industrial drug discovery settings. Should new, robust LOS approaches be identified, translation to the production of ranges of molecules for screening will be required which will perhaps lead to innovative funding models involving both the group initiating the methodology and commercial organisations with the ability to expand the availability and utility of the molecules.

Designers of new methodology wishing to discuss our thoughts on the lead-likeness of molecules derived fom their novel approaches should feel free to contact us.

The crisis of productivity of the drug discovery process has been well documented and dire warnings as to the very existence of the industry, which still employs large numbers of synthetic chemistry PhDs and graduates, have been made. To ensure the continued ability to develop and bring important new drugs to patients, the efficiency of the discovery process must improve and as we have shown here, synthetic chemistry can play a pivotal role in this improvement. Through closer awareness and collaboration, we hope the full might of novel synthetic methodology can be brought to bear to assist in the task of bringing new medicines to the patients that need them.

We are grateful to Professors Adam Nelson (Leeds), Tim Gallagher (Bristol), Tim Donohoe (Oxford), Rob Stockman (Nottingham), Steve Caddick (UCL), Hon-Wai Lam (Edinburgh), and Matt Gaunt (Cambridge) for useful discussions and to Stephen Pickett, Chris Luscombe, Simon MacDonald, Darren Green, Andrew Brewster, and Mythily Vimal (all GSK) for their contributions.

Received: August 18, 2011

[1] J. P. Garnier, *Harv. Bus. Rev.* **2008**, *86*, 68–70.
[2] I. Kola, J. Landis, *Nat. Rev. Drug Discovery* **2004**, *3*, 711–716.
[3] P. D. Leeson, B. Springthorpe, *Nat. Rev. Drug Discovery* **2007**, *6*, 881–890.
[4] P. D. Leeson, J. R. Empfield, *Annu. Rep. Med. Chem.* **2010**, *45*, 393–407.
[5] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Delivery Rev.* **1997**, *23*, 3–25.

[6] M. S. Lajiness, M. Vieth, J. Erickson, *Curr. Opin. Drug Discovery Dev.* **2004**, *7*, 470–477.

[7] M. P. Gleeson, *J. Med. Chem.* **2008**, *51*, 817–834.

[8] P. Gleeson, G. Bravi, S. Modi, D. Lowe, *Bioorg. Med. Chem.* **2009**, *17*, 5906–5919.

[9] J. D. Hughes, J. Blagg, D. A. Price, S. Bailey, G. A. DeCrescenzo, R. V. Devraj, E. Ellsworth, Y. M. Fobian, M. E. Gibbs, R. W. Gilles, N. Greene, E. Huang, T. Krieger-Burke, J. Loesel, T. Wager, L. Whiteley, Y. Zhang, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 4872–4875.

[10] Polar surface area is a calculated property of a molecule which sums the contribution to polarity from polar atoms (mainly oxygen and nitrogen and their attached hydrogen atoms). Utilities to calculate PSA are freely available, e.g. http://www.daylight.com/meetings/emug00/Ertl/tpsa.html. Relationships between polar surface area and the ability of molecules to cross biological barriers have been empirically observed. For example, molecules appear to cross cellular membranes most efficiently when PSA is <140 Å^2 whilst penetration into the central nervous system is usually most efficient for molecules with PSA <70–80 Å^2. See also "Polar Surface Area": P. Ertl in *Molecular Drug Properties* (Ed.: R. Mannhold), Wiley-VCH, Weinheim, **2007**, pp. 111–126.

[11] M. J. Waring, *Expert Opin. Drug Discovery* **2010**, *5*, 235–248.

[12] M. J. Waring, C. Johnstone, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 1759–1764.

[13] M. J. Waring, *Bioorg. Med. Chem. Lett.* **2009**, *19*, 2844–2851.

[14] Various free Log*P* calculation utilities are available, e.g. http://www.acdlabs.com/resources/freeware/chemsketch/logp/.

[15] A. P. Hill, R. J. Young, *Drug Discovery Today* **2010**, *15*, 648–655.

[16] M. C. Wenlock, R. P. Austin, P. Barton, A. M. Davis, P. D. Leeson, *J. Med. Chem.* **2003**, *46*, 1250–1256.

[17] J. R. Proudfoot, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1087–1090.

[18] P. D. Dobson, D. B. Kell, *Nat. Rev. Drug Discovery* **2008**, *7*, 205–220.

[19] S. Oswald, M. Grube, W. Siegmund, H. K. Kroemer, *Xenobiotica* **2007**, *37*, 1171–1195.

[20] P. D. Dobson, K. Lanthaler, S. G. Oliver, D. B. Kell, *Curr. Top. Med. Chem.* **2009**, *9*, 163–181.

[21] S. J. Teague, A. M. Davis, P. D. Leeson, T. Oprea, *Angew. Chem.* **1999**, *111*, 3962–3967; *Angew. Chem. Int. Ed.* **1999**, *38*, 3743–3748.

[22] W. E. Sneader, *Drug Prototypes and Their Exploitation*, Wiley, New York, **1996**.

[23] G. M. Keserü, G. M. Makara, *Nat. Rev. Drug Discovery* **2009**, *8*, 203–212.

[24] M. M. Hann, T. I. Oprea, *Curr. Opin. Chem. Biol.* **2004**, *8*, 255–263.

[25] M. M. Hann, A. R. Leach, J. N. Burrows, E. Griffen, *Compr. Med. Chem. II* **2006**, *4*, 435–458.

[26] M. M. Hann, A. R. Leach, G. Harper, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 856–864.

[27] J. L. Reymond, R. van Deursen, L. C. Blum, L. Ruddigkeit, *MedChemComm* **2010**, *1*, 30–38.

[28] K. T. Nguyen, L. C. Blum, R. van Deursen, J. L. Reymond, *ChemMedChem* **2009**, *4*, 1803–1805.

[29] L. C. Blum, J. L. Reymond, *J. Am. Chem. Soc.* **2009**, *131*, 8732–8733.

[30] M. Congreve, C. W. Murray, R. Carr, D. C. Rees, *Annu. Rep. Med. Chem.* **2007**, *42*, 431–448.

[31] R. A. E. Carr, M. Congreve, C. W. Murray, D. C. Rees, *Drug Discovery Today* **2005**, *10*, 987–992.

[32] M. Congreve, R. Carr, C. Murray, H. Jhoti, *Drug Discovery Today* **2003**, *8*, 876–877.

[33] F. Lovering, J. Bikker, C. Humblet, *J. Med. Chem.* **2009**, *52*, 6752–6756.

[34] T. J. Ritchie, S. J. F. MacDonald, *Drug Discovery Today* **2009**, *14*, 1011–1020.

[35] M. H. Potashman, M. E. Duggan, *J. Med. Chem.* **2009**, *52*, 1231–1246.

[36] L. A. Lor, J. Schneck, D. E. McNulty, E. Diaz, M. Brandt, S. H. Thrall, B. Schwartz, *J. Biomol. Screening* **2007**, *12*, 881–890.

[37] A. Jadhav, R. S. Ferreira, C. Klumpp, B. T. Mott, C. P. Austin, J. Inglese, C. J. Thomas, D. J. Maloney, B. K. Shoichet, A. Simeonov, *J. Med. Chem.* **2010**, *53*, 37–51.

[38] http://ncgc.nih.gov/projects/cruzain/Cruzain_qHTS_Supplemental_Table_Exclusion_Filters.xls.

[39] J. R. Huth, R. Mendoza, E. T. Olejniczak, R. W. Johnson, D. A. Cothron, Y. Liu, C. G. Lerner, J. Chen, P. J. Hajduk, *J. Am. Chem. Soc.* **2005**, *127*, 217–224.

[40] J. B. Baell, G. A. Holloway, *J. Med. Chem.* **2010**, *53*, 2719–2740.

[41] A. Chuprina, O. Lukin, R. Demoiseaux, A. Buzko, A. Shivanyuk, *J. Chem. Inf. Model.* **2010**, *50*, 470–479.

[42] Scifinder is available from CAS via various subscription levels at http://www.cas.org/products/scifindr/index.html.

[43] R. E. Dolle, B. Le Bourdonnec, K. Worm, G. A. Morales, C. J. Thomas, W. Zhang, *J. Comb. Chem.* **2010**, *12*, 765–806.

[44] J. F. Xu, X. Huang, *J. Comb. Chem.* **2009**, *11*, 938–942.

[45] J. Kočí, N. Pudelova, V. Krchnak, *J. Comb. Chem.* **2009**, *11*, 397–402.

[46] G. Dou, D. Shi, *J. Comb. Chem.* **2009**, *11*, 1073–1077.

[47] A. Shaabani, A. H. Rezayan, S. Keshipour, A. Sarvary, S. W. Ng, *Org. Lett.* **2009**, *11*, 3342–3345.

[48] M. Adib, M. R. Kesheh, S. Ansari, H. R. Bijanzadeh, *Synlett* **2009**, 1575–1578.

[49] N. Pudelová, V. Krchňák, *J. Comb. Chem.* **2009**, *11*, 851–859.

[50] C. R. B. Rhoden, D. G. Rivera, O. Kreye, A. K. Bauer, B. Westermann, L. A. Wessjohann, *J. Comb. Chem.* **2009**, *11*, 1078–1082.

[51] M. Baghbanzadeh, M. Molnar, M. Damm, C. Reidlinger, M. Dabiri, C. O. Kappe, *J. Comb. Chem.* **2009**, *11*, 676–684.

[52] J. Meng, P. P. Kung, *Tetrahedron Lett.* **2009**, *50*, 1667–1670.

[53] S. D. Roughley, A. M. Jordan, *J. Med. Chem.* **2011**, *54*, 3451–3479.

[54] T. W. J. Cooper, I. B. Campbell, S. J. F. MacDonald, *Angew. Chem.* **2010**, *122*, 8258–8267; *Angew. Chem. Int. Ed.* **2010**, *49*, 8082–8091.

[55] W. H. B. Sauer, M. K. Schwarz, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 987–1003.

[56] S. L. Schreiber, *Science* **2000**, *287*, 1964–1969.

[57] P. A. Clemons, N. E. Bodycombe, H. A. Carrinski, J. A. Wilson, A. F. Shamji, B. K. Wagner, A. N. Koehler, S. L. Schreiber, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 18787–18792.

[58] D. J. Newman, G. M. Cragg, *J. Nat. Prod.* **2007**, *70*, 461–477.

[59] P. Tosatti, J. Horn, A. J. Campbell, D. House, A. Nelson, S. P. Marsden, *Adv. Synth. Catal.* **2010**, *352*, 3153–3157.

[60] I. S. Young, P. S. Baran, *Nat. Chem.* **2009**, *1*, 193–205.

[61] M. S. Wiehn, N. Jung, S. Braese, *Power Funct. Resins Org. Synth.* **2008**, 437–465.

[62] W. M. Dai, J. Shi, *Comb. Chem. High Throughput Screening* **2007**, *10*, 837–856.