Journal of Medicinal Chemistry

PERSPECTIVE

pubs.acs.org/jmc

The Medicinal Chemist's Toolbox: An Analysis of Reactions Used in the Pursuit of Drug Candidates†

Stephen D. Roughley*,‡ and Allan M. Jordan§

‡Vernalis (R&D) Ltd., Granta Park, Cambridge, CB21 6GB, U.K.

§Cancer Research UK Drug Discovery Unit, Paterson Institute for Cancer Research, University of Manchester, Wilmslow Road, Manchester, M20 4BX, U.K.

1. INTRODUCTION

Industrial medicinal chemistry departments the world over are charged with the rapid delivery of small molecule new chemical entities (NCEs) into the screening process to facilitate the discovery of novel medicines to allow for the prevention, management, or cure of disease. While this headline aim seems straightforward on paper, the reliable, timely, and dependable synthesis of NCEs remains an unpredictable art that calls for the application of robust and reliable chemical transformations to best ensure chances of success and to help alleviate the bottlenecks often caused by synthetic tractability issues within a drug discovery program. It is little wonder that chemists have therefore developed a repertoire of transformations that, to a greater or lesser extent, can be relied upon to furnish the desired derivatives across a variety of chemotypes and in the presence of varied pendent functionalities. This collection of amassed knowledge and experience of robust transformations is often referred to as the "chemist's toolbox", into which they can delve to select the best synthetic strategy to furnish the desired chemical transformation. However, the exact contents of this "toolbox" naturally vary between individual chemists, based largely upon personal experience.

In a recent publication, we noted that "while a recent review has surveyed the reaction types most commonly used in the large scale manufacture of pharmaceuticals, no such review exists for those reactions favored in the small-scale synthesis of drug candidates. However, informal discussions between the authors and a small number of practicing medicinal chemists defined a number of key reaction types that were almost universally considered to be essential in the rapid synthesis of compounds for bio-assay".[1] While this small survey was useful for the intended purpose, we wondered how representative the findings would be of the work being conducted within the wider context of pharmaceutical R&D laboratories around the world. More interestingly, we considered whether this analysis would support many commonly expressed beliefs regarding the nature of this work.

In our experiences, there are several statements frequently expressed by those both outside and within the medicinal chemistry community. For example, discussions with other chemists have revealed that many of our drug discovery colleagues outside the synthetic community perceive our syntheses to consist of typically six steps, predominantly composed of amine deprotections to facilitate amide formation reactions and Suzuki couplings to produce biaryl derivatives. These "typical" syntheses invariably result in large, flat, achiral derivatives destined for screening cascades. We believed these statements to be misconceptions, or at the very least exaggerations, but noted there was little if any hard evidence in the literature to support our case. To this end, we determined to analyze the reaction types used in the pursuit of novel drug candidates and evaluate their frequency of occurrence, alongside other factors such as drug likeness, chirality, and the number of steps to each derivative.

Such a survey can never be truly comprehensive because of a multitude of factors. For example, company confidentiality means a substantial proportion of intellectual output remains within the confines of the organization and never enters the public domain through publication. Furthermore, the wealth of information in the literature is too great to wholly encompass in an analysis such as this. To this end, we elected to analyze a representative subset of the literature that we felt offered a sensible and manageable snapshot of the types of chemistries being applied to medicinal chemistry problems within the pharmaceutical industry. In parallel with the related analysis of reactions used to produce drug candidates themselves on large scale,[2] we elected to analyze the published output from the medicinal chemistry departments of GlaxoSmithKline, Pfizer, and AstraZeneca, assuming that this would offer a wide summary of different therapeutic areas and chemical transformations and allowing a direct and meaningful comparison between the two ends of the chemistry effort in drug discovery and development. We determined that a survey of the three highest impact dedicated medicinal chemistry journals would best represent this output [*Journal of Medicinal Chemistry* ("JMC"; American Chemical Society, impact factor 4.898), *Bioorganic and Medicinal Chemistry* ("BMC"; Elsevier, impact factor 3.075), and *Bioorganic and Medicinal Chemistry Letters* ("BMCL"; Elsevier, Impact factor 2.822)][3] and chose to focus our analysis upon publications dating from 2008, the latest year for which the entire output was available at the start of the analysis process in late 2009. During the preparation of this manuscript, workers at GSK published a related analysis covering 4800 reactions performed specifically during the lead-optimization phase in the Respiratory CEDD at GSK, including reactions performed in the high-throughput parallel array synthesis of compound libraries.[4]

While offering only a representative overview of the reaction types employed and the products produced, this perspective article offers some insight into the chemistries regularly employed in pharmaceutical R&D laboratories around the world. Not only may this analysis prove or dispel some of the many myths and preconceptions that surround the work but we feel it may also prompt areas of further synthetic research, suggesting chemistries that are at present under-represented and in clear need of robust,

Received: February 18, 2011
Published: April 19, 2011

dx.doi.org/10.1021/jm200187y | *J. Med. Chem.* 2011, 54, 3451–3479

Table 1. Summary of References Analyzed

	AstraZeneca (AZ)	GlaxoSmithKline (GSK)	Pfizer	total
total papers	24	65	50	139
JMC	2	12	6	20
BMC	0	1	3	4
BMCL	22	52	41	115
reactions	1117	3897	2301	7315
test compounds	602	1801	1163	3566

reliable, and widely applicable methodology to allow concise and dependable access to new areas of chemical space.

It became apparent to us during the early stages of the reaction analysis process that the results of the analysis could be potentially biased by the common approach of most medicinal chemistry programs of synthesizing a common core motif and then performing multiple derivatizations of this core in order to generate useful structure−activity relationships (SAR) for the project team. The early synthetic stages would only be counted once but potentially be used in the synthesis of multiple compounds. To address this potential bias, we decided to employ in tandem a second analytical approach, in addition to this reaction-based method, in which we considered the functional groups present in the reported molecular structures, in particular, choosing those groups representing the final product of the reaction types used (i.e., reaction = amide formation, functional group = amide; reaction = Sonogashira coupling, functional group = arylacetylene; etc). This latter approach will also under-represent certain reaction types, as their product functional groups are generally not considered to be "druglike" (for example, O-sulfonylation is a widely used reaction due to the synthetic utility of the sulfonate ester products, but there are no sulfonate esters present in the compounds with pharmacological data because of this same reactivity). However, we believe that, in concert with the reaction-based analysis, it will provide a deeper insight into the chemistries employed by medicinal chemists.

2. DATA GATHERING

In order to identify relevant articles, CAS Scifinder[5] searches were performed by institution, using the keywords "AstraZeneca", "GSK", and "Pfizer", and refined by year 2008, followed by analysis by journal name and keeping only records from *Bioorganic and Medicinal Chemistry* (Elsevier, BMC), *Bioorganic and Medicinal Chemistry Letters* (Elsevier, BMCL), and *Journal of Medicinal Chemistry* (American Chemical Society, JMC). Manual analysis removed a small number of papers from this set that did not represent medicinal chemistry SAR publications. Several papers provided details of SAR of a number of compounds without providing synthetic details; in these cases, the compounds are included in the compound-based analysis but not in the reaction based analysis. The distribution of papers, reactions, and compounds among target classes is tabulated in the Supporting Information (Tables S1−S3); kinases and GPCRs (peptidic and aminergic) account for approximately half of the papers and compounds, with a broad spread of other target classes represented across the remainder.

Reactions were retrieved from SciFinder,[5] or in those cases where no reactions were abstracted for a paper, searches were repeated in the Beilstein database[6] on a paper-by-paper basis or, in a small number of cases, abstracted manually. Reactions were then analyzed manually for each paper and the results aggregated for analysis following the broad reaction classifications used in the analysis of reactions used in process (or development) chemistry.[2] In some instances, papers were abstracted with a comment along the lines "analogues prepared in this manner"; here, we referred to the original paper to elucidate how many variations on each step were performed.

For the analysis of compound features, compound structures were retrieved by searching the Beilstein database for each citation and selecting only those compounds where pharmacological data were abstracted (Beilstein query term "PHARM exists").[6] Structures were manually refined to remove synthetic intermediates and reagents that have pharmacological properties reported elsewhere, along with compounds abstracted from figures giving background to the biological target investigated. In a small number of cases, structures had not been abstracted, and in these cases they were entered manually. By use of these methods, the 7315 reactions and 3566 test compounds described in 139 publications were compiled into a data set upon which the required analyses could be performed (Table 1).

Numbers of screened compounds were determined by manual analysis of the original papers. The number of synthetic steps to each compound was determined manually from the text and schemes in the paper. Where more than one route to a compound was described, only the shortest route was included in this part of the analysis. Figures quoted are for the longest linear sequence from the starting material described in the paper, regardless of whether the origins of this material are described as being commercial sources, described elsewhere by citation, or not discussed. As such, they represent a minimum length for the synthetic sequence. It is impossible to ascertain without doubt what materials were available to the chemists performing the individual work from internal corporate collections, and so further elaboration of this is not possible.

"Lipinksi"[7,8] properties [MWt, H-bond donors (HBD) and acceptors (HBA), clogP,[9] number of rotatable bonds (NRot),[10] and topological polar surface area (TPSA)[10,11]], numbers of chiral centers, and the Fsp3 parameter recently proposed by Lovering et al.[12] were calculated using Python, version 2.6.5,[13] and RDKit.[14]

Functional group counts were generated using SMARTS[15] substructure query strings in RDKit. SMARTS strings were validated using SMARTSViewer.[16,17] Aromatic rings were counted following the system adopted by Ritchie and MacDonald,[18] whereby when a ring is part of a fused ring system, each ring is counted separately (e.g., purine = imidazole + pyrimidine). Individual aromatic ring types were counted using a set of SMARTS strings; SMARTS strings for all 6,6-, 6,5-, and 5,5-fused aromatic rings containing C, N, O, or S, whether synthetically feasible or not, were systematically enumerated computationally. Analysis of amine, amide, ether, and halogen environments was performed similarly, using further SMARTS query strings. Counts were aggregated using Knime Desktop, version 2.2.2.[19,20]

3. REACTION ANALYSIS

The reactions in the data set were classified into 10 major categories based on the overall reaction type, e.g., heteroatom substitutions (see Table 2), and within each category into further subtypes, e.g., N-alkylation, O-alkylation, etc. In the following sections, we will discuss the overall picture in the major categories and then discuss each category in detail. We conclude this

Table 2. Occurrence Rates of Reactions within the Data Set[a]

reaction type	no. of reactions	% of total	% of subtype
heteroatom alkylation and arylation	1687	23.1	
N-substitution with alkyl-X	390		23.1
reductive amination	386		22.9
N-arylation with Ar-X	458		27.1
amide N-alkylation	49		2.9
aniline N-alkylation	1		0.05
heteroaryl N-alkylation	44		2.6
O-substitution	319		18.9
S-substitution	30		1.8
acylation and related processes	1635	22.4	
N-acylaton to amide	1165		71.3
N-sulfonylation	163		9.9
N-acylation to urea	155		9.5
carbamate/carbonate formation	42		2.6
amidine formation	4		0.2
O-acylation to ester	13		0.8
O-sulfonylation	75		4.6
other	18		1.1
C−C bond formation	841	11.5	
Suzuki coupling	338		40.2
Heck reaction	3		0.4
Sonogashira reaction	155		18.4
Stille reaction	17		2.0
other Pd-catalyzed reactions (Negishi, Kumada, etc.)	11		1.3
ester condensation	46		5.5
Grignard	47		5.6
Wittig olefination	36		4.3
other organometallic (e.g., organolithium)	30		3.6
Friedel−Crafts acylation	27		3.2
other	131		15.6
heterocycle formation	601	8.2	
N-containing	537		89.4
O-containing	54		8.9
S-containing	10		1.7
protections	225	3.1	
deprotections	1319	18.0	
reductions	406	5.6	
NO_2 to amine	78		19.2
amide to amine	53		13.1
CN or imine to amine	21		5.2
ester to alcohol	48		11.8
ketone to alcohol	52		12.8
alkene to alkane	46		11.3
alkyne to alkane	9		2.2
aryl/hetaryl to fully sat.	6		1.5
other	93		22.9
oxidations	110	1.5	
alcohols to aldehydes	38		34.5
at sulfur	25		22.7
alcohols to acids	16		14.5
at nitrogen	7		6.4
alkene oxidative cleavage	4		3.6
benzylic/allylic oxidation	1		0.9
alkene oxidation	1		0.9

Table 2. Continued

reaction type	no. of reactions	% of total	% of subtype
oxidations (continued)			
other	18		16.4
functional group interconversion (FGI)	413	5.6	
alcohol to halide	68		16.5
amide to imidoyl chloride	7		1.7
acid to acid chloride	39		9.4
nitrile to acid	5		1.2
dehydration	20		4.8
carbonyl to C=N	26		6.3
other	248		60.0
functional group addition (FGA)	78	1.1	
halogenation	37		47.4
nitration	3		3.8
sulfonation	2		2.6
other	36		46.2
total	7315		

[a] Protecting group manipulations are detailed in Table 3.

part of the analysis with a brief discussion of the "top 10" reactions reported in the data set.

3.1. Overview of Classes. Table 2 shows the number of reactions abstracted in each broad class. At this level of detail, the obtained figures correspond well to those outlined in the analysis of reactions used in larger-scale drug candidate synthesis.[2] Some clear trends become evident when considering this summarized data set alone.

The formation of carbon−heteroatom bonds account for almost half of all reactions (45.5%) in the data set, being almost equally divided between acylation reactions (amides, ureas, sulfonamides, etc.; see below) and alkylation and arylation reactions, perhaps providing an initial indication of the perceived dominance of reductive aminations and amide formations in medicinal chemistry. By contrast, only a little over $^{1}/_{10}$ of the transformations resulted in the formation of carbon−carbon bonds. This latter figure is perhaps surprising in view of the supposed preponderance of palladium-catalyzed cross-coupling reactions in medicinal chemistry. Despite advances in chemoselective and tolerant reaction methodologies, the practicing medicinal chemist still appears to rely heavily upon protecting group strategies to enable the construction of the desired chemical entities. While clearly inefficient in terms of time, reagents, and overall yields,[21−24] these facilitating processes account for a further $^{1}/_{5}$ of all reactions, with this figure being dominated by deprotection reactions.

Despite the broad utility of heteroaromatic ring systems as core scaffolds within medicinal chemistry programs,[25−27] only 8.2% of reactions involve heterocycle formation. This must reflect either their formation early in a synthetic scheme or their purchase from commercial sources and subsequent diversification. The remainder of the reactions are accounted for by oxidation state adjustments and other functional group manipulations.

3.2. Heteroatom Alkylation and Arylation. Chemists clearly favor reactions involving heteroatom alkylation and arylation for which robust and reliable methods are readily available and offer a diverse set of widely applicable protocols (Table 2). Almost 80% of all heteroatom alkylation and arylation reactions are derivatizations of nitrogen atoms, with the large majority of the remainder being oxygen functionalization. In view of this and in line with the precedent from the earlier process chemistry analysis,[2] the nitrogen functionalization reactions were further subdivided by reaction type.

3.2.1. N-Alkylation. Alkylation of nitrogen with alkyl halides, despite the contingent issues with overalkylation and side reactions, accounts for almost a quarter of transformations in this class. Reductive aminations,[28,29] offering similar products with a greater degree of control over reactivity and the product thus obtained, are surprisingly only equally popular. These two methods together, however, account for half of all heteroatom derivatizations. It is worthy of note that in a recent paper of lead optimization and array chemistry at GSK,[4] to the surprise of those compiling the array analysis, no examples of reductive amination were reported during the survey period.

3.2.2. N-Arylation with Ar−X. Despite the popularity of the above two transformations, the most common C−N bond forming reaction in this section is the aromatic substitution of an aryl−halogen species by a nitrogen, accounting for more than one-quarter (27%) of the reactions in this category. This displacement of aromatic halogen atoms with amines, both by "traditional" nucleophilic displacement (S_NAr and related mechanisms, e.g., ANRORC[30,31] and $S_{RN}1$[32]) and by Buchwald−Hartwig palladium-catalyzed processes,[33−40] highlights the utility and applicability of this transformation. This utility is no doubt aided by the ease of preparation of the precursor halogenated heterocyclic moieties and their widespread commercial availability, itself almost certainly encouraged by their utility in C−C bond forming processes (see below).

3.2.3. Amide and Heteroaryl N-Alkylation. While amide formation remains ubiquitous (see below), further derivatization of this functionality by N-alkylation is notably less commonplace, with few examples of amide N-alkylation appearing in the analysis. While it is a common SAR manipulation to investigate the importance of a hydrogen bond donor in the derivative, this transformation does not appear to be routinely employed to explore the SAR around the amide bond. This is most likely a result of the ease and popularity of reactions such as reductive amination to prepare the substituted amine prior to the amide formation step. Similarly limited is the N-alkylation of nitrogen containing heterocyclic cores, perhaps limited either by the

important role such unsubstituted systems play as hydrogen bond donors in their interactions with biological macromolecules or by issues of regioselectivity.

3.2.4. Alkylation of Other Heteroatoms. O-Substitution reactions account for almost $^{1}/_{5}$ of all heteroatom substitutions, although this figure includes O-arylations along with alkylation by both the Mitsunobu reaction[41,42] and alkyl halide and sulfonate reagents.

In terms of N-substitution reactions, the frequencies observed largely reflect those reactions employed in the larger scale synthesis of drug candidate molecules,[2] where N-alkylation and reductive amination reactions are highly prevalent. In contrast, both S- and O-substitutions appear to be less common in the research setting compared to the synthesis of molecules selected as potential drug candidates. This cannot be readily attributable to the ease of synthesis of such compounds, as in general there are fewer possible side reactions than in the corresponding N-substitution transformations, and it was anticipated that such couplings would be more commonplace in the small-scale laboratory environment.

3.3. Acylation Reactions. The acylation class of reactions (including N- and O-substitutions with a variety of carbon- and sulfur-based electrophiles) accounts for 22% of the reported reactions in our data set. As such, it occurs at a very similar rate compared to the heteroatom alkylation and arylation reaction types. However, in contrast to that group, the acylation class is dominated by amide formations, which account for over 70% of such processes (Table 2).

3.3.1. Amide Formations. In line with the common perception described in the introduction, amide formation is clearly the most numerous reaction, both within the acylation reactions class (in which 7 out of every 10 reactions are amide formations) and in absolute reaction count across all classes (see below). Likely explanations for this include the wide range of robust methodologies available for the synthesis of amide bonds, as a result of the efforts made in the area of peptide synthesis,[43,44] the ready availability of starting materials by a range of synthetic methods, and the relative ease of purification of the reaction products, factors that contribute to the amenability of the reaction to high-throughput parallel synthesis.[4] We found a wide variety of reagents and conditions used to accomplish this transformation within this data set, with no clear preference among the community for a particular reagent.

3.3.2. Other Acylation Reactions. Additional N-acylations (including the analogous sulfonylations) are also well represented, with the SAR-informative *N*-sulfonamides and ureas being prepared in roughly similar proportions, between them accounting for $^{2}/_{3}$ of the remaining reactions in this class. O-Sulfonylation reactions are the next largest group, accounting for almost 5% of reactions in this category, almost invariably representing the formation of mesylate (OMs), tosylate (OTs), and more rarely, triflate (OTf) sulfonate esters to serve as leaving groups in heteratom alkylations (directly or via conversion to a halide first) and, to a lesser extent, as cross-coupling partners in Pd-catalyzed processes. The remaining reaction types within this class are considerably less well exemplified, most likely because of a combination of their in vivo lability (which is often deliberately exploited in the design of prodrugs[45,46]) or strong basicity. While O-acylation is described both here and elsewhere, the examples listed here are distinct from those O-acylations where the sole aim is protection of functionality. In the instances listed above, the O-acylation is a deliberate step to introduce the ester functionality, either into compounds for biological evaluation or as a synthetic handle for further manipulation, rather than the addition of a protecting group.

The metrics regarding the acylation reactions discussed in sections 3.3.1 and 3.3.2 recapitulate very closely those described in the literature for the synthesis of drug candidates on the process scale,[2] further highlighting their importance in the preparation of biologically relevant molecules. However, it is difficult to ascertain whether this is the case because (i) the product functionalities are critically and uniquely important for the resultant biological interactions or (ii) their inclusion is self-selecting in the early stages of compound evolution because of their broad applicability and reliability, leading to a readily accessible subset of molecules having desirable profiles.

3.4. C−C Bond Formation. Data presented in the literature surveyed clearly demonstrate that palladium-mediated C−C bond formation is the methodology of choice for the construction of such bonds in small molecules for biological evaluation, accounting for almost two-thirds of all transformations in this class. The diversity of available starting materials (a search of the Available Chemicals Directory[47] reveals ∼6600 aromatic boronic acids alone, along with 4800 boronate esters and 750 trifluoroboronate salts and ∼660000 chloro-, bromo-, or iodoarenes), chemoselectivity, and tolerability of these reactions clearly adds to their utility and attractiveness. Additionally, the reactions are, in general, readily amenable to parallel synthesis operations, often in contrast to the other reactions in the C−C bond forming category. The Suzuki (or Suzuki−Miyaura) cross-coupling reaction[40,48−50] is the single most numerous reaction within the C−C bond forming group, accounting for 40% of all such reactions. This popularity is almost certainly due to the almost unique combination of reagent stability[51] and safety.[52] Alternative Pd-catalyzed processes are generally only employed when issues of stability or reactivity prevent a successful outcome in a Suzuki reaction.

The Sonogashira reaction,[53−55] in which a terminal acetylene is coupled directly with a suitable haloarene or haloalkene, ranks surprisingly highly in this list at 18%, the second largest within the category. Again, this is likely to be a combination of the relatively benign nature of the starting materials and the utility of the acetylene linker, both in its own right and as a precursor to other functionalities.

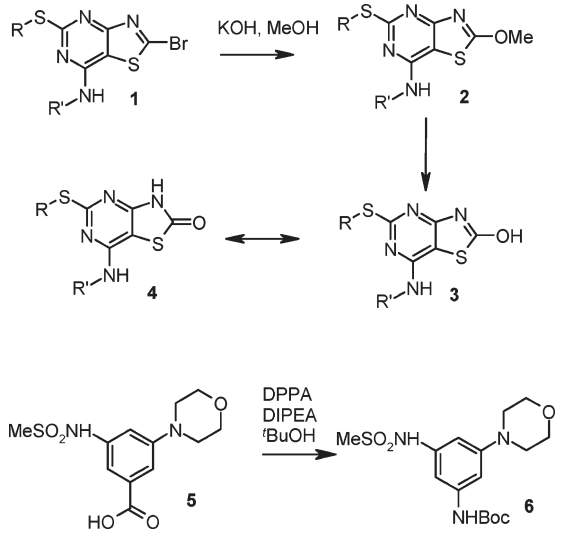
These couplings are not without limitations and care must be taken to ensure that residual palladium does not cause spurious noncompound related artifacts, particularly as a compound progresses into more advanced cellular and in vivo assay systems, where palladium toxicity may prove confounding.[56−58]

Outside the scope of these couplings, a variety of other C−C bond forming reactions are widely represented to a lesser extent, with selected examples of ester condensations, Grignard reactions (including Weinreb amide-style methodology[59−61]), and Friedel−-Crafts acylations comprising the majority of the described transformations. Wittig reactions[62−64] and Grignard reactions,[59−61] both of which we highlighted in our informal survey,[1] each account for around 5% of C−C bond forming reactions (and therefore around 1 in 200 of all reactions), making them some of the most common non-Pd-mediated C−C bond forming reactions but relatively rare overall. We have grouped the various other metal-based processes together in a single "organometallics" category, in contrast to the review of process chemistry reactions,[2] as the individual subtypes were very rare, with only a few examples of each. Among other methodologies not listed specifically in Table 2 are a wide variety of transformations, predominantly aldol-type condensations.[65−70] Once again, the common underlying theme with the majority of these transformations is the accessibility of a wide range of starting materials, some tolerability of other functionality, and the robust,

Table 3. Detailed Analysis of Protecting Group Manipulations within the Reaction Data Set

	protection		deprotection	
protected functionality/group	no.	%	no.	%
NH total	88	39.0	608	46.2
N-Boc	73	32.4	357	27.1
N-Bn	0	0.0	29	2.2
N-Cbz	1	0.4	10	0.8
other NH	14	6.2	212	16.1
RCO_2H	92	40.9	395	29.9
ROH total	41	18.2	279	21.1
OBn	1	0.4	19	1.4
$OSiR_3$	11	4.9	89	6.7
OAc	3	1.3	12	0.9
other OH	26	11.6	159	12.1
RSH			21	1.6
others	4	1.8	16	1.2
total	225	100	1319	100

Scheme 1. Introduction of Protected Functionalities by Non-Protection Routes[a]



[a] Top: Introduction of a methyl ether by S_NAr in a synthesis of CXCR2 antagonists (AZ) and subsequent deprotection.[71] Bottom: Introduction of a Boc-protected amine by Curtius degradation of a carboxylic acid in the synthesis of EphB4 tyrosine kinase inhibitors (AZ).[72]

reliable nature of the chemistry involved. The large proportion of reactions in the "other" category reflects the centrality of the C−C bond forming reaction category to organic synthesis and the diversity of methods available for its execution.

3.5. Heterocycle Formation. A heterocycle synthesis in which multiple different heteroatoms are present in the synthesized ring was classified under each; i.e., pyrroles and imidazoles are classified as N-containing, but thiazoles are classified as both N-containing and S-containing, following the same convention used in the analysis of development chemistry reactions (Table 2).[2] From this analysis, it can be seen that N-containing heterocycle synthesis dominates, accounting for almost 90% of this reaction class, with the majority of the remaining reactions being O-containing heterocycles syntheses. The analysis of compound substructures later in this manuscript discusses in more detail the individual heterocyclic systems reported within this data set.

3.6. Protecting Group Manipulations. Across the protecting groups encountered in the data set (Table 3), all examples showed considerably more (up to 20-fold) deprotections than corresponding protection steps. This trend is in agreement with that reported for process chemistry.[2] There are several explanations for this trend: (1) use of ready-protected commercial building blocks; (2) use of ready-protected noncommercial building blocks from internal archives; (3) the protection of a common motif, followed by parallel diversification and subsequent deprotections of multiple analogues; (4) the introduction of protected functionality by other means, with later deprotection; examples include displacement of a halogen with nitrogen, oxygen, or sulfur nucleophiles[71] and the Curtius degradation of acids to Boc-protected amines (see Scheme 1 for examples).[72] It is noteworthy that ~80% of all protecting group manipulations are related to one or the other component of the amide formation reaction (i.e., acid or amine), reflecting the dominance of this reaction, although it is important to emphasize that this is not the sole use of such products.

Protection of carboxylic acids and NH groups occur in similar numbers, accounting for 80% of all protection reactions. Boc protection of NH groups dominates, with no benzylations and only a single example of a Cbz protection in the data set. The corresponding deprotections show greater diversity, although unsurprisingly Boc still dominates. This is almost certainly due to the broad availability of Boc-protected reagents and the associated simple protection and deprotection conditions.

Carboxylic acids are invariable protected as simple esters, predominantly methyl and ethyl, while occasional use is made of benzyl and *tert*-butyl esters when alternative deprotection conditions are required.

For protection of alcohol and phenol groups (which account for ~20% of both protection and deprotection reactions), silyl ethers are the biggest single category accounting for around one-quarter of all OH protections and around one-third of OH deprotections. Over half of all OH protections and deprotections are categorized as "other", suggesting that beyond the silyl ethers there is a broad range of protection strategies used in this class, often reflecting a combination of "personal favorites" and commercial availability of preprotected starting materials. The corresponding sulfur analogues are more rarely encountered; there are no examples of S-protection in the data set and only 21 deprotections (1.6% of total deprotections), 19 of which are from a single paper from GSK describing a series of thiol-based ACE2 inhibitors.[73] This is almost certainly a consequence of the reactive nucleophilic nature of the deprotected thiol groups, which can disrupt disulfide bridges within proteins and undergo covalent conjugation with free cysteine thiol groups (either directly or via breaking of an existing cystine disulfide bridge), leading medicinal chemists to avoid their incorporation.[74]

3.7. Reduction Reactions. The most common transformation within this group is the reduction of the nitro group to an amine (Table 2), often employing the precursor both as a regiochemical director of installed functionality and as a masked amine group, ready to be derivatized via acylation or reductive amination. Similarly, the reduction of amides, nitriles, and imines are also commonplace, in this instance generating a homologated amine that can also provide useful structure−activity information. Overall, reductions of various functionalities to amino

groups correspond to at least 37% of reactions in this group (see below for "other reductions"), again pointing to the centrality of amine-derived products within medicinal chemistry. These data do not include reductive aminations (another 386 reactions), as in general the imine intermediate is not isolated.

The reduction of ester to alcohol occurs with a frequency similar to that of the analogous amide to amine reduction. Reduction of aldehydes and ketones to alcohols also occurs at a similar level, and in total OH-forming reductions account for almost 25% of reductions.

Interestingly, the reduction of alkene to alkane (11.3%) is much more prevalent than the reduction of alkyne to alkane (2.2%). It has been supposed that both the Heck[75] and Sonogashira[53−55] palladium-mediated couplings are primarily employed specifically to form such alkane linkages postcoupling by simple hydrogenation. However, these figures do not reflect this hypothesis, being inversely proportional to the frequency of such cross-coupling reactions (the Sonogashira coupling is 50-fold more prevalent in this data set than the Heck reaction (Table 2)). The number of Sonogashira reactions (155) far outweighs the instances of alkyne reduction (9), suggesting that the alkynes so-formed are incorporated for inclusion in the final products or for nonreductive derivatization. Conversely, alkene reductions (46) outnumber Heck reactions (3). The most likely explanation for this is that the alkene group is synthesized by alternative methods, such as the Wittig reaction[62−64] or aldol/elimination reaction.[65−70]

Other reductive processes not specifically mentioned in the text, which account for almost 23% of reductions, predominantly included examples of nitroso reductions to the corresponding amine and heterocyclic scaffold reformation, such as the reductive cleavage of an isoxazole and subsequent recyclization to a pyrimidine or pyrazole, allowing expansion of SAR by facile scaffold-hopping.

3.8. Oxidation Reactions. Perhaps the most striking initial observation regarding oxidation reactions is how much rarer they are than reduction reactions (only 110 oxidation reactions compared with 406 reductions (Table 2)). The reasons for this are likely to be complex and depend on the availability of starting materials, along with the hazardous nature of oxidizing reagents in general, coupled with environmental concerns surrounding the disposal and toxicity of the heavy metals employed in many oxidations.

The most common member of this class (35%) is the oxidation of alcohols to aldehydes, useful intermediates for both C−C bond forming reactions and reductive amination reactions. It is noteworthy that this oxidation state is accessed entirely by oxidation within this data set, with no reports of the controlled reduction of nitriles or esters directly to the aldehydes. Many of the alcohol precursors for this oxidation were obtained by reduction of the corresponding ester to the alcohol prior to reoxidation. It is likely that this two-step procedure is preferred, as the aldehydes are generally more reactive than the corresponding esters, resulting in difficulty controlling the reduction route, whereas there are many good methods for controlled oxidation that offer relatively mild selective conditions to achieve the required transformation without significant overoxidation. Clearly, a direct reduction method to aldehydes with the operational simplicity and reliability of the reductions to alcohols would provide an improvement in synthetic efficiency in these cases.[78]

Following this are oxidations occurring at sulfur, sulfide to sulfoxide and to sulfone, accounting for 23% of oxidations. Both of these products are themselves most likely to be employed as leaving groups in S_NAr reactions, where they perform better than their thioether precursors.[76,77]

Oxidation of alcohols to acids is the next biggest single oxidation type reported (16%), again highlighting the importance of the acid group within medicinal chemistry as a precursor to amides and heterocyclic compounds in particular. A broad range of other oxidation processes are also represented, although all at lower levels than the above. It is clear that in the majority of cases, oxidation is used as an adjustment prior to further reaction.

3.9. Functional Group Interconversions. Given the prevalence of certain key reaction types already noted, it is hardly surprising that the key functional group interconversions identified by this analysis are the conversion of an alcohol to a halide (16.5%, Table 2) and carboxylic acid to acid chloride (9.4%), given that these functionalities can both be employed in the highly prevalent derivatization of amines by alkylation or acylation reactions, respectively. The vast majority of examples within this category come in the "other" category, representing the diversity of this group.

Within the "other" category, a number of reactions are featured on a regular basis: (1) diazotization reactions of anilines and their subsequent conversion to iodoarenes (the Sandmeyer reaction); (2) halogen displacement with CN^-; (3) Pd-mediated alkoxy- and amino- carbonylations to esters and simple amides. While these last two transformations could be considered as carbon−carbon bond forming reactions, they have been included here rather than the alternative section, as their use was primarily to adapt or interchange pendent functionality, as opposed to constructing the molecular scaffold.

By far the most prevalent transformation not specifically listed above, however, was the conversion of aromatic halides to boronic acids and boronate esters prior to Suzuki coupling reactions.[48,49] This accounted for almost half of the unlisted "other" transformations, predominantly from a single citation within the data set.[79]

3.10. Functional Group Additions. Halogenations dominate this facet of chemistry, and akin to the functional group interconversions discussed above, their occurrence is anticipated given the prevalence of Pd-mediated couplings observed earlier that depend upon the routine access to halogenated precursor materials. They also benefit from a broad range of reagents, many of which show mild and selective reactivity, in contrast to the other reactions in this group. More surprising, given the utility of the resultant functionality, the incorporation of a nitro moiety is uncommon in the literature surveyed for this analysis. As the nitro to amine reduction is relatively common (78 reactions in the data set), it would appear that nitrated materials, in the context of the syntheses discussed herein, are generally purchased or obtained by prior synthetic manipulation of commercially available nitro compounds rather than being prepared by nitration as part of the published reaction sequence. It is likely that this is due to a combination of the potentially hazardous nature of the nitration reaction, along with the diversity of available nitro-containing building blocks (a search of the Symyx Available Chemical Directory[47] reveals more than of 97 000 nitroarene-containing compounds).

Other notable functional group additions not specifically listed are the introduction of nitroso functionality as a masked amine, which may also go some way toward explaining the surprising rarity of nitration processes, and the introduction of an enamine substituent using dimethylformamide dimethyl acetal as a precursor to heterocyclic ring formation.

Table 4. Top 10 Reactions by Frequency in the 2008 Data Set

reaction	no. of reactions	% of all reactions
N-acylation to amide	1165	16.0
N-containing heterocycle formation	537	7.4
N-arylation with Ar-X	458	6.3
RCO_2H deprotection	395	5.4
N-subs with alkyl-X	390	5.3
reductive amination	386	5.3
N-Boc deprotection	357	4.9
Suzuki cross-coupling reaction	338	4.6
O-substitution	319	4.4
other NH deprotection	212	2.9
total	4557	62.4

3.11. Summary of Reaction Types and a "Top 10" Reactions List. The literature data set shows that while medicinal chemists rely heavily on a relatively small number of reaction types, the remainder of the reactions used covers a wide range of different types in order to achieve the goal of discovering new drug candidate molecules. In general terms, heteroatom alkylation and acylation reactions account for almost half of the reported reactions. Protecting group manipulations account for ~20% of reactions, whereas C−C bond forming and heterocycle-forming reactions each account for ~10% of reactions. However, within this set, a small number of processes dominate (Table 4); notably, ~1 in 6 of all reactions in the data set was an amide formation. While this figure is lower than has been suggested (various sources quote figures of up to 50% of reactions being amide formations[80]), it is still remarkably high for a single transformation. Indeed, amide formations alone account for more reactions than any of the other broad categories except for heteroatom alkylation/arylation and deprotections. Nitrogen-containing heterocycle formation is the next most commonly employed synthetic methodology. Notably, the Suzuki coupling, a perceived favorite, does indeed appear in the top 10 but ranks lower than both *N*-Boc deprotection and carboxylic acid deprotection. Seven of the top 10 list are either acylation reactions or reactions leading to potential amide-forming precursors, accounting for 30% of all reactions. The reactions in the top 10 list themselves account for almost $^2/_3$ of all reactions reported. Taken together, we see that while medicinal chemists use a broad range of synthetic methodologies, the suggestion that there are a small number of "favorites" seems to hold some truth.

Interestingly, of these "top 10" favorites, 5 were highlighted in our previous publication (amide formation, N-arylation with Ar−X, reductive amination, Suzuki reaction, and O-substitution), the selection for which was based on an informal "straw poll".[1] It is likely that the omission of nitrogen-containing heterocycle formation reflects the somewhat broad brush that this category covers, but it is interesting to speculate whether the fact that 3 of the remaining 4 reaction types were deprotection reactions reflects the mindset of the medicinal chemist, in which such processes are almost overlooked when considering synthetic complexity. Removing protecting group manipulations (and ignoring the "other" categories within each category, as these represent a combination of many differing processes) adds sulfonamide formation (163 reactions, 2.2% of total), Sonogashira reaction (155 reactions, 2.1%), urea formation (155 reactions, 2.1%), and nitro to amine reduction (78 reactions, 1.1%) to the top 10 list. Our straw poll correctly identified 7 (sulfonylation and Sonogashira reactions, in addition to those listed above) of this revised top 10 list. However, the Heck reaction is among the 10 least common reactions in our survey (only 3 reactions, accounting for 0.04% of the total). The remaining 2 reactions in our straw poll (Grignard and Wittig reactions) have been discussed above.

In general, the trends are similar to those observed within the field of process chemistry,[2] with a few variations noted in the text, most of which can be attributed to the differing goals of medicinal and process chemistries.

4. COMPOUND ANALYSIS

We have noted throughout the manuscript that some reactions may be over-represented because of their tendency to occur later in the synthesis. To address the question of whether this use of certain well-developed reactions that are readily amenable to parallel synthesis (e.g., amide formations) is biasing the reaction count in favor of these processes in a way not seen for the analysis of process development reactions,[2] where only a single compound is normally being made, we also analyzed all the compounds within the published data set by compound structure. In this case, we counted various functional groups that were likely to be formed by the synthetic reactions described. While this also has some limitations (for example, substructures being formed as reactive functional handles that are subsequently further reacted not being included), we feel that it is perhaps more representative and, taken with the above reaction-based analysis, provides a more complete picture.

4.1. General Points. Table 5 shows the headline data for the occurrence rates of a range of functional groups within those compounds for which biological data were reported (3566 compounds). The data shown are the total number of occurrences of the functional group or substructure within the data set, the mean number of occurrences per compound (across *all* compounds, not just those in which it is found, i.e., the total number of occurrences divided by 3566 compounds), the number and % of compounds containing at least one occurrence of the functional group, and the maximum number of occurrences of that functional group within a single compound within the data set. Perhaps the most obvious point from Table 5 is the universality of aromatic rings within medicinal chemistry; 99% of compounds contain at least one aromatic ring of some sort, 94.3% contain at least one benzene ring, and 72.5% contain at lease one heteroaromatic ring, with an average over the entire data set of almost three aromatic rings per compound. This figure compares well with Ritchie and Macdonald's suggestion of a maximum of three aromatic rings per compound.[18]

Additionally, amides occur in just over 50% of compounds, and both aliphatic amines and biaryl systems occur in ~40% of compounds, while 35% of compounds contain an alkoxyaryl ether group. Thioethers are the least prevalent heteroatom linkage, as expected because of their oxidation potential. Within the halides, 14% contain an alkyl fluoride, while the other alkyl halides are, as is to be expected because of their reactivity, almost completely absent. Aryl fluorides (~20%) and chlorides (~30%) are common, with the heavier and more lipophilic bromo- and iodoarenes being less common. More detailed discussion of each subgroup is presented below.

4.2. Amines. With 42.9% of compounds containing an aliphatic amine, this group is one of the most popular

Table 5. Summary of Functional Group Occurrences across the Data Set of Compounds with Biological Data (3566 Compounds)

subclass	total no. of occurrences of FG in data set	mean no./compd across entire data set	no. of compds with ≥1 occurrence of FG	% of compds with ≥1 occurrence of FG	max. no. of FG in single compd
		Class: Amine			
aliphatic	1719	0.48	1529	42.9	4
aryl	896	0.25	810	22.7	3
diaryl	561	0.16	503	14.1	2
		Class: (Thio)ether			
ROR	609	0.17	530	14.9	6
ROAr	1568	0.44	1231	34.5	5
ArOAr	283	0.08	280	7.9	2
C-S-C	234	0.07	228	6.4	2
		Class: Acyl, Sulfonyl, Etc.			
RCO_2H	242	0.07	234	6.6	2
$ArCO_2H$	134	0.04	133	3.7	2
amide	2376	0.67	1942	54.5	5
sulfonamide	383	0.11	374	10.5	2
ester	108	0.03	101	2.8	4
urea	325	0.09	323	9.1	2
carbamate	68	0.02	68	1.9	1
carbonate	1	0.00	1	0.0	1
amidine	37	0.01	37	1.0	1
		Class: sp^2-sp^2			
biaryl	1682	0.47	1375	38.6	3
arylalkene	116	0.03	87	2.4	2
arylalkyne	253	0.07	152	4.3	2
		Class: Aromatic Ring			
all Ar	10633	2.98	3532	99.0	6
all heteroaromatic	4362	1.22	2586	72.5	4
all 5-mr[a]	2509	0.70	2067	58.0	4
all 6-mr[a]	8124	2.28	3515	98.6	5
benzenoid	6270	1.76	3363	94.3	4
N-containing (5-mr[a])	2171	0.61	1813	50.8	4
N-Containing (6-mr[a])	1853	0.52	1550	43.5	3
O-containing	406	0.11	371	10.4	2
S-containing	480	0.13	473	13.3	2
		Class: Alcohol			
ROH	486	0.14	371	10.4	4
ArOH	232	0.07	179	5.0	3
R/ArSH	39	0.01	39	1.1	1
		Class: Alkyl Halide			
R-F	1588	0.45	496	13.9	9
R-Cl	13	0.00	13	0.4	1
R-Br	0	0.00	0	0.0	0
R-I	0	0.00	0	0.0	0
		Class: Aryl Halide			
Ar-F	912	0.26	685	19.2	4
Ar-Cl	1325	0.37	1071	30.0	3
Ar-Br	93	0.03	92	2.6	2
Ar-I	62	0.02	62	1.7	1
		Class: Miscellaneous			
sulfoxide	4	0.00	4	0.1	1

dx.doi.org/10.1021/jm200187y |J. Med. Chem. 2011, 54, 3451–3479

Table 5. Continued

subclass	total no. of occurrences of FG in data set	mean no./compd across entire data set	no. of compds with ≥1 occurrence of FG	% of compds with ≥1 occurrence of FG	max. no. of FG in single compd
		Class: Miscellaneous (Continued)			
sulfone	88	0.02	86	2.4	2
$ArNO_2$	15	0.00	15	0.4	1

[a] 5-mr and 6-mr refer to five-membered and six-membered rings respectively.

Table 6. Detailed Analysis of Aliphatic Amines in the Data Set

amine type	max.	no.	% of all aliphatic amines	% of subclass
primary amine	2	84	4.9	
secondary amine	2	605	35.2	
secondary NMe	1	120	7	19.8
secondary piperidine	1	94	5.5	15.5
secondary piperidin-4-yl	1	38	2.2	6.3
secondary piperidin-3-yl	1	35	2	5.8
secondary piperidin-2-yl	1	3	0.2	0.5
secondary piperazine[a]	2	89	5.2	14.7
secondary pyrrolidine	1	78	4.5	12.9
secondary pyrrolidin-3-yl	1	60	3.5	9.9
secondary pyrrolidin-2-yl	1	5	0.3	0.8
secondary morpholine[b]	1	59	3.4	9.8
secondary azetidine[c]	1	12	0.7	2.0
secondary azetidin-3-yl	1	6	0.3	1.0
secondary homopiperazine	1	1	0.1	0.2
tertiary amine	2	1030	59.9	
tertiary pyrrolidine	2	235	13.7	22.8
tertiary pyrrolidin-1-yl	1	115	6.7	11.2
tertiary pyrrolidine-1,3-diyl	2	73	4.2	7.1
tertiary pyrrolidine-1,2-diyl	1	7	0.4	0.7
tertiary piperazine	2	196	11.4	19.0
tertiary piperazin-*N*,*N*′-diyl	2	134	7.8	13
tertiary piperidine	3	181	10.5	17.6
tertiary piperidine-1-yl	1	17	1	1.7
tertiary piperidine-1,4-diyl	1	105	6.1	10.2
tertiary piperidine-1,3-diyl	1	3	0.2	0.3
tertiary piperidine-1,2-diyl	1	1	0.1	0.1
tertiary dimethyl	1	169	9.8	16.4
tertiary morpholine[d]	1	72	4.2	7
tertiary diethyl	1	13	0.8	1.3
tertiary azetidine	1	12	0.7	1.2
tertiary azetidin-1-yl	1	4	0.2	0.4
tertiary azetidin-1,3-diyl	1	6	0.3	0.6
tertiary homopiperazine[e]	1	11	0.6	1.1
β-fluoro	4	14	0.8	
γ-fluoro	1	1	0.1	
β-O	3	494	28.7	
γ-O	3	206	12	
β-N	4	884	51.4	
γ-N	2	393	22.9	

[a] All piperazines were linked via N′. [b] All morpholines linked via 2-position. [c] No occurrences of secondary azetidin-2-yl in the data set. [d] All except one example are N-monosubstituted. [e] All N,N′-disubstituted.

functionalities outside the aromatic rings. In many cases, an amino group has been added to the molecule in order to enhance solubility and decrease lipophilicity while attempting to avoid detrimental effects on potency. Visual inspection suggests that these are often introduced in benzylic positions, most likely by reductive amination of a benzaldehyde precursor or via ether

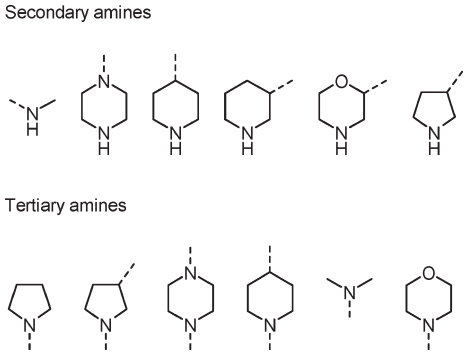


Figure 1. Most common aliphatic amines and favored connectivities (in descending order of occurrence from left to right). Dotted bonds indicate point(s) of attachment.

Table 7. Breakdown of Alkoxyaryl Subtypes

ArOR type	max	no.	% of ArOR (% of parent type)
simple alkyl ethers	4	606	38.6
ArOMe	4	421	26.8 (69.5)
solubilizing	3	314	20.0
β-nitrogen	2	88	5.6 (28.0)
methylenedioxy	1	88	11.2[a]
ethylenedioxy	1	11	1.4[a]
$ArOCF_3$	1	92	5.9
$ArOCH_2CF_3$	1	5	0.3

[a] These percentages are doubled (i.e., 88/1568 = 5.6%), as each alkylidenedioxy group will have counted as 2 ArOR groups in the 1568 total.

linkages to aromatic rings, by O-alkylation of a suitable phenolic precursor. Table 6 shows a detailed breakdown of the amines encountered within the data set. Primary amines are rare; only 5% of the amines fall into this category. Of the remaining 95%, a little over one-third are secondary amines, and the majority (60% of amines) are tertiary. Further examination of the secondary amines shows 20% of them to be terminal NH-methyl groups, with approximately equal numbers of piperazines and piperidines (~15% each). The piperazines are all attached to the rest of the molecule via the reactive N′-position, whereas the piperidines are split almost equally between the 3-postion and 4-position, with only a few examples of 2-substitution. Pyrrolidines are marginally less common (13%), with the 3-position linkage dominating. Morpholines account for ~10% of secondary amines, and all were linked via the 2-position. Homopiperazines and azetidines were rare. It is striking that ~75% of secondary amines are accounted for by only five categories (methyl, piperidine, piperazine, pyrrolidine, and morpholine).

Examination of the tertiary amines shows a similar pattern. Pyrrolidines are the most common, accounting for almost one-quarter of the subclass (23%). Half of these are monosubstituted on the 1-position, with most of the remainder again preferring the 3-position as noted for secondary amines above. Piperazines account for a further 19%, $^2/_3$ of which are substituted only on the more reactive nitrogen atoms. Piperidines occur in similar numbers, but in contrast to the pyrrolidines, only ~10% of piperidines are monosubstituted on the 1-position. The majority (~60%) are disubstituted at the 1,4-positions (most likely because of the symmetry allowing for synthetic accessibility and avoiding issues of chirality), with only 1−2% disubstitued at the 1,3-position or 1,2-position. Dimethylamines represent a further 16% of tertiary amines (in sharp contrast with diethylamines, which are more than 10-fold less common), while 7% are morpholines (almost exclusively monosubstituted on the N position). As observed for secondary amines, azetidines (one-third N-monosubstituted and one-half 1,3-disubstituted) and homopiperazines (exclusively N,N′-disubstituted) are both poorly represented within the data set. Again, ~80% of tertiary amines are accounted for by the same five amine types (pyrrolidine, piperazine, piperidine, dimethyl, and morpholine) as for the secondary amines. Figure 1 summarizes the most common secondary and tertiary amines and their preferred substitution positions. Ritchie et al.[81] have recently investigated the effect of aromatic and aliphatic ring types on compound developability and concluded that heteroaliphatic rings are likely to be beneficial, although they caution that they may also introduce hERG-induced toxicity, which correlates well with the trend seen in this data set for their incorporation.

Half of the amines have a β-nitrogen atom (including the ~15% of amines that are piperazines) and one-quarter a γ-nitrogen atom. Additionally, a little over one-fourth of amines have a β-oxygen atom (including the ~7.5% that are morpholines) and around one in eight a γ-oxygen atom. This almost certainly reflects the utility of amino, amide, and ether linkages in the introduction of the amino groups and possibly also their role as solubilizing groups. Fluorine atoms in the β-position and γ-position are much rarer, despite their modifying properties (see section on halogens below).

In addition to the aliphatic amines, $^1/_3$ of compounds contain at least one aryl- or diarylamino group. Visual inspection suggests that monoarylamines are most often encountered in the form of 2-aminopyridine and 2- or 4-aminopyrimidines (and their fused derivatives), where potential in vivo toxicity is less of a concern than for aniline-derived systems following dealkylation. In the case of diarylamines, the second aromatic ring is a phenyl ring in many cases, as the potential for in vivo metabolism to the free aniline is much reduced.[82−84] Both groups provide, in many cases, a hydrogen bond donor/acceptor pair, forming key interactions with the intended biological target, particularly in the field of kinase inhibitors.[85,86]

4.3. Ethers/Thioethers. Ethers are a commonly represented functionality within the data set; 2694 ethers and thioethers are to be found among the 3566 compounds, an average of 0.76 ethers/compound, with some compounds containing as many as 6 examples of a single subtype within this class (Table 5). While alkoxy aryl ethers (ROAr) account for more than half of this total, all types are well represented, although as mentioned above, thioethers are relatively rare. The ether groups are readily synthesized, can act as H-bond acceptors, and decrease lipophilicity, thereby potentially enhancing aqueous solubility and reducing metabolic liabilities. In the following sections, we analyze these subtypes in more detail.

4.3.1. Alkoxy Aryl Ethers (ROAr). Alkoxy aryl ethers fall broadly into three categories: simple alkyl (Me, Et, ^{i}Pr, ^{n}Pr, ^{c}Pr, etc., up to and including five-carbon atoms and no heteroatoms), "solubilizing" (in which the alkyl portion contains a basic nitrogen atom or $RO(CH_2)_2-$ linkage), and other more complex systems, which we refer to as being part of the scaffold of the molecule. Examination of the compounds within the data set shows that almost 40% of the ethers fall into the simple alkyl category (Table 7), with MeOAr accounting for ~70% of this total (27% of all alkoxyaryls). Additionally, methylenedioxy and ethylenedioxy

Table 8. Dialkyl Ether Subtypes

ROR type	max	no.	% of ROR (% of parent type)
simple RO-alkyl ethers	3	222	36.5
ROMe	2	101	16.6 (45.5)
RO-solubilizing	3	115[a]	0.2
any morpholine	2	191	31.4
mono-N-substituted morpholine	2	129	21.2 (67.5)
furans	1	22	3.6
pyrans	1	48	7.9
4-aminopyrans	1	22	3.6 (45.8)

[a] Only one of these is actually a solubilizing substituent, the remainder being part of the scaffold. See text for details.

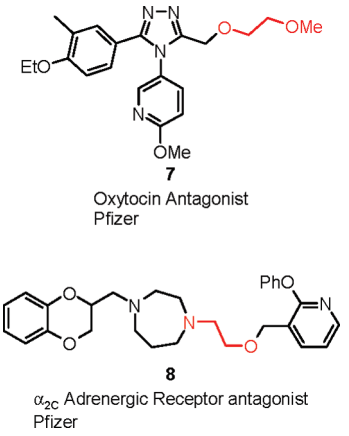


Figure 2. β-Heteroatom-containing dialkyl ethers: (top) only example where used as a point of attachment for a potentially solubilizing group (itself also a β-heteroatom-containing dialkyl ether);[87] (bottom) β-amino dialkyl ether based scaffold from a series of Pfizer α_{2C} adrenergic receptor antagonists.[88]

(both of which count as two alkoxy aryl ethers in the headline totals) account for ~12% of alkoxyaryls, as do trifluoromethyl ethers (F_3COAr). Approximately 20% are in the "solubilizing" category, with 28% of these containing a β-aminoethyl group. The remainder (~25%) are more complex ethers that appear to form part of the core scaffold of the molecule.

From these figures, it is clear that the increased liability of the aromatic ring toward oxidative metabolism on incorporation of electron-donating substituents is offset by the requirement for such groups, which presumably primarily serve as H-bond acceptors, to obtain the required levels of potency. Additionally, the alkoxy aryl ether motif is clearly a useful synthetic handle for the assembly of scaffolds and the attachment of pendent functionality (particularly solubilizing groups) to aromatic templates. This is almost certainly a result of the reliable and clean reactivity of the phenol precursor under both classical (base, alkyl halide, or sulfonate) or Mitsunobu[41,42] alkylation conditions.

4.3.2. Dialkyl Ethers (ROR′). Of the dialkyl ethers, the second largest group within the (thio)ether subgroup, 36% are simple alkyl ethers as defined above (Table 8). Almost half (45%) of these are methyl ethers. In contrast to the above, while the data set contains numerous β- or γ-heteroatom-containing ethers, only one example (7)[87] of these appears to actually function as a point of attachment for a possible solubilizing group. The

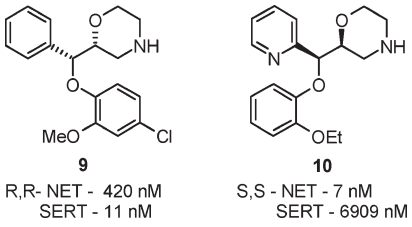


Figure 3. Examples of the use of morpholine as part of the molecular scaffold from Pfizer dual serotonin (SERT)/noradrenaline (NET) (left)[89] and selective noradrenaline (right)[90] reuptake inhibitor programs.

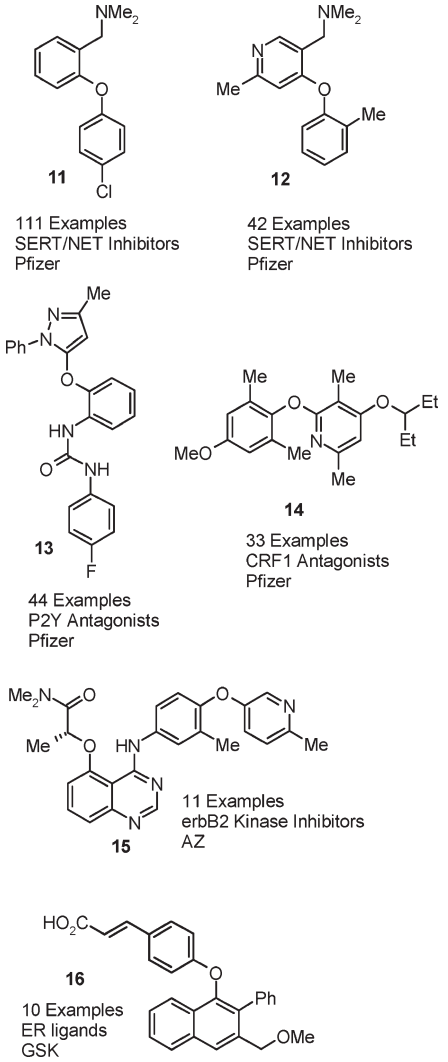


Figure 4. Selected examples of diaryl ether-containing scaffolds contained within the data set.[91−99]

remainder are in general actually part of the solubilizing group (as also in 7, Figure 2)[87] or form part of the core scaffold (e.g., 8, Figure 2).[88] It is worth noting that 8 also shows examples of a diaryl ether and an ethylenedioxy motif in addition to the dialkyl ether.

Tetrahydropyrans and tetrahydrofurans account for a further 12% of the dialkyl ethers (present as both sugar-derived units and simpler systems), most notably the 4-aminopyran solubilizing group, which accounts for almost half of the tetrahydropyrans. While morpholines

Table 9. Breakdown of Thioether Subtypes

thioether type	max	no.	% of thioethers (% of parent type)
simple ArS-alkyl ethers	1	46	19.7
ArSMe	1	20	8.5 (43.5)
simple RS-alkyl ethers	1	12	5.1
RSMe	1	12	5.1 (100)
ArS-solubilizing[a]	1	98	41.9
ArS-β-nitrogen	1	7	3.0 (7.1)
RS-solubilizing	2	3	1.3
RS-β-nitrogen	2	3	1.3 (100)
$ArSCF_3$	1	1	0.4

[a] Examination shows that these are all actually scaffolds, not pendent solubilizers; see text for details.

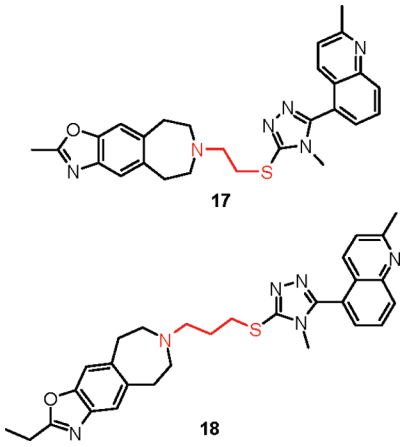


Figure 5. Examples of β- and γ-nitrogen substituted thioethers, in which the thioether groups form part of the scaffold, from a series of D3 receptor antagonists (GSK).[100,101]

are predominantly present as unsubstituted solubilizing groups (representing 21% of all dialkyl ethers in the set), a small number bear methyl modifications while in all likelihood retaining this purpose. However, there are also two reports where they form part of the scaffold (Figure 3), in this case, the chirality affecting the selectivity between serotonin and noradrenaline transporters.[89,90]

4.3.3. Diaryl Ethers (ArOAr′). The 230 of the 283 diaryl ethers found within the data set come from just seven publications, with the remainder consisting of scattered examples of one or two compounds within a series. Examination of the structures of these 230 compounds supports the view that this motif is generally encountered as part of a core scaffold, linking two aromatic rings. Closer examination shows that 153 occurrences all relate to a series of reports from Pfizer of a set of closely related compounds exemplified by **11** and **12** (Figure 4), which were explored for serotonin and noradrenaline reuptake inhibition.[91−95] Further examples include a pyrazolyloxyphenyl ether in a series of P2Y receptor antagonists (e.g., compound **13**),[96] phenoxypyrid-2-yl CRF1 antagonists (e.g., **14**),[97] phenoxypyrid-3-yl erbB2 receptor tyrosine kinase inhibitors (e.g., **15**),[98] and phenoxynaphthyl ER ligands (e.g., **16**),[99] demonstrating a wide array of variations in aromatic motifs and target classes.

4.3.4. Thioethers. Of the thioethers, the least common subtype within the (thio)ether category, ∼25% are simple alkyl ethers, and half of these are methyl thioethers (MeSR or MeSAr), with all the dialkyl thioethers (RSR′) in this group being methyl

Table 10. Analysis of Amide Subtypes

amide type	max	no.	% of total	% of subclass
C-alkylamides	5	1029	43.3	
acetamides	1	40	1.7	3.9
C-solubilized	5	669	28.2	65.0
C-arylamides	3	1347	56.7	
primary $-NH_2$	2	77	3.2	
secondary	3	1454	61.2	
alkyl -NHR	3	972	40.9	66.9
methyl -NHMe	1	65	2.7	4.5
solubilized -NHSol	2	322	13.6	22.1
aryl -NHAr	3	482	20.3	33.1
tertiary	2	779	32.8	
dialkyl $-NR_2$	2	636	26.8	81.6
alkyl-solubilized -NRSol	2	330	13.9	42.4
methylalkyl -N(Me)R	1	103	4.3	13.2
arylalkyl -NRAr	2	143	6.0	18.4
methylaryl -N(Me)Ar	1	9	0.4	1.2
aryl-solubilized -N(Ar)Sol	1	64	2.7	8.2
N-acylamides	2	56	2.4	
imide	1	24	1.0	42.9
N-acylsulfonamide	1	1	0.0	1.7
N-acylurea	2	31	1.3	55.4
lactam	2	304	12.8	
tri- or greater peptide	2	38	1.6	

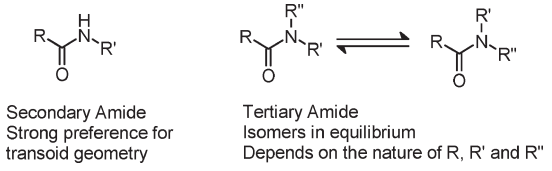


Figure 6. Geometrical preferences for amide bonds.

thioethers (Table 9). The thioether subtype appears at first sight to be dominated by aryl-thioether-linked solubilizing groups. However closer examination of this subset reveals that all of these compounds are scaffold-based aryl thioethers containing a β- or γ-nitrogen within the scaffold (e.g., **17** and **18**; Figure 5).[100,101] In contrast to the case with ethers, aryl thioethers do not occur in the data set as a means of linking solubilizing groups. Again, in contrast with the alkoxy aryl ethers, the CF_3 group (F_3CSAr) occurs only once within the data set.

4.4. Amides, Sulfonamides, and Other Acylation-Type Products. Within this group, in close accord with the trend seen within the reactions analysis, the amide group dominates, being present in over half (54%) of all compounds in the data set (Table 5). Only a small proportion (∼3%, Table 10) of the amides are primary amides. The remainder are split 2:1 secondary/tertiary, suggesting that the presence of an H-bond donor might be significant for their incorporation. Additionally, secondary amides show a clear conformational preference for the transoid amide geometry, whereas there is often little preference for a particular geometric isomer for analogous tertiary amides (Figure 6).

Analysis of the acid and amine parents shows that the acid parents are derived from aliphatic and aromatic acids in almost

dx.doi.org/10.1021/jm200187y |J. Med. Chem. 2011, 54, 3451–3479

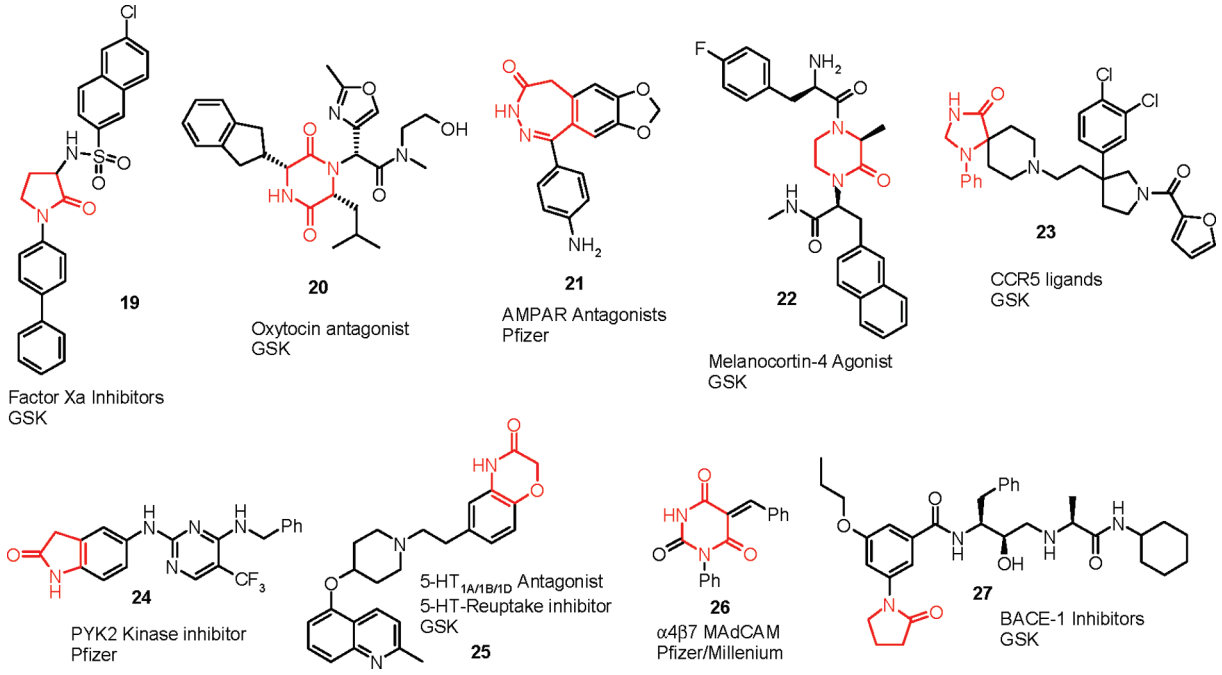


Figure 7. Selected lactam-containing compound series (>10 compounds with motif), showing broad diversity of lactam-containing scaffolds and targets.[102−113]

equal portions (Table 10). Only 4% of the aliphatic-derived amides are acetamides, while 65% can be considered to be "solubilizing" (which we define here as containing an oxygen or nitrogen attached to the α-, β-, or γ-carbon of the aliphatic chain). Many of these are actually substituents, in which it may be the amide that is the actual uncharged solubilizing group (in contrast to amine solubilizing groups) and the chain heteroatom is actually part of a synthetic handle for their attachment.

Unlike the case with the amines described above, no strong preferences for particular amine derivatives were found in the data set. However, almost exactly two-thirds of the secondary amides were derived from aliphatic amines (but only ∼5% from methylamine, the simplest analogue). "Solubilizing" groups (containing a β- or γ-oxygen or nitrogen atom) account for one in five of the secondary amides. Tertiary amides show an even stronger preference for alkyl substituents (dialkylamides accounted for over 80% of the tertiary amides), while there are no diarylamides. This preference is most likely due to a desire to avoid the potential for liberation of potentially toxic anilines by in vivo cleavage of the amide bond by hydrolase enzymes. Around 50% of the tertiary amines bear a solubilizing group. The increased proportion of tertiary amides bearing a solubilizing group relative to secondary amides suggests that the addition of a solubilizing group to the amide nitrogen is a common strategy in situations where the N−H motif has been shown to be unnecessary for biological activity (perhaps by initial synthesis of the *N*-methyl analogue of the parent secondary amine, as suggested by the increased proportion of tertiary amines bearing a methyl substituent).

A surprisingly large proportion of the amides are also cyclic (lactams, ∼12%), arising predominantly from a small number of compound series in which they are present as part of the core scaffold, perhaps artificially elevating their standing within the data set. Even within this set, however, there is considerable diversity within the lactam functionality represented (see Figure 7),[102−107] and two of these scaffolds each contain two lactam substructures (**20** and **26**).[108,109] This motif is also heavily represented by a series of BACE-1 inhibitors bearing an N-linked pyrrolidin-2-one as an aromatic substituent (e.g., **27**)[110,111] and by a series of benzo-fused morpholinones (e.g., **25**) showing combined 5-HT$_{1A/1B/1D}$ antagonism and serotonin (5-HT) reuptake inhibition.[112,113] Many of these scaffolds can also be seen to contain additional amide groups. The remainder of the lactams are present as isolated examples throughout the data set.

In addition to these "simple" amides, a small proportion belongs to more complex functionalities [*N*-acylamides (imides), *N*-acylsulfonamides (only one example of this relatively acidic group in the data set), and *N*-acylureas] which provide a complex array of H-bond donor and acceptor vectors.

Sulfonamides and ureas are both present in ∼10% of compounds, which is in reasonable agreement with the reaction-based figures and reflects their comparison with amides for SAR purposes. The other functionalities within this group are only represented in <5% of compounds, which is unsurprising in view of their chemical reactivity, biological lability, or permeability issues (see discussion in section 3.3.2).

4.5. sp^2−sp^2 Linkages. Biaryl linkages are present in almost 40% of compounds. Again, this reflects the utility of this motif in providing appropriate vectors within the core scaffold and also the prevalence of the Suzuki[40,48−50] and related Pd-mediated cross-coupling reactions[40,50] (40% of C−C bond forming reactions were also Suzuki reactions). Perhaps surprisingly, the products of the Heck[75] and Sonogashira[53−55] reactions (arylalkenes and acetylenes, respectively) are present in only low numbers. While the Heck reaction was rare, the Sonogashira reaction was more common (18.4% of C−C bond-forming reactions), suggesting that the products of this reaction were largely used in further synthetic manipulations

4.6. Aromatic Ring Systems. Aromatic rings, and in particular aromatic heterocycles, are almost ubiquitous in medicinal chemistry, as supported by the finding in this study (Table 5) that 99% of compounds contain at least one aromatic ring. This is at least

Table 11. Occurrences of Individual Heteroaromatic Ring Types within the Data Set

aromatic ring type	no. of heteroatoms	total occurences[a]	% of heteroaromatics	% of subclass	nonfused occurences[c]	% of occurrences nonfused[d]
benzenoid	0	6270			5129	81.8
all 6-mr[b] heteroaromatics		1853	42.5			
pyridine	1	1001	22.9	54.0	636	63.5
pyrimidine	2	627	14.4	33.8	320	51.0
pyrazine	2	161	3.7	8.7	33	20.5
pyridazine	2	38	0.9	2.1	0	0.0
1,2,4-triazine	3	23	0.5	1.2	0	0.0
1,3,5-triazine	3	2	<0.1	0.1	2	100.0
1,2,3-triazine	3	1	<0.1	0.1	0	0.0
all 5-mr[b] aromatics		2509	57.5			
pyrazole	2	616	14.1	24.6	379	61.5
imidazole	2	439	10.1	17.5	178	40.5
pyrrole	1	405	9.3	16.1	27	6.7
thiophene	1	239	5.5	9.5	160	66.9
thiazole	2	215	4.9	8.6	166	77.2
oxazole	2	165	3.8	6.6	64	38.8
1,2,4-triazole	3	158	3.6	6.3	157	99.4
furan	1	99	2.3	3.9	73	73.7
1,3,4-oxadiazole	3	53	1.2	2.1	53	100.0
isoxazole	2	31	0.7	1.2	15	48.4
1,2,4-oxadiazole	3	30	0.7	1.2	30	100.0
1,2,5-oxadiazole	3	28	0.6	1.1	28	100.0
isothiazole	2	25	0.6	1.0	0	0.0
1,2,3-triazole	3	4	0.1	0.2	2	50.0
tetrazole	4	3	0.1	0.1	3	100.0
1,3,4-thiadiazole	3	1	<0.1	<0.1	1	100.0

[a] Note that this includes all unfused and fused occurrences of each ring type. [b] 5-mr and 6-mr refer to five-membered and six-membered rings, respectively. [c] Number of occurrences of the ring system in which it does not form part of a fused aromatic ring system. [d] % of all occurrences of the individual ring system that are not part of a fused aromatic ring system.

in part explained by their ability to provide readily functionalized scaffolds with well-defined vectors for further derivation. The number of heterocycles present (4362 heteroaromatic rings, with 72.5% of compounds containing at least one) far outstrips the number of heterocycle-forming reactions (601 reactions, Table 2). This implies that heterocyclic cores are synthesized prior to derivatization into a number of analogues, purchased, or are previously available within corporate compound collections and then derivatized, in accord with their role as core scaffolds.

We analyzed in more detail the ring systems found within the data set, starting with the individual ring systems (Table 11). Only 23 of the 35 possible heteroaromatic rings (we have only considered N-containing rings in the six-membered series and a maximum of four heteroatoms in any ring) are reported in the data set. In agreement with Murcko's analysis,[26] pyridine is the most common heteroaromatic ring overall, accounting for almost 25% of all heteraromatics. Pyrimidine and pyrazole are the next most common (both ∼15%). In the six-membered series, triazines are less well represented and tetrazines are not represented at all. In the five-membered series, pyrazole is the most common (∼25% of five-membered heteroaromatics). Following this, pyrrole and imidazole are observed at similar levels (both ∼10% of all heteroaromatics). The most common (>250 occurrences) aromatic heterocycles are shown in Figure 8. Thiophenes, furans, oxazoles, thiazoles, 1,3,4-oxadiazoles, and 1,2,4-triazoles are less common, although still well represented. The remaining systems found are only seen in small numbers (<50) of compounds. In general, the occurrence rate falls off as the number of ring heteroatoms increases.

Considering only those examples where the (hetero)aromatic ring does not form part of a fused ring system shows some marked changes (Table 11). Pyridines, pyrimidines, and pyrazoles again dominate this list, with over half of the examples of each of these rings not forming part of a larger fused system. Notably, pyrrole has moved from being one of the most common aromatic rings, considering all occurrences, to having only a small number of isolated examples (27 examples, representing ∼7% of pyrroles). It is likely that this is due to a combination of the reactivity of the pyrrole ring system precluding its incorporation in many circumstances where suitable stabilizing substitution is not tolerated by the receptor, and the preponderance of indoles (see below), which accounts for around half of the fused pyrroles. A number of ring systems, such as the oxadiazole isomers, 1,2,3-triazoles, and tetrazoles, occur exclusively or near exclusively in nonfused settings; in this case the position of heteroatoms precludes the formation of simple stable fused ring systems containing these rings. Benzenoid rings, despite being present in many of the most commonly occurring fused ring systems (see below), are also highly prevalent as nonfused systems. Figure 8 shows the most common (>150 occurrences) nonfused aromatic rings.

While the above analysis starts to show some interesting trends, it is perhaps more informative to consider ring systems wherein two

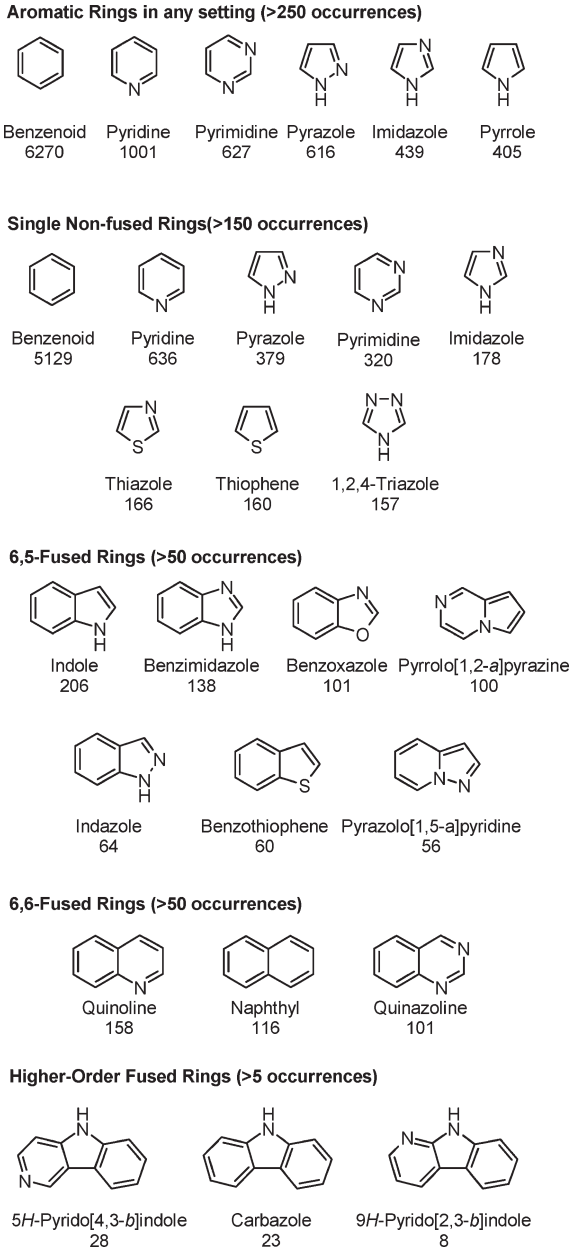


Figure 8. Most common aromatic rings and number of occurrences within the data set.

aromatic rings have been fused together, effectively "spreading" the above data onto a second axis; i.e., pyrimidine may be present as unfused pyrimidine or fused with benzene (quinazolines), imidazole (purines), pyrazine (pteridines), and so on. A recent survey[27] showed that much of the possible heteroaromatic chemical space remains unexplored within the synthetic and medicinal chemistry literature. In view of the already small set of heteroaromatic rings represented in this data set, the possibility of homo- (e.g., 2 × benzene to give naphthyl) and hetero- (e.g., benzene + pyrimidine to give quinazoline) fusions, and the additional complications of multiple possibilities for different fusion isomers within a single pair of heterocycles, it would indeed be surprising if the possible space was well covered within this data set.[114]

Strikingly, 6,5-fused ring systems dominate the data set, while 5,5-fused systems are extremely rare (Table 12). Perhaps more

Table 12. Fused Aromatic Ring Systems Found within the Data Set

fused aromatic ring types	no.	% of total	% of subclass
5,5-fused total	16	1.0	
imidazo[2,1-*b*]thiazole	16	1.0	100.0
6,5-fused total	1143	71.9	
indole	206	13.0	18.0
benzimidazole	138	8.7	12.1
benzoxazole	101	6.4	8.8
pyrrolo[1,2-*a*]pyrazine	100	6.3	8.7
indazole	64	4.0	5.6
benzothiophene	60	3.8	5.2
pyrazolo[1,5-*a*]pyridine	56	3.5	4.9
1*H*-pyrazolo[3,4-*d*]pyrimidine	39	2.5	3.4
pyrazolo[1,5-*b*]pyridazine	36	2.3	3.1
purine	32	2.0	2.8
5-azabenzimidazole	31	2.0	2.7
7-azaindazole	31	2.0	2.7
5-azaindole	28	1.8	2.4
benzisothiazole	25	1.6	2.2
7-azaindole	24	1.5	2.1
imidazo[5,1-*f*][1,2,4]triazine	23	1.4	2.0
7*H*-pyrrolo[2,3-*d*]pyrimidine	19	1.2	1.7
benzothiazole	17	1.1	1.5
thieno[3,2-*d*]pyrimidine	17	1.1	1.5
furo[2,3-*b*]pyridine	16	1.0	1.4
benzisoxazole	16	1.0	1.4
thiazolo[4,5-*d*]pyrimidine	15	0.9	1.3
imidazo[1,2-*a*]pyridine	13	0.8	1.1
1*H*-pyrazolo[4,3-*d*]pyrimidine	11	0.7	1.0
4-azabenzimidazole	6	0.4	0.5
furo[2,3-*d*]pyrimidine	5	0.3	0.4
benzofuran	3	0.2	0.3
furo[2,3-*c*]pyridine	2	0.1	0.2
imidazo[1,2-*b*]pyridazine	2	0.1	0.2
3*H*-[1,2,3]triazolo[4,5-*d*]pyrimidine	2	0.1	0.2
thieno[3,2-*c*]pyridine	1	0.1	0.1
pyrrolo[1,2-*c*]pyrimidine	1	0.1	0.1
[1,2,4]triazolo[2,3-*a*]pyridine	1	0.1	0.1
thieno[3,2-*b*]pyridine	1	0.1	0.1
thiazolo[5,4-*d*]pyrimidine	1	0.1	0.1
6,6-fused total	430	27.1	
quinoline	158	9.9	36.7
naphthyl	116	7.3	27.0
quinazoline	101	6.4	23.5
pteridine	21	1.3	4.9
7-azaquinazoline	12	0.8	2.8
isoquinoline	12	0.8	2.8
quinoxaline	7	0.4	1.6
8-azaquinazoline	2	0.1	0.5
benzo[*d*][1,2,3]triazine	1	0.1	0.2
no. of different rings reported	45		
unique rings reported[a]	28		

[a] Number of unique rings reported refers to the number of ring types reported by only a single company within the data set.

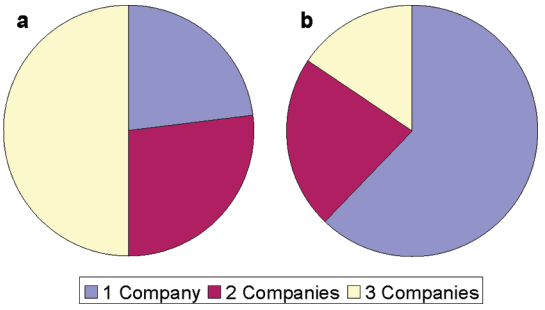


Figure 9. Number of companies reporting aromatic ring types: (a) single heteroaromatic ring types (all occurrences of each ring); (b) fused heteroaromatic ring types.

surprisingly, the 23 heteroaromatic ring types disclosed have only expanded to 45 fused types, clearly only a small portion of the available chemical space. Only 17 of the heteroaromatic rings are utilized in fused systems (alongside benzene). Benzene-fused systems account for 1025 of the fused aromatic systems (from a total of 1589), accounting for 15 of the 45 fused ring systems. The next most common are pyridine and pyrimidine, which are present in 30 of the 45 fused systems (672 compounds). Pyrrole and imidazole account for similar numbers of compounds (639) over 14 fused systems. Figure 8 shows the most common (>50 occurrences) fused aromatic ring systems.

Finally, we considered higher-fused systems (i.e., three or more fused aromatic rings). Only 60 examples of such ring systems were found within the data set, with 59 of those being one of the three scaffolds shown in Figure 8. There were no occurrences of four or more fused fully aromatic ring systems.

In view of the importance of heterocyclic motifs to intellectual property,[25,27,115,116] we considered further the distribution of heterocycle types (Figure 9). While half of all single heteroaromatic rings are reported by all three companies and only ~25% by only one company in this analysis, around two-thirds of the fused ring systems are only reported by one company in the current data set, suggesting that careful choice of heteroaromatic fusion may indeed be key to intellectual property.

4.7. Alcohols. Around 10% of compounds contain an aliphatic alcohol group and 5% a phenolic system. These groups are liable to in vivo oxidation and conjugation reactions but also help increase solubility, decrease lipophilicity, and provide the possibility of both H-bond donor and acceptor interactions.

Inspection of the structures containing aliphatic alcohols groups shows that 16 compounds (comprising a total of 50 ROH groups) contain the alcohol groups as part of a sugar-derived motif. Of the remaining compounds, in many cases, it would appear that the alcohol group forms part of a group intended to increase solubility, either as part of a hydroxyamine motif or as an alternative to a charged amino-type solubilizing group. In contrast, it would appear that the majority of phenolic groups are introduced to modulate enzyme potency. In those cases where multiple phenolic groups are present, they are either in separate aromatic rings[117] or in a meta-relationship (resorcinol) in which oxidation to quinone-type species is not accessible.[118] Thiols, which are considerably more reactive than the corresponding alcohols, are only reported in a single series of compounds.[73,119]

4.8. Halogens. The significance of fluorine in medicinal chemistry has been well documented.[120−122] In particular, the exceptional strength of the C−F bond allows for its incorporation in alkyl chains without introducing undesirable alkylating properties, and its electron-withdrawing nature and small size allow for the replacement of hydrogen atoms with little steric impact but significant electronic and metabolic effects; particularly, its introduction can block oxidative metabolism (either at or adjacent to its site of introduction), modify pK_a of amino groups α-, β-, or γ- to the fluorine(s), affecting both solubility and receptor/off-target binding, and increase lipophilicity. Additionally, it is a good hydrogen-bond (H-Bond) acceptor but almost never functions as a halogen bond ("X-bond") donor.[123]

The distribution of fluorine atoms within the data set is shown in Table 13. It is interesting to note that while ~1 in 3 fluorine atoms are direct aryl substituents, 95% of the alkyl fluorines are found in CF_3 groups, and of these, 75% are present as the aryl substituents $ArCF_3$ or $ArOCF_3$. A further 20% are alkyl CF_3, with the remainder consisting of 28 TFA salts in the data set and a handful of trifluoroacetamides and trifluoromethylsulfonamides. CF_2 groups and aliphatic CF groups together account for only 5% of alkyl fluorides within the data set. Visual inspection suggests that the CF_2 group almost invariably appears in $ArCF_2CF_3$ or $ArCF_2H$ side chains or the corresponding $ArOCF_2CF_3$ groups. This decrease is likely to be at least in part due to the paucity of good generally applicable synthetic methods for their introduction.[124−126]

The heavier halogens offer greater increases in lipophilicity than fluorine while providing similar metabolism-blocking properties, offering increased membrane permeability at the expense of likely increases in off-target effects. While they are generally not considered to be good hydrogen bond acceptors,[123] their ability to take part in halogen bonds (X-bonds) increases in the order Cl < Br < I.[123,127−129] A recent review summarizes these properties in more detail.[130]

In general, the reactive nature of aliphatic chlorine, bromine, and iodine precludes their inclusion in medicinal chemistry compounds. The only exception to this rule in the current data set is a series of clindamycin analogues exemplified by **29** (which showed an antibacterial profile similar to that of clindamycin **28**),[131] containing a natural product-derived hindered secondary alkyl chloride (Figure 10). As clindamycin is clinically useful, presumably this highly hindered alkyl chloride was considered acceptable in this context.

Aromatic halogens, being in general much less reactive than their aliphatic counterparts, feature heavily in the medicinal chemistry literature.[130] Aromatic chlorides are marginally more prevalent than aromatic fluorides (Table 5, 1325 ArCl vs 912 ArF), while bromines are considerably less prevalent. This decline down the group is almost certainly due to the increasing molecular weight, and the solubility and lipophilicity burden introduced as the heavier halogens are incorporated, chlorine appearing to provide the best balance between enhanced membrane permeability and detrimental increases in MWt and log *P* and decreases in aqueous solubility.

Within the heavier aromatic halogens, there is considerable diversity within their molecular environments, and some are documented as having been introduced to modulate physicochemical properties, while others are recorded as enhancing potency relative to other substituents (commonly H− and Me−). However, some other generalizations can also be made. First, only a very small number (42) of chloro substituents are in "activated" positions, where they are liable to undergo S_NAr-type substitution reactions (all were either ortho or para to a ring nitrogen atom (or both). None were activated by $-NO_2$, $-C{\equiv}N$, $-C(=O)R$ etc. susbtituents), and in general, these few examples are expected to

Table 13. Distribution of Fluorine Atoms in the Data Set

fluorine type	max	no.[a]	% of total[b]	% of CF_n
any fluorine	9	2500		
aryl F	4	912	36.5	
alkyl F	9	1588	63.5	
all CF_3	3	506	60.7	
aryl CF_3	3	297		58.7
aryl OCF_3	1	92		18.2
alkyl CF_3 (not TFA salts, esters, amides, CF_3SO_2)	2	105		20.8
$CF_3C(=O)NR_1R_2$	1	4		0.8
$CF_3S(=O)_2NR_1R_2$	1	10		2.0
CF_3CO_2H salt	1	28		5.5
all CF_2	2	30	2.4	
all monofluoro aliphatic	1	10	0.4	

[a] "no." is the number of occurrences of that group, so 1 occurrence of a CF_3 group will account for 3 F atoms in the alkyl F total. [b] Percent of all fluorine atoms; thus, 1 CF_3 group contributes 3 of the 2500 fluorine atoms in the data set.

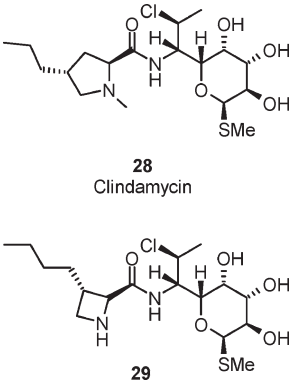


Figure 10. Pfizer clindamycin analogue **29**, showing similar antibacterial profile to clindamycin **28**, containing a hindered secondary alkyl chloride.

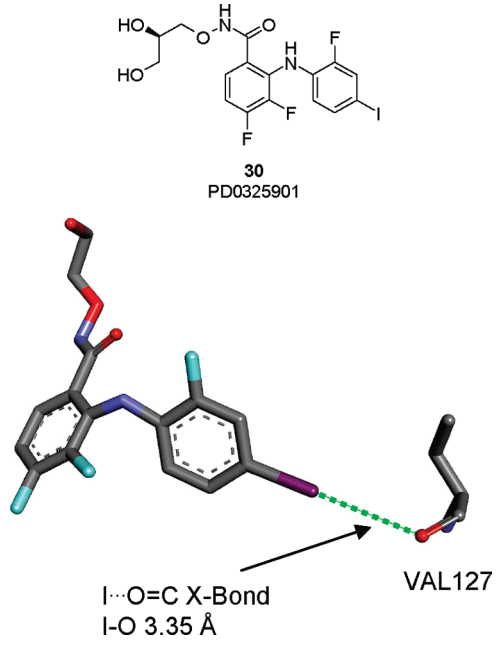


Figure 11. Pfizer iodoarene MEK1 inhibitors: (top) PD0325901 (clinical candidate); (bottom) X-ray structure of closely related compound in complex with MEK1 (PDB code 3DY7).[132,133,135]

have low reactivity (they are predominantly the product of an existing, selective S_NAr process, in which a second halogen was displaced by a nucleophile, concomitantly deactivating the remaining halogen) or potentially synthetic intermediates for which biological assay data were obtained en route to the intended final product. Additionally, F— and Br— tend to occur ortho or para to electron-donating substituents (in both cases with a slight bias toward the para relationship). For Cl— substituents, however, the trend is slightly different, with the majority of substituents being para to an electron donating substituent but meta being slightly favored over ortho. While a meta/para bias is to be expected purely from the point of blocking metabolic sites, as these are the positions activated by the electron-donating substituent, there is also the competing use of halogen substituents to modulate physicochemical properties in SAR-neutral positions and to incorporate halogens to provide specific interactions (either through specific H- or X-bonding interactions or the filling of small hydrophobic pockets) which all serve to mask this effect. The situation is further complicated by the multiple substitutions present on most aromatic rings. Further detailed analysis of the halogen environments is beyond the scope of this review.

A small number of iodoarenes (62) are present in the data set. The majority of these compounds (54) come from two reports describing a series of MEK1 kinase inhibitors, where the iodo substituent was found to be essential for activity.[132,133] Examination of an X-ray structure deposited at the PDB[134] (PDB code 3DY7[135]) suggests that this iodo substituent may form a halogen bond ("X-bond")[123,127−129] to the C═O of VAL127 of MEK1 (Figure 11).

4.9. Miscellaneous (Sulfoxide, Sulfone, and Nitroarenes). All the functional groups in this subclass are only represented in a small number of compounds (<2.5% of compounds contain them). This is in accord with their likely primary reason for introduction relating to their utility as synthetic handles for further synthetic manipulation.

5. MOLECULAR COMPLEXITY

In this section, we analyze three measures of molecular complexity. The first, a measure of synthetic complexity, is the

number of synthetic steps recorded within our published data set to each compound for which biological data were obtained. The second is the number of defined chiral centers, and the third, Fsp3, is a measure of the degree of spatial complexity recently proposed by Lovering et al.[12] The first two measures were also calculated in the review of process chemistry,[2] allowing direct comparison.

5.1. Synthetic Complexity. Analyzing the data set of 3566 compounds reveals that traceable routes are described for 2973 derivatives, where the number of discrete transformations per synthesis can be defined with a reasonable degree of clarity (see section 2 for details of how we determined the number of synthetic steps). These steps equate to 14 309 discrete chemical transformations,[136] averaging 4.8 steps per compound synthesized (Figure 12a). One-quarter of all the compounds were synthesized in three steps, with half being synthesized in three to five steps. This contrasts markedly with the average number of steps per synthesis of drug candidate, of 8.1 steps per derivative.[2] However, the analysis correlates well with differing phases of a medicinal chemistry program, where chemists produce a considerable number of simpler compounds to rapidly explore SAR before making more detailed and complex analyses of molecular architecture to optimize potency, selectivity, and physicochemical parameters. Indeed, it is possible that this survey biases the analysis toward simpler derivatives, as the literature tends to report case studies of rapid SAR expansion through a somewhat disproportionate number of these simpler derivatives, before describing a smaller number of more complex derivatives with properties more akin to a potential drug candidate. Additionally, further bias may be introduced by the reporting of routes from known in-house reagents, which may well have represented the medicinal chemistry starting material but would require synthesis for development-scale work. However, we feel that the data presented generally reflect the relative degrees of synthetic complexity of compounds prepared throughout the lifespan of a medicinal chemistry project. It is testament to the inventiveness of medicinal chemists that such molecular diversity can be accessed with such minimal synthetic effort.

5.2. Chirality. Given the stereochemically defined nature of the targets with which small molecule drug candidates are designed to interact, it has long been seen to be logical to recapitulate this geometry within the derivatives of interest prepared for screening. However, there is a pervasive feeling that such information is not actively captured by the derivatives incorporated into screening collections or designed as part of medicinal chemistry programs.

Despite many preconceptions that medicinal chemists prefer to prepare flat achiral derivatives, these assumptions do not appear to be supported by this analysis. Indeed, of the 3566 compounds described as having been assessed in a biological assay system, around one-third (1093 compounds) possessed at least one chiral center with defined stereochemical integrity (Figure 12b). It is stressed that these numbers do not include compounds prepared as racemates.[137] Furthermore, a considerable number of compounds possessed multiple chiral centers, with almost as many compounds containing two defined chiral centers as containing one. The average number of chiral centers, averaged across all the molecules evaluated, was one defined stereocenter per molecule, indicating a trend toward chemists being equally likely to prepare a chiral derivative as an achiral one.

Analysis of whether the chiral center forms a key part of the core of the compound series or is part of a variable peripheral

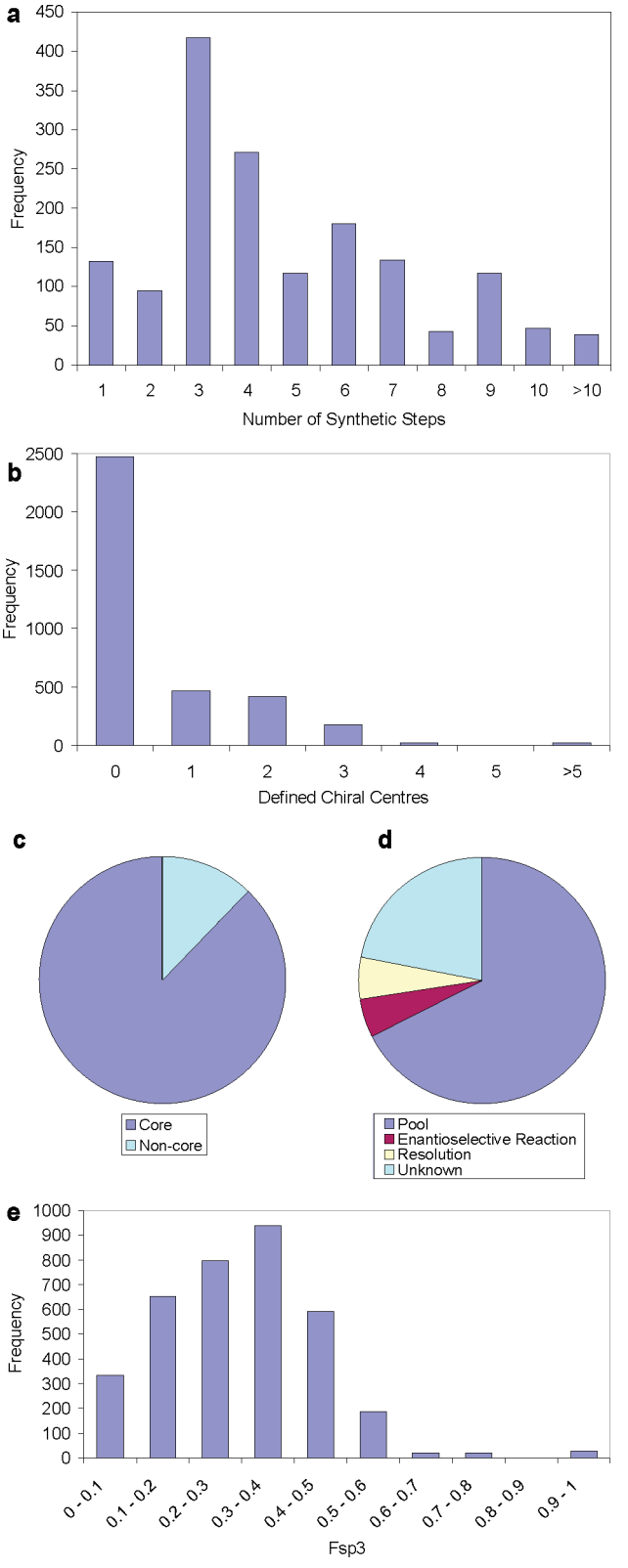


Figure 12. Distributions of parameters of molecular complexity: (a) number of synthetic steps; (b) number of defined chiral centers; (c) distribution of chiral centers, those forming part of the retained "core" of the compounds vs noncore substituents; (d) distribution of the sources of known enantiopure chiral centers; (e) Fsp3,[12] a measure of the degree of saturation. Frequencies are numbers of compounds.

dx.doi.org/10.1021/jm200187y |J. Med. Chem. 2011, 54, 3451–3479

substituent (Figure 12c) shows that the majority of chiral centers arise from being a key part of the core scaffold (~88%), with only a small number being introduced as substituents. While this figure will incorporate a degree of bias due to compound series containing a chiral core increasing this count, it does suggest that there is room for increased consideration of the introduction of chiral substituents within medicinal chemistry programs.

Additionally, we examined the source of chiral centers in the data set (Figure 12 d; see Supporting Information Table S4 for detailed breakdown). It is unsurprising that in cases where the source of chirality could be unambiguously identified, it was introduced through the modification of commercially available enantiomerically pure starting materials in the majority of cases (87%). The remaining enantiopure stereocenters were introduced in almost equal portions by the use of asymmetric process or by resolution processes. In the case of resolutions, in stark contrast with that observed for process chemistry, there were no reports of diastereomeric salt formation and separation by selective crystallization, the preferred method among process chemists.[2]

A recent report by Lovering shows that the proportion of compounds containing at least one chiral center increases from ~50% in the discovery phase to ~60% of approved drugs,[12] perhaps suggesting that medicinal chemists should be further increasing the proportion of chiral compounds synthesized, perhaps as indicated above by use of more chiral substituents or by seeking to modify the core to contain a chiral center.

5.3. "Flatness": Fsp3. In addition to considering chirality, Lovering et al.[12] also proposed a simple measure of the degree of saturation, Fsp3 (defined as the ratio of sp^3-hybridized carbon atoms to all carbon atoms within the structure), which increased from an average value of 0.36 in the discovery phase to 0.47 in approved drugs. The distribution of Fsp3 values in this data set is shown in Figure 12e. It would appear that the majority of compounds have an Fsp3 value in the range 0.2−0.5, suggesting a fair degree of "nonflatness" within the data set, although, in common with the situation for chirality, Lovering's analysis suggests that this is an area that could still be improved significantly within medicinal chemistry programs.

5.4. Aromatic Ring Count. Ritchie and Macdonald have shown that increasing numbers of aromatic rings in a compound correlate with a decrease in compound developability measures, in particular suggesting that compounds with more than three aromatic rings are likely to have an increased risk of attrition during the development process.[18] Figure 13a shows the distribution of aromatic ring counts per compound within our data set. While it can be seen that the majority of compounds fall within this figure (73% of compounds had zero to three aromatic rings), a significant number had four aromatic rings, although the number of rings drops off sharply beyond this.

A more recent and detailed analysis of the effect of ring type on compound developability has suggested that benzenoid rings are more problematic than heteroaromatic rings and indeed that the proportion of heteroaromatic to benzenoid rings in a molecule appears to be significant, increasing proportions of benzenoid rings also being adverse.[81] To investigate how this data set compares with this, we looked initially at the simple distributions of number of benzenoid and heteroaromatic rings in compounds in the data set (Figure 13b). Perhaps concerningly, the distribution of benzenoid rings is weighted toward higher numbers than the distribution of heteroaromatics, with almost half of all compounds containing two benzenoid rings, in contrast to

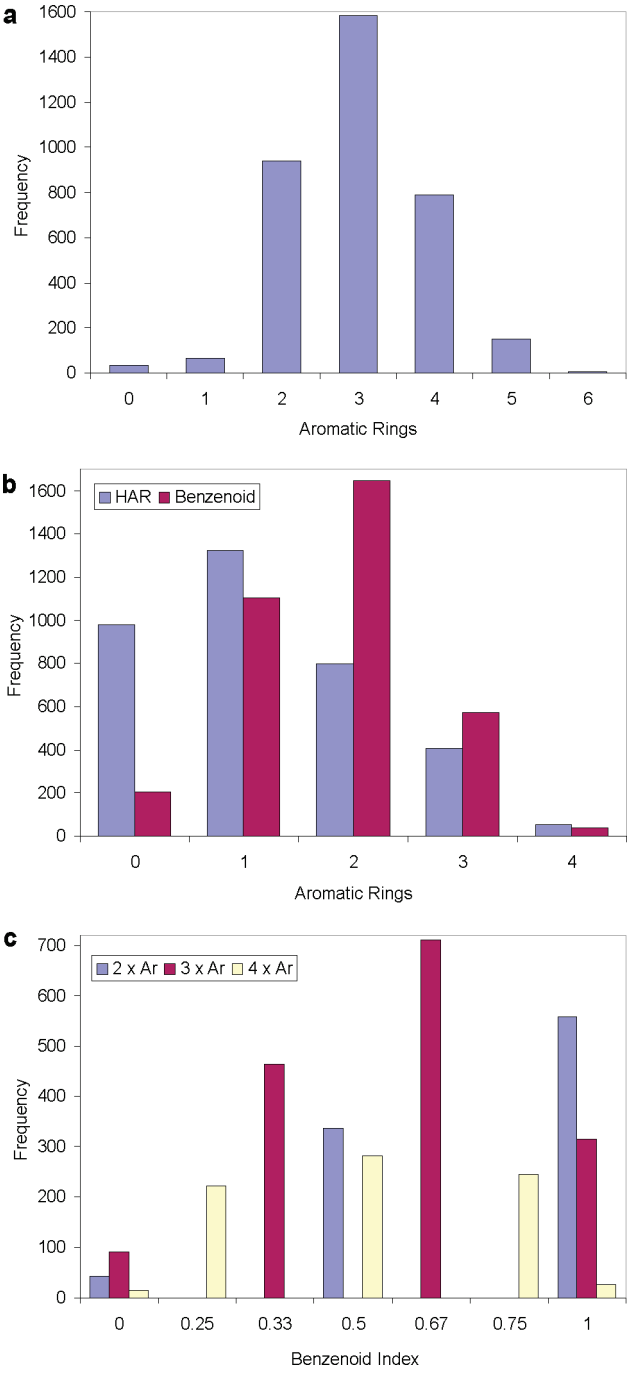


Figure 13. Aromatic ring counts: (a) distribution of number of aromatic rings per molecule; (b) distribution of benzenoid and heteroatomatic (HAR) rings per molecule; (c) benzenoid index (see text) for two to four aromatic-ring-containing compounds.

almost 1000 compounds (27%) containing no heteroaromatic rings, and a maximum in the distribution of 1 heteraromatic ring. In order to quantify this ratio further, we calculated a measure which we refer to as the "benzenoid index", defined in eq 1:

$$\text{benzenoid index (BI)} = \frac{\text{no. of benzenoid rings}}{\text{total no. of aromatic rings}} \quad (1)$$

Thus, a compound containing three aromatic rings, including one heteroaromatic ring, will have a BI of 0.67. By use of this

dx.doi.org/10.1021/jm200187y | J. Med. Chem. 2011, 54, 3451–3479

definition, BI values range from 0 to 1, with 1 indicating a compound in which the aromatic rings are entirely benzenoid. On the basis of Ritchie et al.'s analysis,[81] increasing values are indicated as more likely to suffer from poor compound developability. As different numbers of aromatic rings give a different set of discrete values for this index (a system with three rings can have BI of 0, 0.33, 0.67, or 1, whereas four rings can give 0, 0.25, 0.5, 0.75, and 1), but with some overlap (most notably 0 and 1, which will distort the distribution toward these values), we decided to plot the distribution for each number of rings separately (an alternative approach, particularly for larger data sets, would be to bin the values across all ring counts). We chose to ignore those systems with only a single aromatic ring (as this gives only 0 or 1) and those systems with five or more rings, as the BI values are distributed over a larger number of values for only a relatively small number of compounds in the data set. It can be seen (Figure 13c), however, that for two- and three-ring systems, higher BI values dominate, while for four-ring systems, there is a much flatter distribution. Indeed there is a trend toward lower BI as the number of rings increase. It is clear, however, that medicinal chemists need to be aware of the potential implications for compound development highlighted by Ritchie et al. of both aromatic ring count and type.[18,81]

6. PHYSICOCHEMICAL ("LIPINSKI") PROPERTIES

Following Lipinksi's seminal work,[7] linking a number of simple physicochemical properties (MWt, H-bond donors and acceptors and clogP) to human oral absorption of drug molecules (known as the "rule of 5"),[8,138] the importance of physicochemical properties in drug discovery has received wide recognition. Later papers have added the number of rotatable bonds (NRot)[10] and polar surface area (PSA[10] or TPSA[11]), along with many other more or less esoteric parameters[139] to predict oral bioavailability. Others have attempted to extend this approach to predict brain penetration,[8] crossing of the placenta,[140] and exposure by a plethora of routes including inhalation, opthalmic administration, and transdermal absorption,[141] with varying degrees of success. Particular attention has been paid to the importance of lipophilicity (clogP), which has been linked, in addition to solubility and permeability, to cytochrome P450 inhibition,[142] hERG binding,[143] increased adverse toxicological outcomes,[144] and off-target effects,[95] and recent papers have suggested tighter constraints on this parameter.[144−150] While not the primary focus of this review, we looked at how the compounds in this data set were distributed among these physicochemical properties (Figure 14). The number of compounds with MWt > 500 rapidly tails off, although there are still a significant number of compounds in the 500 < MWt ≤ 600 range (Lipinski's original paper allows for one violation of his rules[7]). The number of compounds violating the proposed limits of H-bond acceptors (≤10) and H-bond donors (≤5) is very small; almost all compounds in the data set fall within these limits and many considerably below them. Roughly 10% of compounds exceed Veber's proposed limit for NRot (≤10),[10] while most surprisingly, significant numbers (∼20%) of compounds exceed even Lipinski's proposed clogP limit of 5. It perhaps provides some vindication of Lipinski's desire to appeal to the pattern recognition skills of medicinal chemists in formulating his "rule of 5" that the most violated of these rules is the clogP constraint, which is also the only one of his parameters not readily calculated directly from the molecular structure without recourse to computational tools. TPSA has the flattest distribution across the data set; however, the majority of compounds fall in the range 50−100 Å^2, which is lower than the originally proposed limit of 140 Å^2.[10]

If one considers the number of Lipinski's rules violated, then 1310 compounds violate at least one rule (a scenario allowed in the original publication[7] but often overlooked). Of these, only 326 (9% of the compounds) break two (318 compounds) or more (three violations, four compounds; four violations, four compounds) of the rules and would thus be categorized as likely to have poor oral bioavailability. Leeson has published similar findings in a more detailed analysis of physicochemical properties of compounds synthesized by leading drug companies.[147] More recent reports proposing harsher cut-offs postdate the publication dates for these compounds, which in most cases will have been synthesized several years prior to their publication, and so they are not considered here except to suggest that the medicinal chemistry community needs to make continued efforts to restrain the lipophilicity of the compounds synthesized.

7. SUMMARY AND CONCLUSIONS

The analysis of this data set has revealed that while the medicinal chemist's perceived reliance on a small number of reactions (amide formations and Suzuki cross-couplings being the most often cited) is generally true (10 reaction types comprise almost two-thirds of all reactions), these processes and their resulting products are not as universal as might be believed. In addition to these processes, a large diversity of reaction types and functionalities are reported by medicinal chemists, although the favoring of a small number of reliable performers is understandable given the pressures of compound delivery. The robust, reliable nature of palladium-mediated cross-couplings and their wide applicability, chemoselectivity, and functional group tolerance make them natural candidates for the rapid generation of compound sets to ask specific questions around the biological importance of key substituents. Indeed, this importance and dependability were recognized most notably by the award of the 2010 Nobel Prize for Chemistry for the pioneering work in this area.[151] Furthermore, the reliance upon a limited subset of reaction methodologies appears to have helped enable and incentivize commercial reagent suppliers to cater to these specific needs, resulting in large numbers of cost-effective, diverse, and readily available starting reagents, facilitating rapid evaluation of SAR without recourse to the in-house synthesis of large numbers of bespoke building blocks. Conversely, it can be argued that lack of commercial availability of reagents may also detract from the popularity of some of the less commonly used reactions. It is interesting to ponder whether those processes that are less commonly used are avoided because of a dearth of reliable methods for their execution or whether such methods have not been developed because of a perceived lack of desire to incorporate their resultant functionalities.

The reliance on a small number of reaction types with the properties described above highlights the need for new methodologies or improvements to existing transformations, making them more generally applicable and amenable to parallel chemistry, as noted previously.[4] This would facilitate their use in the later stages of medicinal chemistry syntheses, potentially broadening the diversity of compounds synthesized. Indeed, the ACS Green Chemistry Initiative Pharmaceutical Roundtable recently issued a "call to arms" to address this very requirement, specifying a number of processes that were felt, across the industry, to be in need of additional research

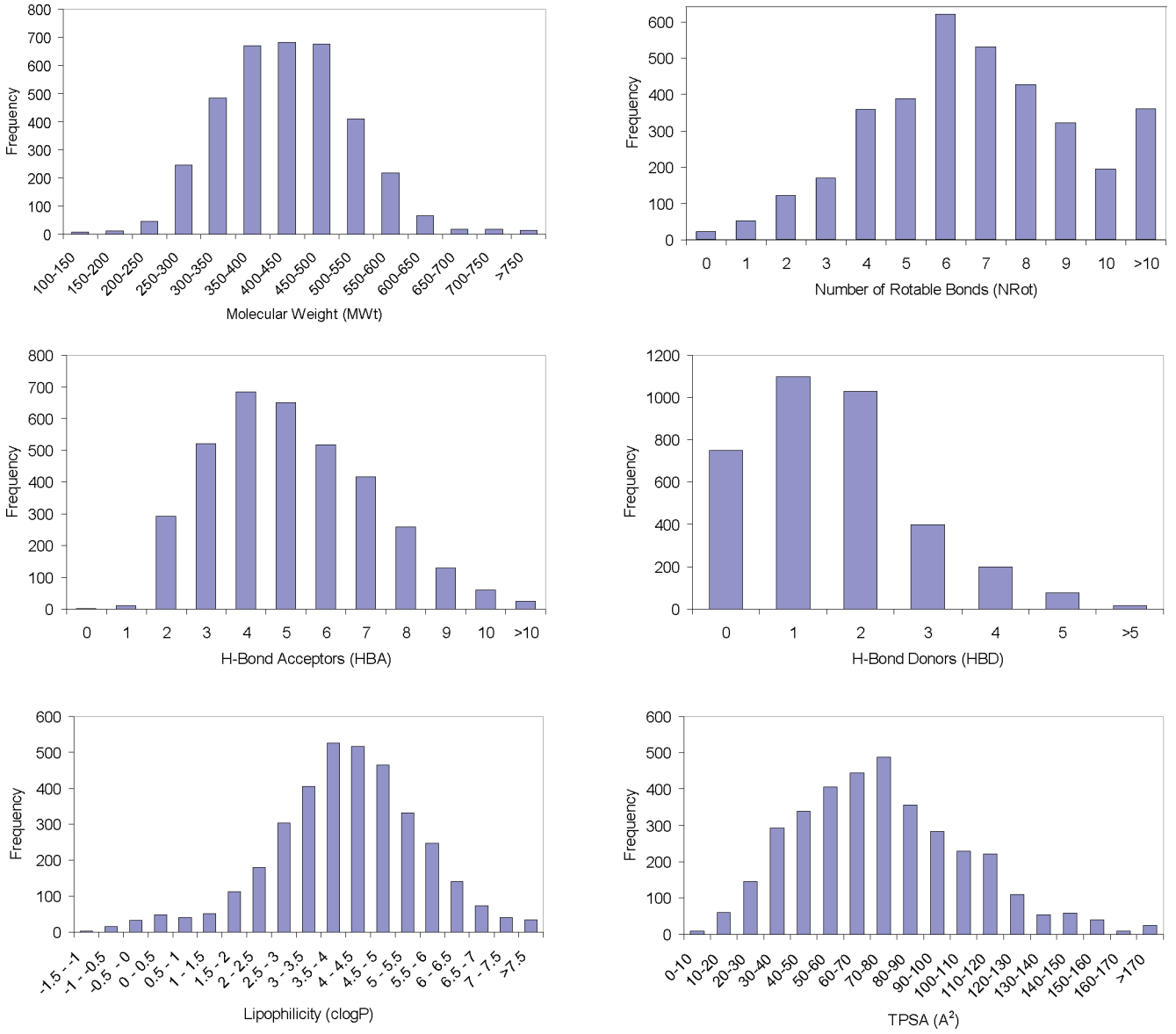


Figure 14. Distribution of physicochemical properties in the data set: MWt, molecular weight; NRot, number of rotatable bonds; HBA, H-bond acceptors; HBD, H-bond donors; TPSA, topological polar surface area.

and optimization.[152] This suggested not only more aspirational ideals to improve upon areas of chemistry that were clearly not fully developed or widely applicable but also requests for further improvements to tried and tested (and widely employed) methodologies, where clear deficiencies in areas such as atom economy and safety were identified. In addition to allowing exploration of additional chemical space and structural diversity, such methodologies would improve the safety, cost effectiveness, and scalability of the transformations employed in the medicinal chemistry laboratory.

The discovery and development of robust conditions for existing and novel transformations that are readily practicable in the context of medicinal chemistry (i.e., amenable to parallel synthesis approaches, using readily available reagents, high yielding, with broad substrate scope, tolerability, and reliability, not requiring extremes of temperature, rigorous exclusion of oxygen and moisture, simple workup and purification) are key areas in which the academic synthetic chemistry community can provide invaluable input to future medicinal chemistry programs.

Despite advances in chemo- and regioselective syntheses and transformations with increasing tolerance of a wide range of functionalities, one in five of all transformations analyzed in this paper are involved in protecting group manipulation. While some notable efforts have been made toward protecting-group-free synthetic strategies,[21−23] this area of synthetic methodology is far from fully developed and a key benefit of novel transformations such as those discussed above would be the ability to run such reactions without the need for recourse to protecting groups and the resultant deprotection steps, which are clearly wasteful in terms of time, reagent costs, and overall yield. Until such methods are more widely developed, the use of protecting groups remains a necessary synthetic burden to facilitate preparation of required molecules, requiring optimization of extra protection and deprotection steps. Until protecting-group-free methodology becomes more widespread, medicinal chemists must tolerate the use of protecting groups while bearing in mind that their use is not always necessary. In the authors' experience, it is often the

case that deprotection can be moved forward in a synthesis prior to later diversification steps while maintaining chemoselectivity by appropriate choice of reaction and workup conditions.

While less common in this data set (7% of reactions), redox processes have also been highlighted by Baran[78] as a source of synthetic inefficiency, in particular, those processes where multiple adjustments are made successively to the redox state of a single group, an area we highlighted in the sections dealing with these transformations. Of particular note in this review was the repeated use of a two-step ester to aldehyde conversion, comprising over-reduction to the intermediate alcohol, followed by oxidation to the required aldehydes. While this transformation can often be achieved directly using diisobutylaluminum hydride, this is often considered a poorly controlled reaction with troublesome workup, requiring inert atmospheres and low temperatures. New conditions for this process with operational simplicity comparable to that of the reduction of ketones with sodium borohydride would be a welcome addition to the medicinal chemists' toolkit. It is a testament to the creativity of medicinal chemists that within these parameters (constraints of time, the need for parallel-amenable chemistry, and speed of access to products), such a broad diversity of biologically and therapeutically relevant compounds are produced from the available transformations.

In recent years, microwave chemistry has found widespread use in the medicinal chemistry community, where its ability to provide dramatic rate enhancements, particularly in the areas of metal-catalyzed reactions, has allowed for significant increases in efficiency.[153−156] However, alternative complementary technologies, such as flow chemistry reactors, are being embraced more slowly by the medicinal chemistry community. Indeed there were no reports in our data set of the use of flow chemistry systems, despite the ability to provide a safe and scalable entrance into otherwise hazardous reactions such as the Curtius degradation.[157−161] Clearly, medicinal chemists must continue to investigate and embrace new enabling technologies, while equipment manufacturers and the academic community need to make new technologies readily accessible to the inexperienced user.

Medicinal chemistry is often perceived of as the synthesis of planar ("flat") aromatic systems lacking chirality. However, the data set reveals a surprisingly large number of chiral compounds, often containing multiple chiral centers, with an average of ∼1 defined stereocenter per molecule, taken across the entire data set. The perception, thus, that medicinal chemistry is entirely "flat" appears misguided, although it is true that aromatic systems are almost universally present, the likely source of this preconception. Furthermore, an assessment of the degree of unsaturation using the Fsp3 parameter[12] shows that between one-quarter and one-half of carbon atoms are sp^3-hybridized. Despite this, Lovering's 2009 analysis suggests that the incorporation of a greater proportion of chiral compounds with higher degrees of unsaturation may improve clinical outcomes; Lovering's Fsp3 parameter provides a simple way of assessing this for any given structure.[12] As chiral substituents are in the minority, it would seem that the routine inclusion of a greater range of chiral substituents may be a simple and achievable means of addressing this, while consideration should also be given to modification of aromatic core templates to saturated or partially saturated systems, containing chiral centers at the point(s) were substituents are attached.

As noted in a recent publication,[27] the extent of coverage of heterocycle space is surprisingly limited, although it is clear in this analysis that different companies have favored alternative heterocyclic systems, presumably in order to gain access to clear intellectual property areas. A relatively small number of aromatic heterocycles (three systems of pyridine, pyrimidine, and pyrazole) account for half of all heteroaromatics, while fusion increases the diversity of heterocyclic motifs by a remarkably small amount, albeit into areas with much less overlap between each company in this analysis. Again, our analysis suggests that medicinal chemists should actively seek a broader range of novel heterocyclic motifs, and in particular, care needs to be taken regarding both the overall aromatic ring count and the relative proportions of benzenoid and heteroaromatic rings (indicated by our proposed BI index). New synthetic methods will need to be developed in order to access novel ring systems, but the academic synthetic community is perhaps unlikely to rise to the challenge unless it can be justified by evidence of biological activity of such ring systems. The incorporation of cyclic amine solubilizing groups, a common approach within medicinal chemistry to improve solubility and bioavailability of compounds (and indeed one of the perceptions we sought to investigate), has been reported to be beneficial for developability,[81] as perhaps expected because of the improved solubility conferred by such modification. However, there is considerable scope for increasing the diversity of amine groups used (see section 4.2) for this purpose. Furthermore, recent reports have also suggested that such groups may contribute to potential hERG toxicity and other off-target promiscuity,[147,148] suggesting that, where possible, nonbasic solubilizing strategies may be preferable.

Halogens, in particular fluorine and chlorine, are often incorporated to solve metabolic problems or modulate off-target effects. Their use, as such, is often to "fix" an otherwise suboptimal molecule. The increasing potential for the heavier halogens to participate in halogen bonding interactions with biological targets was alluded to in section 4.8.[123,127−129] This interaction, in particular, its designed use, is gaining slow recognition within drug discovery. Indeed there were only two closely related reports in our data set of such an interaction, and even this was not explicitly referred to as a halogen bonding interaction. Greater awareness of this potential interaction, along with the development of improved methods for the incorporation of halogens (particularly fluorine, as highlighted in section 4.8) within the academic community, may allow for the design of improved molecules, in which a halogen serves *in place of* an interacting aromatic substituent, often an electron-rich group that introduces metabolic liabilities, reducing metabolic liability a priori instead of ameliorating it later with minimal potency benefit.

The problems of drug attrition during clinical trials have been well documented.[162−164] While the continued use of molecular fragments known to be "safe" (i.e., found in approved drugs) is understandable, the exploration of novel areas of chemical space, in which the current reliance in particular on amines, amides, and aromatic rings, may be biasing the chemical output into areas of space prone to development problems.

Finally, having considered some of the common preconceptions, it is interesting to consider what this data set suggests is a "typical" medicinal chemistry compound and synthesis. In terms of physicochemical properties, the analysis suggests a molecular weight in the range 350−550, with four to six HBAs and one to two HBDs, NRot in the range 6−8, and clogP between 3.5 and 5.5. TPSA has a broader distribution but typically is 60−90 $Å^2$. The compound will possess one to two chiral centers, with 30−50% of C atoms in the sp^3 hybridization state, and contain a biaryl bond between a fused aromatic system and a second ring (with one of the

rings being a benzenoid system). The molecule is also likely to possess a solubilizing group (probably morpholine or piperazine, linked via an ethylene linker to an aryl ether or via an $-CH_2-$ linker to an aromatic ring), an amide, and an aromatic fluoride or chloride. Synthetically, it will have been accessed in four to six steps, which include an amide formation, a deprotection step (most likely of a *N*-Boc group, introduced as part of a commercially available building block), and a Pd-catalyzed C−C bond formation (most likely a Suzuki coupling). The solubilizing group will have been introduced by either a reductive amination or an O-alkylation step. A search of the data set showed that there is no actual molecule within the set fulfilling all of these "typical" criteria!

We hope that this article will serve to put many of the preconceptions surrounding industrial medicinal chemistry into context, providing a snapshot based on recently published data, and suggest areas that medicinal chemists and academic synthetic chemists may explore in order to benefit the drug discovery process and to stimulate discussion within and between those communities.

Accession Codes

[†]The following PDB code is referenced in this article: 3DY7.

■ ASSOCIATED CONTENT

Supporting Information. Tables detailing the numbers of references and compounds by target class; table listing breakdown of the sources of chirality. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: (+44) 01223 895555. Fax: (+44) 01223 895556. E-mail: s.roughley@vernalis.com.

■ BIOGRAPHIES

Stephen D. Roughley obtained his M.A. in 1995 and Ph.D. (with Prof. Andrew B. Holmes, on the modeling and application of nitrone cycloaddition reactions to the synthesis of histrionicotoxin alkaloids) in 1999 from the University of Cambridge, U.K. During his studies he undertook placements in both Medicinal and Process Chemistry Departments at GlaxoWellcome. In 1999, he joined RiboTargets (later Vernalis) as a medicinal chemist, where he has been involved in a range of drug discovery and technology programs in anti-infective, CNS, and oncology disease areas. Following a secondment in NMR-based fragment screening, he returned to medicinal chemistry as a Principal Scientist and maintains a broad range of interests in drug discovery technologies, medicinal and synthetic chemistry, and cheminformatics.

Allan M. Jordan gained his B.Sc. from UMIST, Manchester, U. K., in 1993. After a short period as a graduate teaching assistant at Arizona State University, he returned to UMIST where he completed his Ph.D. in 1997, investigating taxane-derived anticancer agents with Nick Lawrence and Alan McGown. After postdoctoral studies with Helen Osborn at Reading University, U.K., he moved to Cambridge to join RiboTargets (which later became part of Vernalis), where he contributed to a number of CNS, oncology, and anti-infective programs. In 2009 he left Vernalis to return to Manchester, U.K., taking up his present position as Head of Chemistry in the newly formed Cancer Research UK Drug Discovery Unit at the Paterson Institute for Cancer Research.

■ ACKNOWLEDGMENT

The many colleagues past and present who have made comments in informal discussions on the topics covered in this manuscript or on various drafts of the manuscript are gratefully acknowledged by the authors. Dr. James Davidson (Vernalis (R&D) Ltd.) is particularly thanked for introducing one of the authors (S.D.R.) to Python, RDKit, and Knime. The authors are indebted to the many scientists at AstraZeneca, GlaxoSmithKline, and Pfizer who have taken the time to publish the results of their work in peer-reviewed journals and without whom this analysis would not have been possible. The authors thank the anonymous reviewers for the helpful comments and suggestions made on the manuscript during the review process.

■ ABBREVIATIONS USED

5-HT, 5-hydroxytryptamine (serotonin); 5-mr, five-membered ring; 6-mr, six-membered ring; Ac, acetyl; ACE2, angiotensin-converting enzyme 2; ACS, American Chemical Society; Alk, alkyl group; AMPAR, α-amino-3-hydroxy-5-methyl-4-isoxazole-propionic acid receptor; ANRORC, addition (nucleophilic) ring-opening ring closure; Ar, generic aromatic group; AZ, AstraZeneca; BACE-1, β-secretase 1; BMC, *Bioorganic and Medicinal Chemistry*; BMCL, *Bioorganic and Medicinal Chemistry Letters*; Bn, benzyl; Boc, *tert*-butoxycarbonyl; Boc, *tert*-butoxycarbonyl; CAS, Chemical Abstracts Service; Cbz, benzyloxycarbonyl; CCR5, C−C chemokine receptor 5; clogP, calculated $\log_{10}$ of the octanol/water partition coefficient; ^{c}Pr, cyclopropyl; CRF1, corticotropin releasing hormone receptor 1; CXCR2, CXC chemokine receptor 2; DIPEA, diisopropylethylamine; DPPA, diphenylphosphorylazide; EphB4, ephrin type B receptor 4; erbB2, erythroblast leukaemia viral oncogene B2 (aka HER2, human epidermal growth factor receptor 2); ER, estrogen receptor; Et, ethyl; FG, functional group; FGA, functional group addition; FGI, functional group interconversion; GSK, GlaxoSmithKline; HAR, heteroaromatic; HBA, hydrogen bond acceptor; HBD, hydrogen bond donor; hERG, human ether-a-go-go; IF, impact factor; ^{i}Pr, isopropyl; JMC, *Journal of Medicinal Chemistry*; MAdCAM, mucosal addressin cell adhesion molecule; Me, methyl; MEK1, Raf-activated MAP/ERK kinase; Ms, mesylate ($MeSO_2^-$); MWt, molecular weight; NCE, new chemical entities; NET, noradrenaline transporter; ^{n}Pr, *n*-propyl; NRot, number of freely rotatable bonds; PDB, Protein Data Bank; PSA, polar surface area; PYK2, proline-rich tyrosine kinase 2; R, R′, etc., generic substituents (alkyl or unspecified); SAR, structure −activity relationship; SERT, serotonin (5-HT) transporter; S_NAr, nucleophilic aromatic substitution; Sol, solubilizing group; $S_{RN}1$, radical aromatic substitution; ^{t}Bu, *tert*-butyl; Tf, triflate ($F_3CSO_2^-$); TFA, trifluoroacetic acid; TPSA, topological polar surface area; Ts, tosylate (p-$MeC_6H_4SO_2^-$); UMIST, University of Manchester Institute of Science and Techonology; VAL, valine; X, halogen; *n*-mr, *n*-membered ring

■ REFERENCES

(1) Jordan, A. M.; Roughley, S. D. Drug discovery chemistry: a primer for the non-specialist. *Drug Discovery Today* **2009**, *14*, 731–744.

(2) Carey, J. S.; Laffan, D.; Thomson, C.; Williams, M. T. Analysis of the reactions used for the preparation of drug candidate molecules. *Org. Biomol. Chem.* **2006**, 2337–2347.

(3) Thomson Reuters ISI impact factors for 2008 were retrieved from the journal Internet homepages (JMC, http://pubs.acs.org/

journal/jmcmar; BMC, http://journals.elsevier.com/09680896/bioorganic-and-medicinal-chemistry/; BMCL, http://www.elsevier.com/wps/find/journaldescription.cws_home/972/description#description) at the time of data compilation.

(4) Cooper, T. W. J.; Campbell, I. B.; Macdonald, S. J. F. Factors determining the selection of organic reactions by medicinal chemists and the use of these reactions in arrays (small focussed libraries). *Angew. Chem., Int. Ed.* **2010**, *49*, 2–12.

(5) CAS SciFinder. http://www.cas.org/products/scifindr/index.html (accessed Feb 7, 2011).

(6) Reaxys Elsevier Beilstein Database. http://www.reaxys.com/info (accessed Feb 7, 2011).

(7) Lipinksi, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **2001**, *46*, 3–26.

(8) Lipinksi, C. A. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discovery Today: Technol.* **2004**, *1*, 337–341.

(9) log*P* is calculated in RDKit using the atomic contribution algorithm of Crippen: Wildman, S. A.; Crippen, G. M. Prediction of physicochemical parameters by atomic contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868–873.

(10) Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **2002**, *45*, 2615–2623.

(11) Ertl, P.; Rohde, B.; Selzer, P. 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.

(12) Lovering, F.; Bikker, J.; Humblet, C. Escape from Flatland: increasing saturation as an approach to improving clinical success. *J. Med. Chem.* **2009**, *52*, 6752–6756.

(13) Python. http://www.python.org (accessed Mar 19, 2010).

(14) RDKit: Open-Source Cheminformatics. http://www.rdkit.org (accessed Apr 25, 2010).

(15) SMARTS—A Language for Describing Molecular Patterns. http://www.daylight.com/dayhtml/doc/theory/index.pdf (accessed Jan 20, 2011).

(16) Schomburg, K.; Ehrlich, H. C.; Stierand, K.; Rarey, M. From structure diagrams to visual chemical patterns. *J. Chem. Inf. Model.* **2010**, *50*, 1529–1535.

(17) SMARTSViewer. http://smartsview.zbh.uni-hamburg.de/ (accessed Jan 20, 2011).

(18) Ritchie, T. J.; Macdonald, S. J. F. The impact of aromatic ring count on compound developability: Are too many aromatic rings a liability in drug design? *Drug Discovery Today* **2009**, *14*, 1011–1020.

(19) Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kötter, T.; Meinl, T.; Ohl, P.; Sieb, C.; Thiel, K.; Wiswedel, B. KNIME: The Konstanz Information Miner. In *Studies in Classification, Data Analysis, and Knowledge Organization*; Springer: Berlin, Germany, 2007.

(20) KNIME: The Konstanz Information Miner. http://www.knime.org (accessed Jan 20, 2011).

(21) Baran, P. S.; Maimone, T. J.; Richter, J. M. Total synthesis of marine natural products without using protecting groups. *Nature* **2007**, *446*, 404–408.

(22) Gaich, T.; Baran, P. S. Aiming for the ideal synthesis. *J. Org. Chem.* **2010**, *75*, 4657–4673.

(23) Shenvi, R. A.; O'Malley, D. P.; Baran, P. S. Chemoselectivity: the mother of invention in total synthesis. *Acc. Chem. Res.* **2009**, *42*, 530–541.

(24) Young, I. S.; Baran, P. S. Protecting-group-free synthesis as an opportunity for invention. *Nat. Chem.* **2009**, *1*, 193–205.

(25) Ertl, P.; Jelfs, S.; Mühlbacher, J.; Schuffenhauer, A.; Selzer, P. Quest for the rings. In silico exploration of ring universe to identify novel bioactive heteroaromatic scaffolds. *J. Med. Chem.* **2006**, *49*, 4568–4573.

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

(27) Pitt, W. R.; Parry, D. M.; Perry, B. G.; Groom, C. R. Heteroaromatic rings of the future. *J. Med. Chem.* **2009**, *52*, 2952–2963.

(28) Abdel-Magid, A. F.; Mehrman, S. J. A review on the use of sodium triacetoxyborohydride in the reductive amination of ketones and aldehydes. *Org. Process Res. Dev* **2006**, *10*, 971–1031.

(29) Baxter, E. W.; Reitz, A. B. Reductive aminations of carbonyl compounds with borohydride and borane reducing agents. *Org. React.* **2002**, *59*, 1–714.

(30) Makosza, M. Nucleophilic substitution of hydrogen in electron-deficient arenes - a general process of great practical value. *Chem. Soc. Rev.* **2010**, *39*, 2855–2868.

(31) Van der Plas, H. C. The SN(ANRORC) mechanism: a new mechanism for nucleophilic substitution. *Acc. Chem. Res.* **1978**, *11*, 462–468.

(32) Rossi, R. A.; Pierini, A. B.; Santiago, A. N. Aromatic Substitution by the $S_{RN}1$ Reaction. *Org. React.* **2004**, 1–271.

(33) Hartwig, J. F. Carbon−heteroatom bond-forming reductive eliminations of amines, ethers, and sulfides. *Acc. Chem. Res.* **1998**, *31*, 852–860.

(34) Hartwig, J. F. Transition metal catalyzed synthesis of arylamines and aryl ethers from aryl halides and triflates: scope and mechanism. *Angew. Chem., Int. Ed.* **1998**, *37*, 2046–2−67.

(35) Hartwig, J. F. Carbon−heteroatom bond formation catalysed by organometallic complexes. *Nature* **2008**, *455*, 314–322.

(36) Surry, D. S.; Buchwald, S. L. Biaryl phosphane ligands in palladium-catalyzed amination. *Angew. Chem., Int. Ed.* **2008**, *47*, 6338–6361.

(37) Wolfe, J. P.; Wagaw, S.; Marcoux, J.-F.; Buchwald, S. L. Rational development of practical catalysts for aromatic carbon−nitrogen bond formation. *Acc. Chem. Res.* **1998**, *31*, 805–818.

(38) Wolfe, J. P.; Buchwald, S. L. Palladium-Catalyzed Amination of Aryl Halides and Aryl Triflates: *N*-Hexyl-2-methyl-4-methoxyaniline and *N*-Methyl-*N*-(4-chlorophenyl)aniline. *Organic Syntheses*; Wiley & Sons: New York, 2004; Collect. Vol. 10, p 423.

(39) Yang, B. H.; Buchwald, S. L. Palladium-catalyzed amination of aryl halides and sulfonates. *J. Organomet. Chem.* **1999**, *576*, 125–146.

(40) Corbet, J.-P.; Mignani, G. Selected patented cross-coupling reaction technologies. *Chem. Rev.* **2006**, *106*, 2651–2710.

(41) Hughes, D. L. The Mitsunobu reaction. *Org. React.* **1992**, *42*, 335–656.

(42) Mitsunobu, O. The use of diethyl azodicarboxylate and triphenylphoshine in synthesis and transformation of natural products. *Synthesis* **1981**, 1–28.

(43) Han, S. Y.; Kim, Y.-A. Recent development of peptide couping reagents in organic synthesis. *Tetrahedron* **2004**, *60*, 2447–2467.

(44) Montalbetti, C. A. G. N.; Falque, V. Amide bond formation and peptide coupling. *Tetrahedron* **2005**, *61*, 10827–10852.

(45) Ettmayer, P.; Amidon, G. L.; Clement, B.; Testa, B. Lessons learned from marketed and investigational prodrugs. *J. Med. Chem.* **2004**, *47*, 2393–2404.

(46) Rautio, J.; Kumpulainen, H.; Heimbach, T.; Oliyai, R.; Oh, D.; Järvinen, T.; Savolainen, J. Prodrugs: design and clinical applications. *Nat. Rev. Drug Discovery* **2008**, *7*, 255–270.

(47) Accelrys (formerly Symyx) Available Chemicals Directory, version 2010.08. http://accelrys.com/products/databases/sourcing/available-chemicals-directory.html (accessed Feb 2, 2011).

(48) Kotha, S. Recent applications of the Suzuki−Miyaura cross-coupling reaction in organic synthesis. *Tetrahedron* **2002**, *58*, 9633–9695.

(49) Miyaura, N.; Suzuki, A. Palladium-catalyzed cross-coupling reactions of organoboron compounds. *Chem. Rev.* **1995**, *95*, 2457–2483.

(50) Stanforth, S. P. Catalytic cross-coupling reactions in biaryl synthesis. *Tetrahedron* **1998**, *54*, 263–303.

(51) Many boronic acids and related derivatives are stable solids that can be easily weighed and stored, often for periods of years without serious degradation, in contrast with the organozinc and Grignard reagents of the Negishi and Kumada couplings.

(52) Compare with the Stille reaction of organostannanes, which are highly toxic, often significantly volatile, and generate toxic tin-based byproducts that can be troublesome to separate from the desired

product and represent a serious environmental hazard in their storage and disposal.

(53) Negishi, E.-I. A genealogy of Pd-catalyzed cross-coupling. *J. Organomet. Chem.* **2002**, *653*, 34–40.

(54) Negishi, E.-I. Palladium-catalyzed alkynylation. *Chem. Rev.* **2003**, *103*, 1979–2017.

(55) Sonogashira, K. Development of Pd-Cu catalyzed cross-coupling of terminal acetylenes with sp^2-carbon halides. *J. Organomet. Chem.* **2002**, *653*, 46–49.

(56) Brammertz, A.; Augthun, M. Toxicity of palladium in animals. *Wiss. Umwelt* **1992**, 285–289.

(57) Iavicoli, I.; Bocca, B.; Fontana, L.; Caimi, S.; Bergamaschi, A.; Alimonti, A. Distribution and elimination of palladium in rats after 90-day oral administration. *Toxicol. Ind. Health* **2010**, *26*, 183–189.

(58) Kielhorn, J.; Melber, C.; Keller, D.; Mangelsdorf, I. Palladium, a review of exposure and effects to human health. *Int. J. Hyg. Environ. Health* **2002**, *205*, 417–432.

(59) Nahm, S.; Weinreb, S. M. *N*-Methoxy-*N*-methylamides as effective acylating agents. *Tetrahedron Lett.* **1981**, *22*, 3815–3818.

(60) Labeeuw, O.; Phansavath, P.; Genêt, J.-P. Synthesis of modified Weinreb amides: *N*-*tert*-butoxy-*N*-methylamides as effective acylating agents. *Tetrahedron Lett.* **2004**, *45*, 7107–7110.

(61) Davies, S. G.; Goodwin, C. J.; Hepworth, D.; Roberts, P. M.; Thomson, J. E. On the origins of diastereoselectivity in the alkylation of enolates derived from *N*-1-(1′-naphthyl)ethyl-*O*-*tert*-butylhydroxamates: chiral Weinreb amide equivalents. *J. Org. Chem.* **2010**, *75*, 1214–1227.

(62) Boutagy, J.; Thomas, R. Olefin synthesis with organic phosphonate carbanions. *Chem. Rev.* **1974**, *74*, 87–99.

(63) Maercker, A. The Wittig reaction. *Org. React.* **1965**, *14*, 270–490.

(64) Maryanoff, B. E.; Reitz, A. B. The Wittig olefination reaction and modifications involving phosphoryl-stabilised carbanions. Stereochemistry, mechanism and selected synthetic aspects. *Chem. Rev.* **1989**, *89*, 863–927.

(65) Cowden, C. J.; Paterson, I. Asymmetric aldol reactions using boron. *Org. React.* **1997**, *51*, 1–200.

(66) Guillena, G.; Najera, C.; Ramon, D. J. Enantioselective direct aldol reaction: the blossoming of modern organocatalysis. *Tetrahedron: Asymmetry* **2007**, *18*, 2249–2293.

(67) Mukaiyama, T. The directed aldol reaction. *Org. React.* **1982**, 203–331.

(68) Palomo, C.; Oiarbide, M.; Garcia, J. M. The aldol addition reaction: and old transformation at constant rebirth. *Chem.—Eur. J.* **2002**, *8*, 36–44.

(69) Schetter, B.; Mahrwald, R. Modern aldol methods for the total synthesis of polyketides. *Angew. Chem., Int. Ed.* **2006**, *45*, 7506–7525.

(70) Trost, B. M.; Brindle, C. S. The direct catalytic asymmetric aldol reaction. *Chem. Soc. Rev.* **2010**, *39*, 1600–1632.

(71) Walters, I.; Austin, C.; Austin, R.; Bonnert, R.; Cage, P.; Christie, M.; Ebden, M.; Gardiner, S.; Grahames, C.; Hill, S.; Hunt, F.; Jewell, R.; Lewis, S.; Martin, I.; Nicholls, D.; Robinson, D. Evaluation of a series of bicyclic CXCR2 antagonists. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 798–803.

(72) Bardelle, C.; Coleman, T.; Cross, D.; Davenport, S.; Kettle, J. G.; Ko, E. J.; Leach, A. G.; Mortlock, A.; Read, J.; Roberts, N. J.; Robins, P.; Williams, E. J. Inhibitors of the tyrosine kinase EphB4. Part 2. Structure-based discovery and optimization of 3,5-bis substituted anilinopyrimidines. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5717–5721.

(73) Deaton, D. N.; Gao, E. N.; Graham, K. P.; Gross, J. W.; Miller, A. B.; Strelow, J. M. Thiol-based angiotensin-converting enzyme 2 inhibitors: P1 modifications for the exploration of the S1 subsite. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 732–737.

(74) Jiang, C.; Chang, J.-Y. Unfolding and breakdown of insulin in the presence of endogenous thiols. *FEBS Lett.* **2005**, *579*, 3927–3931.

(75) Heck, R. F. Palladium-catalyzed vinylation of organic halides. *Org. React.* **1982**, *27*, 345–390.

(76) Wan, Z.; Yan, H.; Hall, R. F.; Lin, X.; Livia, S.; Respondek, T.; Widdowson, K. L.; Zhu, C.; Callahan, J. F. Design and development of arrayable syntheses to accelerate SAR studies of pyridopyrimidinone and pyrimidopyrimidinone. *Tetrahedron Lett.* **2009**, *50*, 370–372.

(77) Yan, H.; Boehm, J. C.; Jin, Q.; Kasparec, J.; Li, H.; Zhu, C.; Widdowson, K. L.; Callahan, J. F.; Wan, Z. An improved and highly convergent synthesis of 4-substituted-pyrido[2.3-*d*]pyrimidin-7-ones. *Tetrahedron Lett.* **2007**, *48*, 1205–1207.

(78) Burns, N. Z.; Baran, P. S.; Hoffman, R. W. Redox economy in organic synthesis. *Angew. Chem., Int. Ed.* **2009**, *48*, 2854–2867.

(79) Baldwin, I.; Bamborough, P.; Haslam, C. G.; Hunjan, S. S.; Longstaff, T.; Mooney, C. J.; Patel, S.; Quinn, J.; Somers, D. O. Kinase array design, back to front: biaryl amides. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5285–5289.

(80) The authors have heard various figures mentioned in conference presentations and other discussions but never with supporting data or citations.

(81) Ritchie, T. J.; Macdonald, S. J. F.; Young, R. J.; Pickett, S. D. The impact of aromatic ring count on compound developability: further insights by examining carbo- and hetero-aromatic and -aliphatic ring types. *Drug Discovery Today* **2011**, *16*, 164–171.

(82) Jaeckel, H.; Klein, W. Prediction of mammalian toxicity by quantitative structure−activity relationships: aliphatic amines and anilines. *Quant. Struct.−Act. Relat.* **1991**, *10*, 198–204.

(83) Karki, S. B.; Dinnocenzo, J. P.; Jones, J. P.; Korzekwa, K. R. Mechanism of oxidative amine dealkylation of substituted *N*,*N*-dimethylanilines by cytochrome P-450: application of isotope effect profiles. *J. Am. Chem. Soc.* **1995**, *117*, 3657–3664.

(84) Zok, S.; Görge, G.; Kalsch, W.; Nagel, R. Bioconcentration, metabolism and toxicity of substituted anilines in the zebrafish (*Brachydanio rerio*). *Sci. Total Environ.* **1991**, *109−110*, 411–421.

(85) Ghose, A. K.; Herbertz, T.; Salvino, J. M.; Mallamo, J. P. Knowledge-based chemoinformatic approaches to drug discovery. *Drug Discovery Today* **2006**, *11*, 1107–1114.

(86) Akritopoulou-Zanze, I.; Hajduk, P. J. Kinase-targeted libraries: the design and synthesis of novel, potent, and selective kinase inhibitors. *Drug Discovery Today* **2009**, *14*, 291–297.

(87) Brown, A.; Brown, L.; Brown, T. B.; Calabrese, A.; Ellis, D.; Puhalo, N.; Smith, C. R.; Wallace, O.; Watson, L. Triazole oxytocin antagonists: identification of aryl ether replacements for a biaryl substituent. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5242–5244.

(88) Patel, S. D.; Habeski, W. M.; Min, H.; Zhang, J.; Roof, R.; Snyder, B.; Bora, G.; Campbell, B.; Li, C.; Hidayetoglu, D.; Johnson, D. S.; Chaudhry, A.; Charlton, M. E.; Kablaoui, N. M. Identification and SAR around *N*-{2-[4-(2,3-dihydro-benzo[1,4]dioxin-2-ylmethyl)-[1,4]diazepan-1-yl]-ethyl}-2-phenoxy-nicotinamide, a selective α2C adrenergic receptor antagonist. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5689–5693.

(89) Fish, P. V.; Deur, C.; Gan, X.; Greene, K.; Hoople, D.; MacKenny, M.; Para, K. S.; Reeves, K.; Ryckmans, T.; Stiff, C.; Stobie, A.; Wakenhut, F.; Whitlock, G. A. Design and synthesis of morpholine derivatives. SAR for dual serotonin & noradrenaline reuptake inhibition. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 2562–2566.

(90) Xu, W.; Gray, D. L.; Glase, S. A.; Barta, N. S. Design and synthesis of reboxetine analogs morpholine derivatives as selective norepinephrine reuptake inhibitors. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5550–5553.

(91) Whitlock, G. A.; Blagg, J.; Fish, P. V. 1-(2-Phenoxyphenyl)methanamines: SAR for dual serotonin/noradrenaline reuptake inhibition, metabolic stability and hERG affinity. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 596–599.

(92) Middleton, D. S.; Andrews, M.; Glossop, P.; Gymer, G.; Hepworth, D.; Jessiman, A.; Johnson, P. S.; MacKenny, M.; Pitcher, M. J.; Rooker, T.; Stobie, A.; Tang, K.; Morgan, P. Designing rapid onset selective serotonin re-uptake inhibitors. 2: Structure−activity relationships of substituted (aryl)benzylamines. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 4018–4021.

(93) Mente, S.; Gallaschun, R.; Schmidt, A.; Lebel, L.; Vanase-Frawley, M.; Fliri, A. Quantitative structure−activity relationship of phenoxyphenyl-methanamine compounds with 5HT2A, SERT and hERG activities. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 6088–6092.

(94) Fish, P. V.; Ryckmans, T.; Stobie, A.; Wakenhut, F. [4-(Phenoxy)pyridin-3-yl]methylamines: a new class of selective noradrenaline reuptake inhibitors. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 1795–1798.

(95) Whitlock, G. A.; Fish, P. V.; Fary, M. J.; Stobie, A.; Wakenhut, F. Pyridyl-phenyl ether monoamine reuptake inhibitors: impact of lipophilicity on dual SNRI pharmacology and off-target promiscuity. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 2896–2899.

(96) Pfefferkorn, J. A.; Choi, C.; Winters, T.; Kennedy, R.; Chi, L.; Perrin, L. A.; Lu, G.; Ping, Y.-W.; McClanahan, T.; Schroeder, R.; Leininger, M. T.; Geyer, A.; Schefzick, S.; Atherton, J. P2Y1 receptor antagonists as novel antithrombotic agents. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 3338–3343.

(97) Chen, Y. L.; Braselton, J.; Forman, J.; Gallaschun, R.; Mansbach, R.; Schmidt, A. W.; Seeger, T. F.; Sprouse, J. S.; Tingley, F. D.; Winston, E.; Schulz, D. W. Synthesis and SAR of 2-aryloxy-4-alkoxypyridines as potent orally active corticotropin-releasing factor 1 receptor antagonists. *J. Med. Chem.* **2008**, *51*, 1377–1384.

(98) Barlaam, B.; Ballard, P.; Bradbury, R. H.; Ducray, R.; Germaine, H.; Hickinson, D. M.; Hudson, K.; Kettle, J. G.; Klinowska, T.; Magnien, F.; Ogilvie, D. J.; Olivier, A.; Pearson, S. E.; Scott, J. S.; Suleman, A.; Trigwell, C. B.; Vautier, M.; Whittaker, R. D.; Wood, R. A new series of neutral 5-substituted 4-anilinoquinazolines as potent, orally active inhibitors of erbB2 receptor tyrosine kinase. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 674–678.

(99) Fang, J.; Akwabi-Ameyaw, A.; Britton, J. E.; Katamreddy, S. R.; Navas, F.; Miller, A. B.; Williams, S. P.; Gray, D. W.; Orband-Miller, L. A. Synthesis of 3-alkyl naphthalenes as novel estrogen receptor ligands. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5075–5077.

(100) Micheli, F.; Bonanomi, G.; Braggio, S.; Capelli, A. M.; Celestini, P.; Damiani, F.; Di Fabio, R.; Donati, D.; Gagliardi, S.; Gentile, G.; Hamprecht, D.; Petrone, M.; Radaelli, S.; Tedesco, G.; Terreni, S.; Worby, A.; Heidbreder, C. New fused benzazepine as selective D3 receptor antagonists. Synthesis and biological evaluation. Part one: [h]-fused tricyclic systems. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 901–907.

(101) Micheli, F.; Bonanomi, G.; Braggio, S.; Capelli, A. M.; Damiani, F.; Di Fabio, R.; Donati, D.; Gentile, G.; Hamprecht, D.; Perini, O.; Petrone, M.; Tedesco, G.; Terreni, S.; Worby, A.; Heidbreder, C. New fused benzazepine as selective D3 receptor antagonists. Synthesis and biological evaluation. Part 2: [g]-Fused and hetero-fused systems. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 908–912.

(102) Young, R. J.; Borthwick, A. D.; Brown, D.; Burns-Kurtis, C. L.; Campbell, M.; Chan, C.; Charbaut, M.; Chung, C.-W.; Convery, M. A.; Kelly, H. A.; King, N. P.; Kleanthous, S.; Mason, A. M.; Pateman, A. J.; Patikis, A. N.; Pinto, I. L.; Pollard, D. R.; Senger, S.; Shah, G. P.; Toomey, J. R.; Watson, N. S.; Weston, H. E. Structure and property based design of factor Xa inhibitors: pyrrolidin-2-ones with biaryl P4 motifs. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 23–27.

(103) Young, R. J.; Borthwick, A. D.; Brown, D.; Burns-Kurtis, C. L.; Campbell, M.; Chan, C.; Charbaut, M.; Convery, M. A.; Diallo, H.; Hortense, E.; Irving, W. R.; Kelly, H. A.; King, N. P.; Kleanthous, S.; Mason, A. M.; Pateman, A. J.; Patikis, A. N.; Pinto, I. L.; Pollard, D. R.; Senger, S.; Shah, G. P.; Toomey, J. R.; Watson, N. S.; Weston, H. E.; Zhou, P. Structure and property based design of factor Xa inhibitors: biaryl pyrrolidin-2-ones incorporating basic heterocyclic motifs. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 28–33.

(104) Walker, D. P.; Bi, C. F.; Kalgutkar, A. S.; Bauman, J. N.; Zhao, S. X.; Soglia, J. R.; Aspnes, G. E.; Kung, D. W.; Klug-McLeod, J.; Zawistoski, M. P.; McGlynn, M. A.; Oliver, R.; Dunn, M.; Li, J. C.; Richter, D. T.; Cooper, B. A.; Kath, J. C.; Hulford, C. A.; Autry, C. L.; Luzzio, M. J.; Ung, E. J.; Roberts, W. G.; Bonnette, P. C.; Buckbinder, L.; Mistry, A.; Griffor, M. C.; Han, S.; Guzman-Perez, A. Trifluoromethylpyrimidine-based inhibitors of proline-rich tyrosine kinase 2 (PYK2): structure−activity relationships and strategies for the elimination of reactive metabolite formation. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 6071–6077.

(105) Micale, N.; Colleoni, S.; Postorino, G.; Pellicanò, A.; Zappalà, M.; Lazzaro, J.; Diana, V.; Cagnotto, A.; Mennini, T.; Grasso, S. Structure−activity study of 2,3-benzodiazepin-4-ones noncompetitive AMPAR antagonists: identification of the 1-(4-amino-3-methylphenyl)-3,5-dihydro-7,8-ethylenedioxy-4*H*-2,3-benzodiazepin-4-one as neuroprotective agent. *Bioorg. Med. Chem.* **2008**, *16*, 2200–2211.

(106) Tian, X.; Switzer, A. G.; Derose, S. A.; Mishra, R. K.; Solinsky, M. G.; Mumin, R. N.; Ebetino, F. H.; Jayasinghe, L. R.; Webster, M. E.; Colson, A. O.; Crossdoersen, D.; Pinney, B. B.; Farmer, J. A.; Dowty, M. E.; Obringer, C. M.; Cruze, C. A.; Burklow, M. L.; Suchanek, P. M.; Dong, L.; Dirr, M. K.; Sheldon, R. J.; Wos, J. A. Discovery of orally bioavailable 1,3,4-trisubstituted 2-oxopiperazine-based melanocortin-4 receptor agonists as potential antiobesity agents. *J. Med. Chem.* **2008**, *51*, 6055–6066.

(107) Kazmierski, W. M.; Aquino, C.; Chauder, B. A.; Deanda, F.; Ferris, R.; Jones-Hertzog, D. K.; Kenakin, T.; Koble, C. S.; Watson, C.; Wheelan, P.; Yang, H.; Youngman, M. Discovery of bioavailable 4,4-disubstituted piperidines as potent ligands of the chemokine receptor 5 and inhibitors of the human immunodeficiency virus-1. *J. Med. Chem.* **2008**, *51*, 6538–6546.

(108) Liddle, J.; Allen, M. J.; Borthwick, A. D.; Brooks, D. P.; Davies, D. E.; Edwards, R. M.; Exall, A. M.; Hamlett, C.; Irving, W. R.; Mason, A. M.; McCafferty, G. P.; Nerozzi, F.; Peace, S.; Philp, J.; Pollard, D.; Pullen, M. A.; Shabbir, S. S.; Sollis, S. L.; Westfall, T. D.; Woollard, P. M.; Wu, C.; Hickey, D. M. B. The discovery of GSK221149A: a potent and selective oxytocin antagonist. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 90–94.

(109) Harriman, G. C.; Brewer, M.; Bennett, R.; Kuhn, C.; Bazin, M.; Larosa, G.; Skerker, P.; Cochran, N.; Gallant, D.; Baxter, D.; Picarella, D.; Jaffee, B.; Luly, J. R.; Briskin, M. J. Selective cell adhesion inhibitors: barbituric acid based α4β7—MAdCAM inhibitors. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 2509–2512.

(110) Clarke, B.; Demont, E.; Dingwall, C.; Dunsdon, R.; Faller, A.; Hawkins, J.; Hussain, I.; MacPherson, D.; Maile, G.; Matico, R.; Milner, P.; Mosley, J.; Naylor, A.; O'Brien, A.; Redshaw, S.; Riddell, D.; Rowland, P.; Soleil, V.; Smith, K. J.; Stanway, S.; Stemp, G.; Sweitzer, S.; Theobald, P.; Vesey, D.; Walter, D. S.; Ward, J.; Wayne, G. BACE-1 inhibitors part 1: identification of novel hydroxy ethylamines (HEAs). *Bioorg. Med. Chem. Lett.* **2008**, *18*, 1011–1016.

(111) Clarke, B.; Demont, E.; Dingwall, C.; Dunsdon, R.; Faller, A.; Hawkins, J.; Hussain, I.; MacPherson, D.; Maile, G.; Matico, R.; Milner, P.; Mosley, J.; Naylor, A.; O'Brien, A.; Redshaw, S.; Riddell, D.; Rowland, P.; Soleil, V.; Smith, K. J.; Stanway, S.; Stemp, G.; Sweitzer, S.; Theobald, P.; Vesey, D.; Walter, D. S.; Ward, J.; Wayne, G. BACE-1 inhibitors part 2: identification of hydroxy ethylamines (HEAs) with reduced peptidic character. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 1017–1021.

(112) Serafinowska, H. T.; Blaney, F. E.; Lovell, P. J.; Merlo, G. G.; Scott, C. M.; Smith, P. W.; Starr, K. R.; Watson, J. M. Novel 5-HT1A/1B/1D receptors antagonists with potent 5-HT reuptake inhibitory activity. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5581–5585.

(113) Bromidge, S. M.; Bertani, B.; Borriello, M.; Faedo, S.; Gordon, L. J.; Granci, E.; Hill, M.; Marshall, H. R.; Stasi, L. P.; Zucchelli, V.; Merlo, G.; Vesentini, A.; Watson, J. M.; Zonzini, L. 6-[2-(4-Aryl-1-piperazinyl)ethyl]-2*H*-1,4-benzoxazin-3(4*H*)-ones: dual-acting 5-HT1 receptor antagonists and serotonin reuptake inhibitors. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5653–5656.

(114) With 23 heteroaromatic rings and benzene exemplified within the data set, even without the possibility of different fusion bond isomers, there are 24 × 24 = 576 possible ring types (comprising 64 6,6-fused, 256 5,5-fused, and 128 6,5-fused systems). Fusion isomers increase this figure dramatically (for example, the relatively symmetrical pyridine and pyrrole pairing can generate seven different fusion isomers).

(115) Lewell, X. Q.; Jones, A. C.; Bruce, C. L.; Harper, G.; Jones, M. M.; Mclay, I. M.; Bradshaw, J. Drug rings database with Web interface. A tool for identifying alternative chemical rings in lead discovery programs. *J. Med. Chem.* **2003**, *46*, 3257–3274.

(116) Southall, N. T.; Ajay Kinase patent space visualisation using chemical replacements. *J. Med. Chem.* **2006**, *49*, 2103–2109.

(117) Jin, J.; Wang, Y.; Shi, D.; Wang, F.; Davis, R. S.; Jin, Q.; Fu, W.; Foley, J. J.; Webb, E. F.; Dehaas, C. J.; Berlanga, M.; Burman, M.; Sarau,

H. M.; Morrow, D. M.; Rao, P.; Kallal, L. A.; Moore, M. L.; Rivero, R. A.; Palovich, M.; Salmon, M.; Belmonte, K. E.; Busch-Petersen, J. Discovery of novel and long acting muscarinic acetylcholine receptor antagonists. *J. Med. Chem.* **2008**, *51*, 4866–4869.

(118) Kung, P. P.; Funk, L.; Meng, J.; Collins, M.; Zhou, J. Z.; Johnson, M. C.; Ekker, A.; Wang, J.; Mehta, P.; Yin, M. J.; Rodgers, C.; Davies, J. F., II; Bayman, E.; Smeal, T.; Maegley, K. A.; Gehring, M. R. Dihydroxylphenyl amides as inhibitors of the Hsp90 molecular chaperone. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 6273–6278.

(119) Deaton, D. N.; Graham, K. P.; Gross, J. W.; Miller, A. B. Thiol-based angiotensin-converting enzyme 2 inhibitors: P1′ modifications for the exploration of the S1′ subsite. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 1681–1687.

(120) Purser, S.; Moore, P. R.; Swallow, S.; Gouverneur, V. Fluorine in medicinal chemistry. *Chem. Soc. Rev.* **2008**, *37*, 320–330.

(121) O'Hagan, D. Understanding organofluorine chemistry. An introduction to the C−F bond. *Chem. Soc. Rev.* **2008**, *37*, 308–319.

(122) Hodgetts, K. J.; Combs, K. J.; Elder, A. M.; Harriman, G. C. The role of fluorine in the discovery and optimization of CNS agents: modulation of drug-like properties. *Annu. Rep. Med. Chem.* **2010**, *45*, 429–448.

(123) Bissantz, C.; Kuhn, B.; Stahl, M. A medicinal chemist's guide to molecular interactions. *J. Med. Chem.* **2010**, *53*, 5061–5084.

(124) Zhu, L.; Li, Y.; Zhao, Y.; Hu, J. Nucleophilic (phenylsulfonyl)-difluoromethylation of alkyl halides using $PhSO_2CF_2SiMe_3$: preparation of gem-difluoroalkenes and trifluoromethyl compounds. *Tetrahedron Lett.* **2010**, *51*, 6150–6152.

(125) Baumann, M.; Baxendale, I. R.; Martin, L. J.; Ley, S. V. Development of fluorination methods using continuous-flow microreactors. *Tetrahedron* **2009**, *65*, 6611–6625.

(126) Hagooly, Y.; Cohen, O.; Rozen, S. A general route for constructing difluoromethyl ethers. *Tetrahedron Lett.* **2009**, *50*, 392–394.

(127) Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S. Halogen bonds in biological molecules. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16789–16794.

(128) Lu, Y.; Ting, S.; Wang, Y.; Yang, H.; Yan, X.; Luo, X.; Jiang, H.; Zhu, W. Halogen bonding: a novel interaction for rational drug design? *J. Med. Chem.* **2009**, *52*, 2854–2862.

(129) Zhou, P.; Lv, J.; Zou, J.; Tian, F.; Shang, Z. Halogen −water−hydrogen bridges in biomolecules. *J. Struct. Biol.* **2010**, *169*, 172–182.

(130) Hernandes, M. Z.; Cavalcanti, S. M. T.; Moreira, D. R. M. Halogens atoms in the modern medicinal chemistry: hints for the drug design. *Curr. Drug. Targets* **2010**, *11*, 303–314.

(131) O'Dowd, H.; Lewis, J. G.; Trias, J.; Asano, R.; Blais, J.; Lopez, S. L.; Park, C. K.; Wu, C.; Wang, W.; Gordeev, M. F. Novel antibacterial azetidine lincosamides. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 2645–2648.

(132) Barrett, S. D.; Bridges, A. J.; Dudley, D. T.; Saltiel, A. R.; Fergus, J. H.; Flamme, C. M.; Delaney, A. M.; Kaufman, M.; LePage, S.; Leopold, W. R.; Przybranowski, S. A.; Sebolt-Leopold, J.; Van Becelaere, K.; Doherty, A. M.; Kennedy, R. M.; Marston, D.; Howard, W. A., Jr.; Smith, Y.; Warmus, J. S.; Tecle, H. The discovery of the benzhydroxamate MEK inhibitors CI-1040 and PD 0325901. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 6501–6504.

(133) Warmus, J. S.; Flamme, C.; Zhang, L. Y.; Barrett, S.; Bridges, A.; Chen, H.; Gowan, R.; Kaufman, M.; Sebolt-Leopold, J.; Leopold, W.; Merriman, R.; Ohren, J.; Pavlovsky, A.; Przybranowski, S.; Tecle, H.; Valik, H.; Whitehead, C.; Zhang, E. 2-Alkylamino- and alkoxy-substituted 2-amino-1,3,4-oxadiazoles—*O*-Alkyl benzohydroxamate esters replacements retain the desired inhibition and selectivity against MEK (MAP ERK kinase). *Bioorg. Med. Chem. Lett.* **2008**, *18*, 6171–6174.

(134) Protein Data Bank (PDB). http://www.rcsb.org/ (accessed Feb 13, 2011).

(135) Tecle, H.; Shao, J.; Li, Y.; Kothe, M.; Kazmirski, S.; Penzotti, J.; Ding, Y. H.; Ohren, J.; Moshinsky, D.; Coli, R.; Jhawar, N.; Bora, E.; Jacques-O'Hagan, S.; Wu, J. Beyond the MEK-pocket: Can current MEK kinase inhibitors be utilized to synthesize novel type III NCKIs? Does the MEK-pocket exist in kinases other than MEK? *Bioorg Med. Chem. Lett.* **2009**, *19*, 226–229.

(136) Note that this figure, being almost twice the total number of reactions (7315) in our data set, reflects the general approach of building a common scaffold, which is subsequently used to prepare multiple analogues.

(137) Manual inspection revealed a further 339 compounds containing 1 (315 compounds) or 2 (34 compounds) undefined or only partially defined stereocenters, a total of 383 additional stereocenters. Many of these were the result of the addition of simple substituents (Me, OH, OMe, etc.) to prochiral centers or from a small number of scaffolds with a racemic center. Presumably, the vast majority of these were not deemed of sufficient interest to invest in asymmetric synthesis or resolution.

(138) While Lipinski's rules relate to oral absorption, in most cases no intended dose route was identified in many of the papers examined. We assume for the purposes of this analysis that oral dosing would be the desired "gold standard".

(139) Lipinski notes in ref 7 the importance of their goal of generating simple parameters that are easily accessible to medicinal chemists using their pattern recognition skills.

(140) Giaginis, C.; Zira, A.; Theocharis, S.; Tsantili-Kakoulidou, A. Simple physicochemical properties as effective filters for risk estimation of drug transport across the human placental barrier. *Epitheor. Klin. Farmakol. Farmakokinet., Int. Ed.* **2008**, *22*, 146–148.

(141) Choy, Y.-B.; Prausnitz, M. R. The rule of five for non-oral routes of drug delivery: ophthalmic, inhalation and transdermal. *Pharm. Res.* **2011**, *28*, 943–948.

(142) Lewis, D. F. V.; Jacobs, M. N.; Dickins, M. Compound lipophilicity for substrate binding to human P450s in drug metabolism. *Drug Discovery Today* **2004**, *9*, 530–537.

(143) Waring, M. J.; Johnstone, C. A quantitative assessment of hERG liability as a function of lipophilicity. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 1759–1764.

(144) Hughes, J. D.; Blagg, J.; Price, D. A.; Bailey, S.; DeCrescenzo, G. A.; Devraj, R. V.; Elsworth, E.; Fobian, Y. M.; Gibbs, M. E.; Giles, R. W.; Greene, N.; Huang, E.; Krieger-Burke, T.; Loesel, J.; Wager, T.; Whiteley, L.; Zhang, Y. Physicochemical drug properties associated with in vivo toxicological outcomes. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 4872–4875.

(145) Gleeson, M. P. Generation of a set of simple, interpretable ADMET rules of thumb. *J. Med. Chem.* **2008**, *51*, 817–834.

(146) Gleeson, P.; Bravi, G.; Modi, S.; Lowe, D. ADMET rules of thumb II: a comparison of the effects of common substituents on a range of ADMET parameters. *Bioorg. Med. Chem.* **2009**, *17*, 5906–5919.

(147) Leeson, P. D.; Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discovery* **2007**, *6*, 881–890.

(148) Leeson, P. D.; Empfield, J. R. Reducing the risk of drug attrition associated with physicochemical properties. *Annu. Rep. Med. Chem.* **2010**, *45*, 393–407.

(149) Price, D. A.; Blagg, J.; Jones, L.; Greene, N.; Wager, T. Physicochemical drug properties associated with in vivo toxicological outcomes: a review. *Expert Opin. Drug Metab. Toxicol.* **2009**, *5*, 921–931.

(150) Waring, M. J. Lipophilicity in drug discovery. *Expert Opin Drug Discovery* **2010**, *5*, 235–248.

(151) The Nobel Prize in Chemistry 2010—Richard F. Heck, Ei-ichi Negishi, Akira Suzuki, "For Palladium-Catalyzed Cross Couplings in Organic Synthesis". http://nobelprize.org/nobel_prizes/chemistry/laureates/2010/ (accessed Feb 7, 2011).

(152) Constable, D. J. C.; Dunn, P. J.; Hayler, J. D.; Humphrey, J. L., Jr.; Linderman, R. J.; Lorenz, K.; Manley, J.; Pearlman, B. A.; Wells, A.; Zaks, A.; Zhang, T. Y. Key green chemistry research areas, a perspective from pharmaceutical manufacturers. *Green Chem.* **2007**, *9*, 411–420.

(153) Alcazar, J.; Oehrlich, D. Recent applications of microwave irradiation to medicinal chemistry. *Future Med. Chem.* **2010**, *2*, 169–176.

(154) Kappe, O. C.; Dallinger, D. The impact of microwave synthesis on drug discovery. *Nat. Rev. Drug Discovery* **2006**, *5*, 51–63.
(155) Mavandadi, F.; Pilotti, A. The impact of microwave-assisted organic synthesis in drug discovery. *Drug Discovery Today* **2006**, *11*, 165–174.
(156) Santagada, V.; Frecentese, F.; Perissuttu, E.; Fiorino, F.; Severino, B.; Caliendo, G. Microwave assisted synthesis: a new technology in drug discovery. *Mini-Rev. Med. Chem.* **2009**, *9*, 340–358.
(157) Ahmed-Omer, B.; Brandt, J. C.; Wirth, T. Advanced organic synthesis using microreactor technology. *Org. Biomol. Chem.* **2007**, *5*, 733–740.
(158) Mason, B. P.; Price, K. E.; Steinbacher, J. L.; Bogdan, A. R.; McQuade, D. T. Greener approaches to organic synthesis using microreactor technology. *Chem. Rev.* **2007**, *107*, 2300–2318.
(159) Jones, R. V.; Csajabi, C.; Szekelyhidi, Z.; Kovacs, I.; Borcsek, B.; Urge, L.; Darvas, F. Flow reactos for drug discovery flow for reaction optimization, library synthesis and scale up. *Chim. Oggi* **2008**, *26*, 10–12.
(160) Wiles, C.; Watts, P. Improving chemical synthesis using flow reactors. *Expert Opin Drug Discovery* **2007**, *2*, 1487–1503.
(161) Wiles, C.; Watts, P. Enhanced chemical synthesis in flow reactors. *Chim. Oggi* **2009**, *27*, 34–36.
(162) Borshell, N.; Papp, T.; Congreve, M. Deal watch: valuation benefits of structure-enabled drug discovery. *Nat. Rev. Drug Discovery* **2011**, *10*, 166.
(163) Czerepak, E. A.; Ryser, S. Drug approvals and failures: implications for alliances. *Nat. Rev. Drug Discovery* **2008**, *7*, 197–198.
(164) Kola, I.; Landis, J. Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discovery* **2004**, *3*, 711–715.