Journal of Medicinal Chemistry

ARTICLE

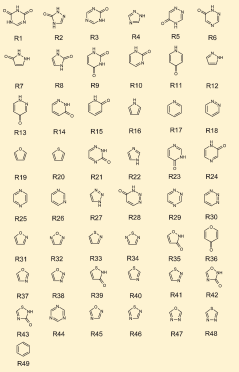
pubs.acs.org/jmc

Systematic Enumeration of Heteroaromatic Ring Systems as Reagents for Use in Medicinal Chemistry

Richard A. Ward* and Jason G. Kettle

Oncology iMED, AstraZeneca, Mereside, Alderley Park, Macclesfield, SK10 4TG, U.K.

ABSTRACT: The availability of suitable chemical building blocks, or reagents, is a key factor that determines the degree of effort required to make a target molecule. If a reagent is not available and requires synthesizing, this increases the total number of synthetic steps in the route and may result in a less attractive synthetic target. This can impact most in compound collection enhancement activities or early lead identification (LI) where typically not enough information or data are available to commit to such long multistep syntheses. In lead optimization (LO) projects, having access to commonly used reagents may improve the efficiency of building structure−activity relationships (SARs) and structure−property relationships (SPRs) around a core scaffold. This paper describes the systematic enumeration of key heteroaromatic reagent classes and the subsequent analysis of the availability of these in a number of commonly used databases.



INTRODUCTION

There are a number of publications that describe strategies for improving the quantity and or quality of compounds in corporate screening collections.[1] This is largely driven by the popular usage of high-throughput (HTS) and subset screening as hit finding tools for drug discovery projects. What has not been as widely reported are strategies to develop novel reagents or building blocks that medicinal chemists can then exploit to synthesize molecules that either were not previously accessible or were only accessible by a lengthy synthetic route.[2] Increasing the diversity of available reagents should allow access to new areas of chemical space for compound collection enhancement work and also allow for a greater efficiency of SAR exploration in early and late stage drug discovery projects. It may be argued that increasing the diversity of available reagents is a more efficient way of improving the diversity of a compound collection, as one novel reagent can be exploited in many different ways to produce a larger number of novel target compounds. As part of our current work to develop our in-house reagent database, we have recently reported an approach to identify embedded secondary amines in biologically active molecules.[3]

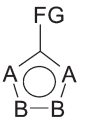
The exhaustive computational enumeration of heteroaromatic ring systems was first described by Pitt.[4] Independently we have also described our focused enumeration of ring systems around an anilinoquinazoline template.[5] Herein we report the use of a modified protocol of our algorithm BOOMSLANG[6] to exhaustively enumerated five- and six-membered aromatic rings on a number of commonly used functional groups (Figure 1).

The functional group was attached to the ring at a carbon atom. These specific groups are amenable to library synthesis and are therefore commonly used in collection enhancement work and drug discovery projects.[7] A recent publication from GlaxoSmithKline has also highlighted a range of reactions that have been demonstrated to be reliable when synthesizing chemical libraries using arrays.[8] As part of the enumeration we exhaustively include carbon, oxygen, nitrogen, and sulfur atoms along with ring carbonyl groups in five- and six-membered rings. Methyl substituents were enumerated on all C and N positions, as these groups can have a significant affect on the conformation of the substituent when attached to a core. For instance, an ortho methyl group may change the geometric preference of the substituent through steric interactions with the rest of the molecule, whereas methylation of a ring NH would remove a hydrogen bond donor, therefore changing the interactions possible for that ring system. The addition of methyl groups can also modulate properties such as lipohilicity, stability, and clearance.[9]

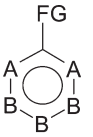
In total 50 612 potential reagent structures were enumerated across the functional groups described in Table 1. This exhaustive output was filtered down to 5759 ring systems with 443 examples per functional group using property and substructural based protocols described in the section Methods. Further analyses were carried out on the 443 enumerated ring structures, where 295 unique ring systems were identified when the functional group is removed and the substitution position is therefore disregarded. When only considering ring systems without methyl substituents, then the total number of systems enumerated after the filtering above is 48. Twenty-six of these ring systems are five-membered rings and 22 are six-membered rings. The relative frequency of these rings in a number of databases is detailed in the later section.

Received: March 23, 2011
Published: May 18, 2011

dx.doi.org/10.1021/jm200338a | J. Med. Chem. 2011, 54, 4670–4677



5-membered heterocycles



6-membered heterocycles

Label	Atoms Enumerated[b]
FG[a]	See Table 1
A	c, c(C), n, n(H), n(C), o, s
B	c, c(C), n, n(H), n(C), c(=O), o, s

Figure 1. Schematic of ring enumeration scheme: [a]FG denotes a functional (reactive) group, and these are exemplified in Table 1; [b]atoms in parentheses denote substitutions on the ring.

RESULTS AND DISCUSSION

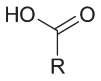
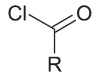
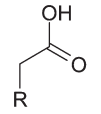
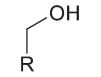
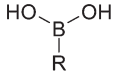
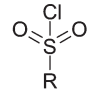
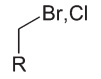
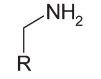
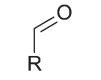
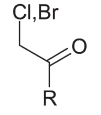
Analysis of Enumerated Reagents in the Public Domain. These 5759 potential reagents were searched for, as exact matches, in the Available Chemicals Directory[10] (ACD) to identify reagents that could be readily purchased (Table 2). We also wanted to evaluate which of these potential reagents had been reported in the wider literature, which could reveal a synthetic route or further give confidence that they could be synthesized successfully. We therefore searched for matches within an internally curated database of over 34 million structures that contains a number of published small molecule databases, including ACD,[10] PubChem,[11] ChemSpider,[12] eMolecules,[13] Prous Integrity,[14] ChEMBL,[15] DrugBank,[16] GVKBio,[17] and our internal compound and reagent collections. Any reagent that does not have a match in any of these databases is identified in Table 1 as a novel reagent. A previously published in-house algorithm[18] was used for the exact match searches.

What is evident from this analysis is that the coverage of these ring systems in databases such as ACD varies widely across the reagent types. Heteroaromatic acids and heteroaromatic amines for example are well populated in ACD (43% and 39%) of enumerated ring systems have exact matches respectively), whereas the heteroaromatic haloacetophenones (2% and 5%) and heteroaromatic sulfonyl chlorides (9%) are less frequent. These observations may highlight areas of significant opportunity for design and synthesis of novel reagents but in many examples may indicate incompatibility between functional (reactive) group and heterocyclic ring system. As would be expected, the abundance of the reagents in ACD is, in general, inversely related to the number of novel reagents for a particular reagent type when searched across all databases (Figure 2).

In terms of total ring system coverage of the enumeration scheme, 20% of the enumerated structures are present in ACD and 61% are novel across the databases described. This set of novels, consisting of 3511 potential reagents, represents a significant gap in a relatively simple but potentially useful reagent set. The analysis also identifies that 19% of the enumerated reagents are not available in ACD but have evidence that they can be made through having a match in one of the described small molecule databases. It is likely that a synthetic route has been published for a number of these reagents, along with other examples that could be purchased from alternative suppliers.

Prioritization of Novel Reagents for Internal Synthesis. Reagents that were present in ACD but not our in-house reagent database were evaluated by medicinal chemists, and a number of examples were purchased to supplement our collection. The in-house reagent database referred to is a proprietary collection of reagents available to chemists at AstraZeneca. Additionally, reagents that were found in one of the small molecule databases detailed but are not available in ACD were assessed and a number of these have been synthesized and made available for use across medicinal chemistry projects. For the remaining novel ring systems these were further prioritized by medicinal chemists and sets of reagents were identified for synthesis. This selection

Table 1. Functional Groups Enumerated on Five- and Six-Membered Aromatic Ring Templates within This Study

Reagent Identifier	Reagent Name	Schematic[a]
A	Heteroaromatic amines	
B	Heteroaromatic acids	
C	Heteroaromatic acid chlorides	
D	Heteroaromatic acetic acids	
E	Heteroaromatic methyl alcohols	
F	Heteroaromatic boronic acids	
G	Heteroaromatic sulphonyl chlorides	
H	Heteroaromatic methyl halides	
I	Heteroaromatic methyl amines	
J	Heteroaromatic aldehydes	
K	Heteroaromatic Haloacetophenones	

[a] R is the enumerated five- and six-membered heterocyclic ring as described in Figure 1.

dx.doi.org/10.1021/jm200338a |J. Med. Chem. 2011, 54, 4670–4677

Table 2. Exact Matches of the Enumerated Reagents in ACD and the Proportion of Novel Reagents within Each Reagent Type

reagent identifier	reagent name	enumerated aromatic rings	ACD matches	ACD frequency,[a] %	novel reagents	novelty frequency,[b] %
A	heteroaromatic amines	443	172	39	119	27
B	heteroaromatic acids	443	192	43	143	32
C	heteroaromatic acid chlorides	443	55	12	300	68
D	heteroaromatic acetic acids	443	80	18	271	61
E	heteroaromatic methyl alcohols	443	110	25	216	49
F	heteroaromatic boronic acids	443	56	13	348	79
G	heteroaromatic sulfonyl chlorides	443	38	9	341	77
H1	heteroaromatic methyl halides (Cl)	443	91	21	246	56
H2	heteroaromatic methyl halides (Br)	443	44	10	312	70
I	heteroaromatic methyl amines	443	126	28	247	56
J	heteroaromatic aldehydes	443	171	39	204	46
K1	heteroaromatic haloacetophenones (Cl)	443	8	2	397	90
K2	heteroaromatic haloacetophenones (Br)	443	20	5	367	83
total		5759	1163	20	3511	61

[a] ACD frequency is the percentage of reagents matched in ACD. [b] Novelty frequency shows the percentage of novel ring systems.

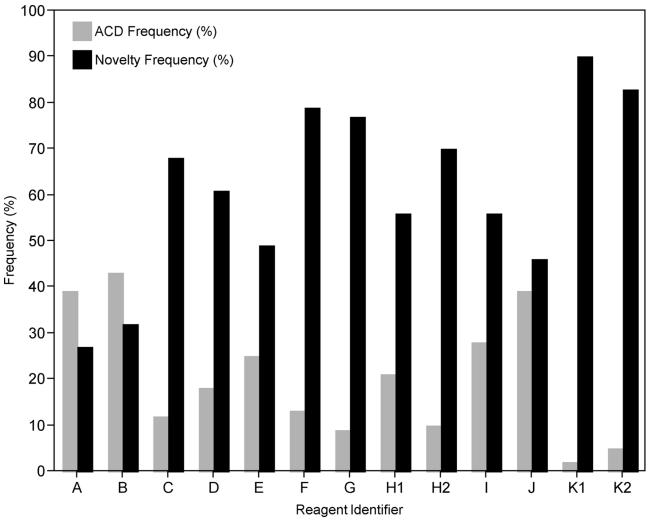


Figure 2. Comparison of the frequency of enumerated reagents in ACD (gray) within each reagent class against the frequency that were identified as being novel (black).

step also provides an opportunity to ensure that redundant reagents are not commissioned for synthesis, such as bromo- and chloroacetophonone equivalents of reagent. Further prioritization can also be made at this stage if, for example, a proposed reagent does not appear to offer significant scope for the synthesis of novel compounds beyond what could be made using commercially available reagents of other reagent classes. Table 3 shows the breakdown of the enumerated rings systems against our in-house reagent database. Also shown are the number of reagents across these classes that have had their syntheses proposed and the number of reagents subsequently delivered by synthesis at an external contract research organization. Any requested reagents submitted for synthesis would not have previously been present in our in-house reagent database and would not have been available in ACD.

Although the additional reagents now available within this enumeration scheme is relatively small when compared to the total output of the enumeration (2%), the examples synthesized have added 97 reagents out of the 653 currently available to the chemist within this reagent set (15%). Out of these 97 reagents already delivered, 8 examples (8%) are classified as novel and not present in any of the described small molecule databases at the time of synthesis. A further observation is that 97 out of the 709 currently proposed reagent structures have been delivered so far through this process (14%). There may be several reasons why reagents have not been delivered into the in-house collection, for example, not being chemically accessible or stable or simply that it was deprioritized by a team of medicinal chemists. However, a number of these proposed reagents will still be working their way through the evaluation stage or scoping out of the synthesis. This analysis also highlights the popular reagent classes that are commonly used in drug discovery projects and ones in which medicinal and synthetic chemists have commonly requested for synthesis. In particular these are heteroaromatic amines (14%), heteroaromatic acids (15%), heteroaromatic acetic acids (21%), heteroaromatic methylamines (18%), and heteroaromatic aldehydes (15%). Focusing our efforts toward functional groups that are most commonly requested is also helpful to ensure that reagents delivered are regularly used in drug discovery projects.

Examples of Reagents Delivered at AstraZeneca. Figure 3 highlights four reagents that were identified when this set of ring systems was first enumerated (September 2005). These are from the heteroaromatic amine and heteroaromatic acid reagent types. When these rings systems were searched for in ACD, they were not available. These were then synthesized and added to our internal reagent collection. Since then, these four reagents have become available in ACD.

Figure 4 shows additional ring systems that were synthesized early on as part of this study. These were not present in ACD when synthesized, and at the time of writing this manuscript they are still not available in ACD. These reagents and others from this work have become routinely used in drug discovery projects, and some of these have already been remade because of demand.

Frequency Analysis of Enumerated Rings in Commercial Databases. To further identify ring systems that could be exploited through the synthesis of new reagents, we looked at the frequency of the enumerated rings, as substructures, across a number of databases. The frequency of the ring systems were

Table 3. Summary of Enumerated Ring Systems and the Exact Matches Present across Internal Databases

reagent identifier	reagent name	enumerated aromatic rings	matches in internal database[a]	reagents proposed in-house[b]	reagents delivered in-house[c]
A	heteroaromatic amines	443	130	101	6
B	heteroaromatic acids	443	138	103	25
C	heteroaromatic acid chlorides	443	23	0	0
D	heteroaromatic acetic acids	443	41	148	17
E	heteroaromatic methyl alcohols	443	56	12	1
F	heteroaromatic boronic acids	443	36	76	17
G	heteroaromatic sulfonyl chlorides	443	23	26	3
H1	heteroaromatic methyl halides (Cl)	443	26	7	0
H2	heteroaromatic methyl halides (Br)	443	12	6	2
I	heteroaromatic methylamines	443	85	126	19
J	heteroaromatic aldehydes	443	73	103	7
K1	heteroaromatic haloacetophenones (Cl)	443	2	1	0
K2	heteroaromatic haloacetophenones (Br)	443	8	0	0
total		5759	653	709	97

[a] Number of matches in our internal reagent collection. [b] Number of reagents from that reagent type which have been proposed for synthesis. [c] Number of reagents synthesized and delivered to AstraZeneca from each reagent class.

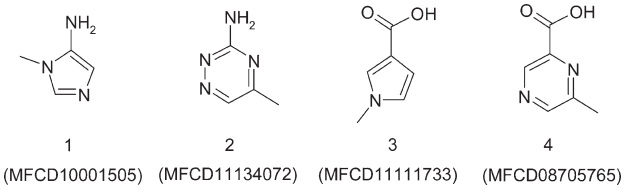


Figure 3. Examples identified from reagent enumeration that have subsequently become available in ACD.

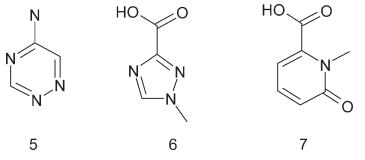


Figure 4. Selected examples of reagents synthesized from reagent enumeration analysis that are not available in ACD.

assessed specifically in ACD and DrugBank (small molecule version) but also our internally curated database of over 34 million structures using in-house software (Figure 5).

Note that the functional group was removed and only ring systems without methyl substituents were used for this substructural search, as examples of those ring systems with methyl atom(s) attached will be picked up in the analysis. No information on the position of connectivity to the functional group was maintained. The analysis shows that all of these 48 ring systems have been reported in the public domain in one or more of these databases (Table 4). For comparison purposes phenyl has been added to this analysis (ring R49), using the search restrictions described in the section Methods to avoid additional fused rings.

The frequency of each of these rings systems in ACD gives an indication of the number of reagents that may be commercially available. However, clearly not all compounds identified in ACD with a particular ring system will be useful reagents, but this analysis does provide some insight into reagent availability across the cores. The broader search for the presence of these rings systems in our combined set of databases may allow information to be gathered from references with synthetic routes. It is worth highlighting that if a molecule is present in multiple databases within the combined set, such as a molecule that is present in GVKBio and Pubchem, this has only been counted as one match. If a ring system was present multiple times in the same molecule, that was also counted as a single match. Although the enumerated ring systems in this study can all be found as embedded substructures in these database, the earlier analyses in this paper have demonstrated that there are a significant number of gaps in terms of the availability of these rings with preferred functional groups at the various possible substitution positions. Further observations can be made from this analysis, such as ring systems that are found with high frequencies in reagent databases such as ACD but exemplified infrequently or not at all in DrugBank. There could be several explanations for these observations, such as lack of the specific key reagents, but it could also be due to the ring system having a particular issue making it unsuitable as part of a drug molecule (e.g., stability issues or potential toxicity). Examples of such ring systems that appear underrepresented in DrugBank include R24, R32, R33, and R46. An alternative way of using this information is to highlight ring systems such as R35. Although there are only 77 matches for this substructure in ACD and the substructure is relatively infrequent across all databases (4151 matches), the ring is found in five molecules in DrugBank. This is an example of a ring system where it may be prudent to consider the synthesis of key reagent sets. Finally, ring systems such as R30, R28, and R42 have very few examples in ACD and across the databases in general; however, there are some matches that may indicate that synthetic routes exist to allow the population of reagents containing these rings. Broadening the availability of reagents containing these ring systems may then lead to their increased usage in drug discovery projects.

METHODS

Enumeration Protocol. The ring systems were enumerated using BOOMSLANG. In total 50 612 ring systems with unique

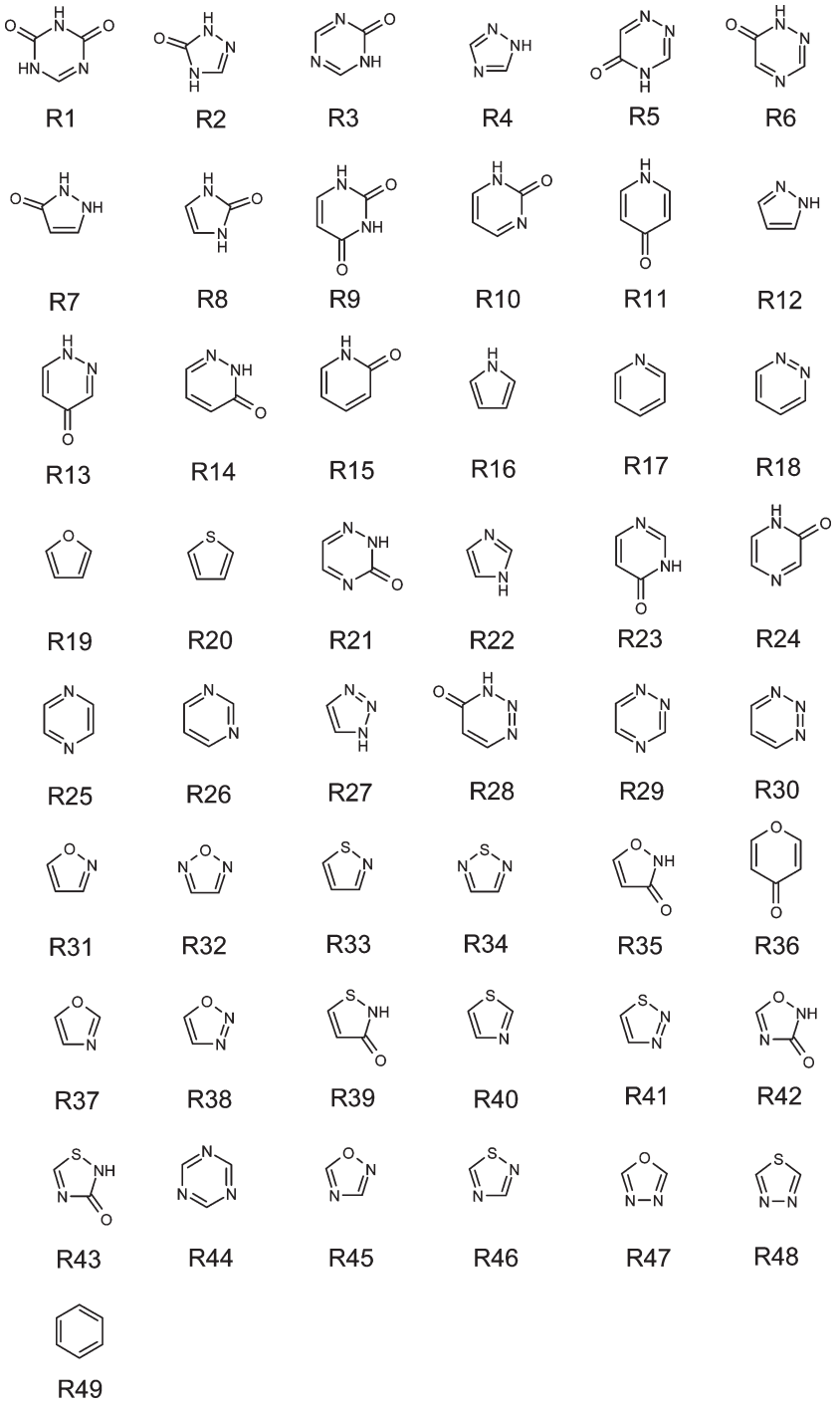


Figure 5. Enumerated ring systems with no methyl substituents attached and functional groups removed. The protonation pattern of each ring shown is a selected structure and does not represent a preferred form. The phenyl ring, R49, has been added as a reference.

SMILES[19] were generated across the 11 described reagent classes. The generated SMILES (that included all possible tautomers, protonation states, and methyl substituent positions) were then canonicalized[19] and exact matches resulting from symmetry and hydrogen placement removed by text matching. Only aromatic rings were required for this analysis, so ring systems that were not aromatic, using the Openeye definition of aromaticity,[20] were also removed. The heteroaromatic methyl halides (**H**) and heteroaromatic haloacetophenones (**K**) were enumerated with both bromine and chlorine, since both are equally useful and one or both may be commercially available. The focus of this work was to identify rings that were less lipophilic than phenyl; therefore, the ClogP (version 4.3) of the enumerated rings was calculated without the functional group being present. We then removed rings with a ClogP > 2. For comparison, the ClogP of phenyl is 2.14. In a previous report of our in-house strategic reagent design[3] we have focused on groups where the added substituent has a molecular weight of below 200; however, all these enumerated rings fell within this criteria. A set of SMARTS[21] based substructural filters was used to

dx.doi.org/10.1021/jm200338a |*J. Med. Chem.* 2011, 54, 4670–4677

Table 4. Frequency Analysis of 48 Enumerated Core Ring Systems across a Range of Compound Databases[a]

ring	ring SMILES	ACD matches[b]	all database matches[c]	DrugBank matches[d]
R49	**c1ccccc1**	**716347**	**24687578**	**1740**
R17	c1ccncc1	80947	2630314	185
R9	c1c[nH]c(=O)[nH]c1=O	3092	156158	106
R22	c1cnc[nH]1	14786	700244	82
R40	c1cscn1	27995	1033334	65
R26	c1cncnc1	20833	901864	62
R20	c1ccsc1	39538	1381671	54
R10	c1c[nH]c(=O)nc1	1536	41246	41
R12	c1c[nH]nc1	39429	1310596	24
R31	c1cnoc1	7158	351432	19
R19	c1ccoc1	25630	1071853	18
R4	c1[nH]ncn1	17289	736104	17
R16	c1cc[nH]c1	12188	473140	12
R25	c1cnccn1	2437	181802	12
R15	c1cc[nH]c(=O)c1	4301	165734	9
R27	c1cnn[nH]1	3348	130690	9
R44	c1ncncn1	1437	117372	8
R23	c1cnc[nH]c1=O	3650	117159	7
R7	c1c[nH][nH]c1=O	1179	67503	7
R35	c1co[nH]c1=O	77	4151	5
R37	c1cocn1	2137	189978	4
R2	c1[nH]c(=O)[nH]n1	583	34210	4
R45	c1ncon1	5091	244804	3
R48	c1nncs1	3558	240156	3
R47	c1nnco1	5760	215111	3
R18	c1ccnnc1	1718	114637	2
R11	c1c[nH]ccc1=O	413	23708	2
R3	c1[nH]c(=O)ncn1	63	3905	2
R14	c1cc(=O)[nH]nc1	1355	91963	1
R41	c1csnn1	1093	29100	1
R8	c1c[nH]c(=O)[nH]1	163	19361	1
R29	c1cnncn1	767	14986	1
R36	c1coccc1=O	147	14117	1
R34	c1cnsn1	57	11272	1
R39	c1cs[nH]c1=O	34	3085	1
R1	c1[nH]c(=O)[nH]c(=O)n1	45	2633	1
R32	c1cnon1	1206	31229	0
R46	c1ncsn1	287	30987	0
R33	c1cnsc1	401	21336	0
R24	c1cncc(=O)[nH]1	130	20196	0
R5	c1c(=O)[nH]cnn1	954	13268	0
R13	c1c[nH]ncc1=O	122	7020	0
R38	c1conn1	86	3085	0
R21	c1cn[nH]c(=O)n1	277	2811	0
R30	c1cnnnc1	0	1421	0
R6	c1c(=O)[nH]ncn1	31	732	0
R43	c1nc(=O)[nH]s1	27	259	0
R42	c1nc(=O)[nH]o1	1	62	0
R28	c1cnn[nH]c1=O	1	54	0

[a] The data for phenyl (R49) has been added and is shown in bold. The table results are ordered by number of hits in DrugBank. [b] Number of ring matches in ACD. [c] Number of ring matches in our internal curated small molecule database of 35 million structures. [d] Number of ring matches in DrugBank (small molecule version).

remove ring systems with unwanted connection patterns such as sulfur to sulfur and sulfur to oxygen bonds. All six-membered rings containing sulfur were removed, along with six-membered rings containing oxygen where there were additional heteroatoms in the ring (allowing, for example, pyranones but no azapyranones). Ring esters and thioesters (where the enumeration placed a carbonyl group adjacent to either ring oxygen or sulfur) were removed because of general concerns around stability. Ring carbonyl groups were not enumerated on the 2-position of a five-membered ring or the ortho position on a six-membered ring to minimize interference with the properties of the functional group. Methyl groups were enumerated in all available positions on ring carbon and nitrogen atoms, although examples with more than two methyl substituents around the ring were subsequently removed because of concerns over potentially increased metabolic liability. Finally, any ring system with four or more heteroatoms within the enumerated ring was removed because of potential issues with ring stability. In this procedure a ring carbonyl or the functional group itself was not considered as part of the ring. This step may remove some reasonable ring systems (such as tetrazole); however, it is an efficient way of removing significant numbers of likely unstable ring systems to prioritize examples of most interest. Although the physicochemical properties of the filtered rings are broadly similar, the positioning and relative strengths of the hydrogen bonding groups offer significant diversity in this set. Addition of methyl substituents also increases the diversity in the ring shape and conformation when added onto a template as a substituent. These searches produced a final set of 5759 ring systems with 443 examples per functional group. Note that no specific assessment was made as to whether all these ring systems would be synthetically accessible or stable beyond the filtering protocol described. Within a specific reagent set of 443 enumerated reagent structures, there are 295 unique ring systems when the functional group is removed, and the substitution position is therefore not considered. If you take only those ring systems without methyl substituents, then the total number of ring systems enumerated after the filtering above is 48. Twenty-six of these ring systems are five-membered rings and 22 are six-membered rings.

Frequency Searching of Enumerated Cores. With the in-house software,[22] each of the 48 identified ring systems were searched for by using the ring SMILES notation as a substructural query (i.e., SMARTS), where each ring atom was designated with two ring connections (i.e., "R2" using Openeye definitions[23]). This avoids ring systems being matched which are fused onto other ring systems. All hydrogen atoms attached to nitrogen (represented as [nH]) were removed in the SMARTS query, thereby allowing molecules substituted at any nitrogen atom to be found as matches. A further modification required was to ensure that all ring carbons (which were not enumerated as part of a carbonyl group) in the query would not hit molecules in the database that had a ring carbonyl in this position. Therefore, all ring carbons not in a carbonyl group were coded by SMARTS not to hit a molecule in the database if the atom was double bonded to an aliphatic atom outside the ring. This process is summarized in Figure 6, using the pyrimidine ring system (R26) as an example. This also avoids ring systems such as imines being found as hits when searching for the presence of these enumerated ring systems in a database.

The final modification required to search accurately for the relevant substructures was to explicitly cover a ring carbonyl along with the hydroxyl tautomer equivalent, as molecules could

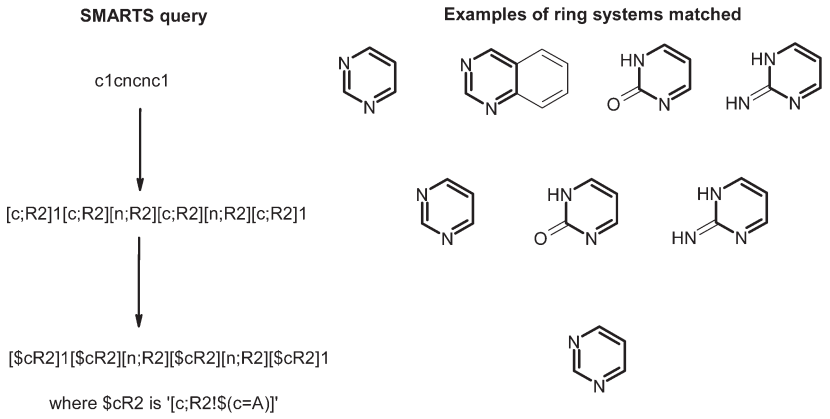


Figure 6. Exemplified SMARTS string used to specifically encode the pyrimidine ring system as part of the ring searching protocol (R26).

be represented in a database in either form. We therefore modified the SMARTS queries to not only match the ring carbonyl but also the neutral and deprotonated hydroxyl forms. We did not consider molecules with further substituents on the oxygen as matches for the ring systems.

CONCLUSIONS

This work details a method to systematically enumerate sets of target reagents. These proposed reagents were then matched against what is commercially available or reported in the literature to highlight key gaps of interest in reagent chemistry. These types of analyses can be used to increase the diversity of available reagents to allow more rapid and efficient SAR exploitation and access new compounds for collection enhancement work. From this study a number of reagents were purchased from ACD, and reagents published in the literature but not in ACD were synthesized where it was felt they were useful additions. Perhaps the most valuable reagents targeted here were reagents that were not known in either ACD or the literature but from subsequent evaluation and prioritization have been synthesized and registered in our in-house reagent database. Many of these reagents have already been used in drug discovery projects and collection enhancement work. Their ability to modulate and reduce the lipophilicity of compounds in a proposed design set makes them of particular value to medicinal chemists. Furthermore, these ring systems were searched for as substructures in reagent focused databases such as ACD but also databases of small molecules that can identify a synthetic route being discovered. This analysis has also been used to identify ring systems present in known drug molecules but that are poorly exemplified in reagent databases with the aim of increasing the reagent coverage in those areas.

AUTHOR INFORMATION

Corresponding Author

*Phone: +44(0)1625 519045. E-mail: Richard.A.Ward@AstraZeneca.com.

ACKNOWLEDGMENT

We thank the colleagues at AstraZeneca who have been involved in this work, especially Andrew Poirrette and Sorel Muresan for curation of the internal databases.

ABBREVIATIONS USED

LI, lead identification; LO, lead optimization; SAR, structure−activity relationship; SPR, structure−property relationship; HTS, high-throughput screening; ACD, Available Chemicals Directory

REFERENCES

(1) (a) Jacoby, E.; Schuffenhauer, A.; Popov, M.; Azzaoui, K.; Havill, B.; Schopfer, U.; Engeloch, C.; Stanek, J.; Acklin, P.; Rigollier, P.; Stoll, F.; Koch, G.; Meier, P.; Orain, D.; Giger, R.; Hinrichs, J.; Malagu, K.; Zimmermann, J.; Roth, H.-J. Key aspects of the Novartis compound collection enhancement project for the compilation of a comprehensive chemogenomics drug discovery screening collection. *Curr. Top. Med. Chem.* **2005**, *5*, 397–411. (b) Schuffenhauer, A.; Popov, M.; Schopfer, U.; Acklin, P.; Stanek, J.; Jacoby, E. Molecular diversity management strategies for building and enhancement of diverse and focused lead discovery compound screening collections. *Comb. Chem. High Throughput Screening* **2004**, *7*, 771–781. (c) Drewry, D. H.; Macarron, R. Enhancements of screening collections to address areas of unmet medical need: an industry perspective. *Curr. Opin. Chem. Biol.* **2010**, *14* (3), 289–298.

(2) (a) Wang, J.; Hou, T. Drug and drug candidate building block analysis. *J. Chem. Inf. Model.* **2010**, *50*, 55–67. (b) Welsch, M. E.; Snyder, S. A.; Stockwell, B. R. Privileged scaffolds for library design and drug discovery. *Curr. Opin. Chem. Biol.* **2010**, *14* (3), 347–361. (c) Lewell, X. Q.; Judd, D. B.; Watson, S. P.; Hann, M. M. RECAP—retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 511–522.

(3) Kettle, J. G.; Ward, R. A.; Griffen, E. Data-mining patent literature for novel chemical reagents for use in medicinal chemistry design. *Med. Chem. Commun.* **2010**, *1*, 331–338.

(4) Pitt, W. R.; Parry, D. M.; Perry, B. G.; Groom, C. R. Heteroaromatic rings of the future. *J. Med. Chem.* **2009**, *52* (9), 2952–2963.

(5) Kettle, J. G.; Ward, R. A. Towards the comprehensive systematic enumeration and synthesis of novel kinase inhibitors based on a 4-anilinoquinazoline binding mode. *J. Chem. Inf. Model.* **2010**, *50* (4), 525–533.

(6) Cosgrove, D. A.; Kenny, P. W. BOOMSLANG: a program for combinatorial structure generation. *J. Mol Graphics* **1996**, *14* (2), 1–5.

(7) 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**, *4*, 2336–2347.

(8) Cooper, T. W. J.; Campbell, I. B.; Macdonald, S. J. F. *Angew. Chem., Int. Ed.* **2010**, *49*, 8082–8091.

(9) Leach, G. A.; Jones, H. D.; Cosgrove, D. A.; Kenny, P. W.; Ruston, L; MacFaul, P. J.; Wood, M.; Colclough, N.; Law, B. Matched molecular pairs as a guide in the optimization of pharmaceutical

properties: a study of aqueous solubility, plasma protein binding and oral exposure. *J. Med. Chem.* **2006**, *49* (23), 6672–6682.

(10) Symyx. Available Chemicals Directory. Version 2010.02. http://www.symyx.com/products/databases/sourcing/acd/index.jsp (accessed July 1, 2010).

(11) National Centre For Biotechnology Information. PubChem. http://pubchem.ncbi.nlm.nih.gov (accessed Mar 4, 2011).

(12) ChemSpider. http://www.chemspider.com (accessed Mar 4, 2011).

(13) eMolecules. http://www.emolecules.com (accessed Mar 4, 2011).

(14) Prous Integrity. http://thomsonreuters.com/products_services/science/science_products/a-z/integrity (accessed Mar 4, 2011).

(15) Warr, W. A. ChEMBL. An interview with John Overington, team leader, chemogenomics at the European Bioinformatics Institute Outstation of the European Molecular Biology Laboratory (EMBL-EBI). *J. Comput.-Aided Mol. Des.* **2009**, *23* (4), 195–198.

(16) Wishart, D. S.; Knox, C.; Guo, A. C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* **2008**, *36*, D901–D906 (Small Molecule Structures, version 2.5).

(17) GVKBio Databases. http://www.gvkbio.com (accessed Mar 4, 2011).

(18) Blomberg, N.; Cosgrove, D. A.; Kenny, P. W.; Kolmodin, K. Design of compound libraries for fragment screening. *J. Comput.-Aided Mol. Des.* **2009**, *23* (8), 513–525.

(19) Weininger, D.; Weininger, A.; Weininger, J. L. SMILES. 2. Algorithm for generation of unique SMILES notation. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 97–101.

(20) Openeye Scientific Software. OEChem Toolkit (version 1.7.0). http://www.eyesopen.com/docs/docs-v1.7.0-2/html/OEChemTK-python/aromaticity.html (accessed Nov 1, 2010).

(21) Daylight Chemical Information Systems, Inc. Daylight Theory Manual. http://www.daylight.com/dayhtml/doc/theory/index.html (accessed Nov 1, 2010).

(22) In-house software developed by Dave A. Cosgrove (unpublished results).

(23) Openeye Scientific Software. OEChem Toolkit (version 1.7.0). http://www.eyesopen.com/docs/docs-v1.7.0-2/html/OEChemTK-python/ring.html (accessed Nov 1, 2010).