ELSEVIER

Teaser The rational design of novel, high quality building blocks can accelerate drug discovery projects and improve compound quality, but has been overlooked in the medicinal chemistry literature.

Designing novel building blocks is an overlooked strategy to improve compound quality

Frederick W. Goldberg[1], Jason G. Kettle[1], Thierry Kogej[2], Matthew W.D. Perry[2] and Nick P. Tomkinson[1]

[1] AstraZeneca, Alderley Park, Macclesfield SK10 4TG, UK
[2] AstraZeneca, SE-43183 Mölndal, Sweden

One pragmatic way to improve compound quality, while enhancing and accelerating drug discovery projects, is the ability to access a high quality, novel, diverse building block collection. Here, we outline general principles that should be applied to ensure that a building block collection has the greatest impact on drug discovery projects, by discussing design principles for novel reagents and what types of reagents are popular with medicinal chemists in general. We initiated a program in 2009 to address this, which has already delivered three candidate drugs, and the success of that program provides evidence that focussing on building block design is a useful strategy for drug discovery.

Frederick Goldberg is a medicinal chemist in the oncology department at AstraZeneca. He received his MSc from Cambridge University in 2000 and his PhD from Imperial College London in 2004, then received a Fulbright scholarship to perform postdoctoral studies with Philip Magnus at the University of Texas at Austin before returning to the UK in 2006 to work for AstraZeneca. He has worked across all discovery phases in diabetes and oncology disease areas, and his research interests include kinases, structure-based design, compound collection enhancement, chemical biology, and metabolism. In 2013, he was awarded the EFMC Prize for a Young Medicinal Chemist in Industry.

Introduction

Numerous analyses have been published on the importance of 'compound quality' or 'drug-likeness', both in the context of high-quality screening collections to improve success rates of high-throughput screening (HTS) and the quality of the resulting hits, and in the context of the developability of candidate drugs. Although compound quality is loosely defined [1,2], these analyses have typically focussed on factors such as lead-like properties [3], more general consideration of lipophilicity and other physicochemical properties [4–7], diversity [8–10], novel or diverse coverage of chemical space [11,12], privileged structures for drugs [13], or structures that have favourable physical properties or metabolic stability [14,15]. One pragmatic way to improve the quality of both candidate drugs and screening collections is by improving the quality of the building blocks (reagents) that are used to synthesise them. Although this is widely recognised among medicinal chemists, access to diverse, high quality reagent sets and the design principles that should govern both strategic acquisitions and custom synthesis of such building blocks have been rather overlooked in the medicinal chemistry literature. Reagents that are chosen to be medicinally relevant and designed to impart favourable physical properties when used to synthesise project compounds can not only accelerate drug discovery programmes, by avoiding lengthy syntheses, but also improve the quality of the designed molecules, by focussing on substructures and properties known to have imparted biological activity and good 'drug-like' properties in the past. One approach is to purchase these reagents from commercial suppliers,

Corresponding author: Goldberg, F.W. (frederick.goldberg@astrazeneca.com)

1359-6446/06/$ - see front matter http://dx.doi.org/10.1016/j.drudis.2014.09.023

GLOSSARY

Available Chemicals Directory (ACD) a regularly updated list of commercially available chemicals from diverse chemical suppliers
Electronic Lab Notebook (ELN) a searchable electronic repository and database of synthetic reactions that have been performed
Fsp^3 fraction of carbons in the molecule that are sp^3 hybridised, which therefore measures the degree of saturation in a given molecule
Simplified molecular-input line-entry system (SMILES) a method of representing a molecular structure with a text string
Smiles arbitrary target specification (SMARTS) a method of encoding substructures using SMILES notation (http://www.daylight.com/dayhtml/doc/theory/)

particularly because the availability and breadth of both Available Chemicals Directory (ACD; see Glossary) listed reagents and non-ACD commercial reagents is continually improving. However, it is still our experience that restricting drug discovery programs to readily available commercial reagents does not provide a sufficiently thorough structure–activity relationship (SAR), neither does it provide sufficient access to innovative, novel structures that could provide the step-jump improvement in properties that is often required to deliver a candidate drug or *in vivo* tool compound.

Similar arguments can be made when considering the requirements of an ideal set of building blocks for enhancing a screening compound collection by synthesising novel libraries, where novelty and efficient, diverse scoping of chemical space are desirable goals. In addition, access to novel building blocks can provide a competitive advantage, by creating compound collections and project compounds that are differentiated from those synthesised by competitors, with potential intellectual property advantages.

Within AstraZeneca (AZ), we have attempted to address this with custom synthesis of novel building blocks, where novelty in this context is defined as not in ACD. In 2009, we launched a 'strategic reagent initiative' (SRI), to harness the learning from our internal program and external medicinal chemistry literature [16]. This initiative has proved to be successful, as judged by the widespread impact of the project and direct incorporation of SRI reagents into three candidate drugs, and has led us to conclude that this is an important method to enhance compound quality in drug discovery.

Here, we discuss our conclusions from this program, using a data set of 3044 reagents delivered on (typically) 20 g scale, with a CRO partner (WuXi AppTec) from 2009 to 2012. We highlight methods that can be used to generate such reagents, the desirable properties of such a building block collection, and what types of reagent have proven to be popular with medicinal chemists, by analysing the contribution of these reagents to our corporate compound collection and our internal electronic lab notebook of 690,000 reactions performed by our chemistry teams.

Design strategies and guidelines, the 'rule of two'

A successful building block program requires a broad input of ideas from diverse sources, and ideally uses a large number of experienced medicinal chemists to generate those ideas. To achieve this, we set up a global team that represented AZ discovery chemistry groups in UK, US, Sweden, France, and India. Examples of two of the methods we have used have already been published, namely

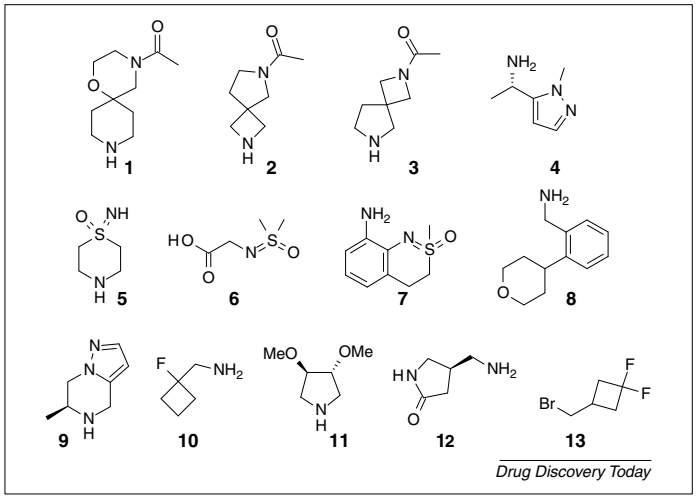

FIGURE 1

Examples of novel (defined as non-ACD) designed reagents that have been incorporated into final screening compounds in the AstraZeneca compound collection.

systematic enumeration of aromatic rings [17] and data mining of the patent literature [18]. Other approaches we took included saturated and chiral [19] reagents to 'escape from flatland' [20] (e.g. spirocyclic examples **1**–**3**, Fig. 1), key motifs of utility for medicinal chemistry (e.g. alpha-methyl benzylamines and heteroaromatic analogues, such as **4**), bioisosteric groups of commonly used functional groups in medicinal chemistry [21] (e.g. sulfoximines **5**–**7**, designed as isosteres of sulfones or sulfonamides with potential to improve solubility) [22,23], SAR sets of common functional groups enumerated on five- and six-membered aromatic rings (e.g. THP-substituted phenyl **8**) and inspiration from internal projects (e.g. **9**). Pragmatically, we also chose to mine areas that we knew from usage statistics to be popular, such as appropriately designed amines to modulate basicity or log *D* (e.g. **10** and **11**), groups that have the potential to lower log *D* and improve metabolic stability relative to the analogous cycloalkane (e.g. **12** and **13**), and common substructures in known drugs and bioactive compounds [24]. All of these reagents have been incorporated into final compounds by drug discovery projects within AZ. For example, **12** has been incorporated into 209 final compounds in the AZ compound collection, including a disclosed example in a recent 11β-hydroxysteroid dehydrogenase type 1 (11β-HSD1) publication [25], and **13** has been incorporated into 59 final compounds, including a disclosed example in a recent γ-secretase publication [26]. When they were synthesised, none of the structures in Fig. 1 was listed in ACD, although **12** and **13** have since been added. Not being listed in ACD does not necessarily mean that reagents are not commercially available, because many commercial suppliers prefer to not list their building blocks in ACD. However, many of the reagents that we have synthesised are not, to the best of our knowledge, commercially available outside of ACD either, and many can also be considered to be novel, as judged by a SciFinder search. In addition, it is our experience that reagents that are not listed in ACD, but are nevertheless commercially available, are often only available at high cost and with relatively long lead times.

With such diverse methods of producing ideas, we generated several times more targets than we could synthesise, so prioritisation of the targets became important. This was primarily done by visual inspection of every structure by experienced medicinal chemists to assess their general desirability and broad applicability. To achieve this, we experimented with different models to achieve the right balance of conservative and innovative structures. A simple voting system among the global team was used initially, although this had a tendency to filter out the more innovative structures, because unusual structures can generate divergent opinions among the team. This voting system was then replaced with a 'design team' model, where the design lead was in control of the final prioritisation, in consultation with the views of the team. Reagents inspired by a specific project were only prioritised if it was felt that there was likelihood that the reagent could have a broader utility across projects. Projects often have particular requirements for building blocks; for example, activity in that project might depend on a particular warhead, an unusual mechanism, or an unusual route of administration. In that case, the project team should develop their own project-specific set of building blocks as a complementary activity to the more general approach advocated here.

To aid the prioritisation process, we developed a set of guidelines to provide more concrete guidance to the chemistry team. An analysis was performed on the reagents that had been used for library synthesis under collaboration with ChemBridge Corporation. This analysis showed that molecular weight (MW) and clogP were important factors in the frequency of use of reagents. Other parameters, such as PSA, hydrogen bond acceptor (HBA) count, hydrogen bond donor (HBD) count, and rotatable bond count, were less important. From this analysis, we defined a simple guideline that popular reagents could be defined as those that typically do not add more than 200 Da in MW or 2 units of clogP. We also aimed for a HBD count $\leq$2 and HBA count $\leq$4, which in combination with the MW and clogP guidelines, gave a 'rule of 2' mnemonic (Table 1), to be applied as a guideline rather than as a strict cut-off. This 'rule of 2' can be compared to the 'rule of 5' for orally administered drugs [27], the 'rule of 4' that has been proposed for lead-like properties [28], and the 'rule of 3' that has been proposed to guide selection of fragments for fragment-based lead generation [29], in that it is in part a consequence of our desire as medicinal chemists to design candidate drugs with physical properties that are suitable for oral administration.

There is an additional driver in this case, that smaller, simpler reagents are also more efficient in scoping out the SAR on a project scaffold. A typical parallel synthesis run on a project scaffold of 50–100 final compounds necessitates the selection of relatively simple reagents, otherwise the SAR from the library will be difficult to interpret. These general principles are in agreement with conclusions published by a group at GlaxoSmithKline (GSK) [30], where they introduced the concept of 'lead-oriented synthesis'. In their review, the authors discussed the need to apply and develop suitable synthetic methodologies to control physicochemical properties, and also presented data that showed that there was a log *P* drift in parallel (library) chemistry. This drift arises because the more lipophilic compounds in the library have a higher chance of being successfully synthesised (and isolated) and, thus, the compounds delivered are more lipophilic on average than intended. Although the authors in that article focussed on the properties of final compounds and general synthetic methodology, rather than building blocks, there is some degree of consensus between their conclusions and our approach. The importance of designing lead-like compounds in library design is also addressed by a previous AZ paper from a group at Charnwood, UK [8].

Other design guidelines are as listed in Table 1. The preference of one commonly used reactive group was not strictly applied, but our view is that capping groups are more likely to have broad generic utility across projects and target classes, whereas bifunctional reagents that are often used as cores can be more target class specific. However, bifunctional reagents were allowed if they were considered particularly medicinally relevant, such as protected amino acids or bis-amines. These design guidelines were applied along with more generic strategic goals (Table 1), and ensured that we focussed on simpler reagents that were more likely to be popular with medicinal chemists, because they were more likely to have broad utility and give desirable final compounds when enumerated onto a project scaffold.

It is apparent that some of the strategic goals listed conflict with each other when choosing which reagents to prosecute. In particular, the goal of novelty is not always aligned with choosing simple reagents that have the greatest chance of wide utility across numerous projects. Combining these goals is achievable, but requires creativity and rigour, because the tendency when focussing on novelty is to increase MW and complexity, with the concomitant risk of synthesising esoteric reagents that are less likely to be selected by medicinal chemists.

TABLE 1

Design guidelines and strategic goals when designing novel reagents for drug discovery projects, including the 'rule of two'

Design guidelines	Strategic goals
One commonly used reactive group[a]	Novelty
Not readily available from ACD	Simplicity and/or general utility
Rule of 2: MW <200, clogP <2, HBD ≤2, HBA ≤4[b]	Substructures with good PK and/or physical properties
Pure enantiomers preferred	Chirality and/or 3D shape
Synthetically tractable	Lowering lipophilicity

[a] Bifunctional reagents were allowed if they were considered particularly medicinally relevant (e.g. protected amino acids or bis-amines).

[b] MW, clogP, HBA and/or HBD count were calculated on the fragment that is added when it is used for coupling (e.g. for aryl bromides, the bromine atom was removed before calculating the properties, but for amines the whole reagent was used for the calculations).

Which reagents are popular among medicinal chemists?

What constituted a commonly used or popular reactive group was initially a subject of some debate within the team. We were able to refine our criteria as the program progressed, on the basis of which reagent classes proved to be popular and, more recently, by a more systematic analysis of our internal electronic lab notebook (see below). Amines (especially secondary amines), acids, and boronic acids and/or boronate esters were ranked most highly on desirability, followed by aryl halides, alkyl and/or benzyl halides,

aldehydes, alcohols, and anilines, and finally, sulfonyl halides, ketones and isocyanates. Isocyanates and sulfonyl halides are useful in medicinal chemistry, but were given lower priority primarily because of their instability on long-term storage. Small numbers of more esoteric reagents (e.g. hydrazines and lactams) were also prioritised on occasion when there was sufficient rationale.

This prioritisation was principally based on the popularity of these reagent classes within AZ, although we believe it is also broadly consistent with the usage across the pharmaceutical industry. We continually analysed which reagent classes proved to be most popular, so as to inform which reagent classes to invest in for future projects. We are not aware of a similar analysis by reagent type in the literature. However, our prioritised reagent classes are broadly consistent with the types of coupling reaction used to synthesise new marketed drugs [31–34] and with the several published analyses on the most popular chemical transformations in the pharmaceutical industry. These include an analysis of the candidate drugs of Pfizer, GSK, and AZ [35], an analysis of a selection of published structures from the medicinal chemistry groups of the same companies [36], a data set of reactions performed by the respiratory department at GSK [37], and a computational deconstruction of both marketed drugs and of the GVK-BIO database of bioactive compounds [38].

The classification of building blocks into reagent classes is not trivial because of the possibility of building blocks containing more than one unprotected functional group. We created a hierarchy of functional groups, defined according to how likely the group is to be used in the first synthetic step. This hierarchy is given in Supplementary information, using SMARTS to encode the substructures (see Glossary). If a reagent matches two different functional groups, the hierarchy determines which reagent class it is assigned to. For example, if a reagent contains both amine and boronic acid functionalities, it is defined as a boronic acid, because boronic acid matching is higher in the hierarchy compared with an unprotected amine. This reflects the assumption that a chemist is more likely to select a reagent containing both amine and boronic acid functionality initially because of the boronic acid functionality (e.g. for Suzuki or Chan-Lam coupling as the first step) rather than for the amine group. There will be some exceptions to this method, but we have found this hierarchical system to be the most practical classification method to ensure that all building blocks can be assigned to a single reagent class. Building blocks with two highly reactive functionalities (e.g. unprotected amino acids), where the classification would be problematic, were not in any case synthesised for this initiative. Bifunctional reagents can be useful building blocks for drug discovery, but typically one of the reactive functional groups needs to be protected to allow selective, high-yielding reactions.

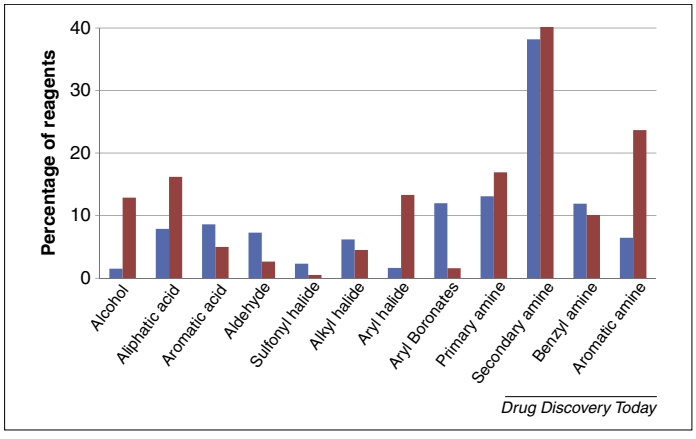

FIGURE 2

A breakdown of reagent classes that were delivered by custom synthesis in the AstraZeneca initiative (SRI, blue) versus those available commercially through ACD (red).

The proportions of different reagent classes that were delivered during 2009–2012 (3044 reagents) are shown in blue in Fig. 2, with the comparison to ACD reagents in red. It is apparent that our delivery as desired was focussed on amines, acids, and boronic acids and/or boronates. ACD also has good numerical coverage of amines and acids, but it is noteworthy that our focus on boronic acids and/or boronates is not observed in ACD, a gap that we identified early on in the programme.

We then compared the properties of our SRI collection to ACD (Fig. 3). Our designed reagents target lower MW and clogP distribution compared with ACD, presumably as a consequence of our 'rule of 2'. The HBA and/or HBD count distributions (and consequently PSA) were similar between our designed reagents and ACD (not shown), with most reagents having one to four acceptors and zero to two donors. Overall, it is striking that 80% of our designed SRI collection conforms to the 'rule of 2', compared with only 7% of ACD building blocks (ACD building blocks defined as reagents in ACD with MW <350 that match to one of the substructures in Fig. 2). Our designed reagents have a markedly different rotatable bond distribution, not only because of the lower MW distribution, but also as a consequence of the fact that experienced medicinal chemists have filtered the ideas by eye, and tend to avoid prioritising structures with a lot of rotatable bonds.

A strategic focus on 3D shape and physical properties is apparent in the ring distribution (not shown), with a greater focus on aliphatic rings versus aromatic rings. Analysis of Fsp^3 (no. of sp^3 carbons/total no. of carbons) [20] shows that the SRI collection, relative to ACD, has focussed on two distinct sets. There is a slightly greater proportion of low Fsp^3 reagents, typically SAR sets of desirable aromatic substituents, particularly heteroaromatic reagents, because those are sometimes hard to obtain through ACD, or those with novel and/or isosteric functional groups on the aromatic ring. There are also a greater proportion of high Fsp^3 reagents that incorporate more saturated ring systems and/or substituents to increase 3D shape and improve properties. Examples of low (**14**–**16**) and high (**17** and **18**) Fsp^3 building blocks that we synthesised are shown in Fig. 4, none of which are in ACD, and again all building blocks shown have been used by projects and elaborated into screening compounds in the AZ compound collection.

We analysed the impact and popularity of the AZ synthesised (SRI) and commercial (ACD) reagents in two ways: first by analysing novel compounds in the AZ screening collection that contained the SRI reagent as a substructure, and second by monitoring usage of reagents from both sets in our electronic lab notebook (ELN). Both approaches have limitations because of assumptions made in the analysis. The first approach (analysis of final screening

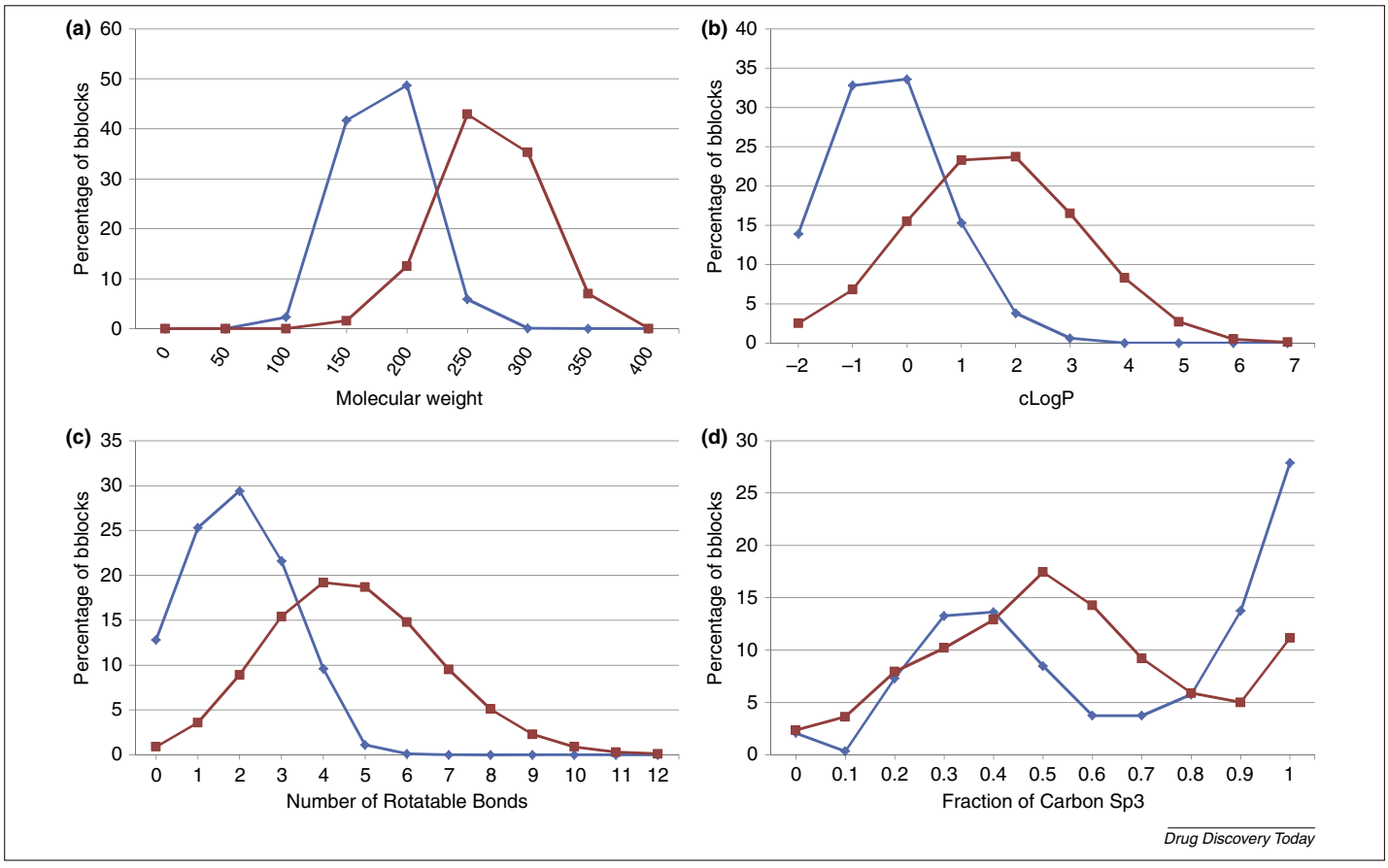

FIGURE 3

The AstraZeneca strategic reagent initiative (SRI; blue) reagents target **(a)** lower clogP, **(b)** molecular weight (MW), and **(c)** rotatable bond count compared with ACD (red), and have a different Fsp^3 distribution **(d)**. As before, these descriptors are all on the added fragment when coupled; for example, for aryl bromides, the bromine atom has been removed.

compounds by substructure) was only possible for the SRI set, because simple ACD reagents would spuriously match to many compounds that did not really use the reagent. For the SRI set, this analysis might slightly overpredict the true usage because some of the final compounds might have been made using alternative reagents to generate the same substructure, or the same reagent might have been made independently by the project team. We limited the impact of those occurrences by excluding examples where significant numbers of final compounds matched, that could not have been from the building block that we provided (see Supporting information for methods), and also using the registration date to only include final compounds that were registered after the building block was available in our reagent stores.

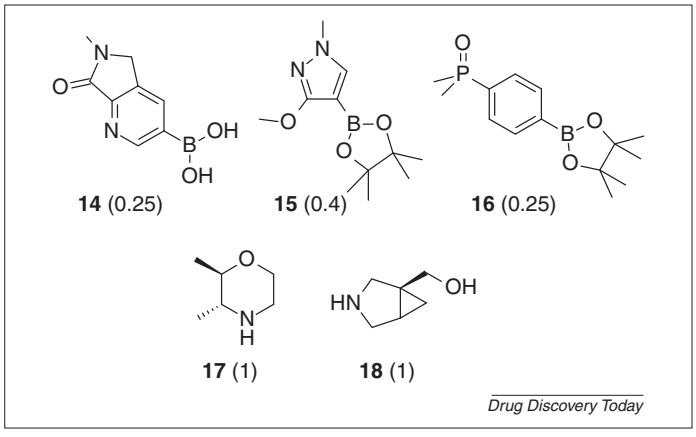

FIGURE 4

Examples of low Fsp^3 (top) and high Fsp^3 (bottom) synthesised reagents, not found in ACD. Fsp^3 values for the added fragment when coupled are shown in parentheses.

The second approach (analysis of the usage in our ELN) might not reflect positive impact on projects; in particular, we have not considered for this article whether the reaction was successful and whether it ultimately led to a screening compound. However, with those caveats, we regard both approaches to be useful to inform our reagent strategy. The first approach will tell us which SRI reagents had the most project impact and also more generally which substructures are popular to AZ projects, and the second approach will tell us directly which reagents (whether ACD or not) are used most by AZ chemists. In addition, the data sets for both approaches are of a sufficient size (130,000 final compounds in the AZ compound collection match to SRI reagents, and the ELN database used contained 690,000 reactions) and diversity (numerous disease areas and sites and/or countries) to provide meaningful conclusions.

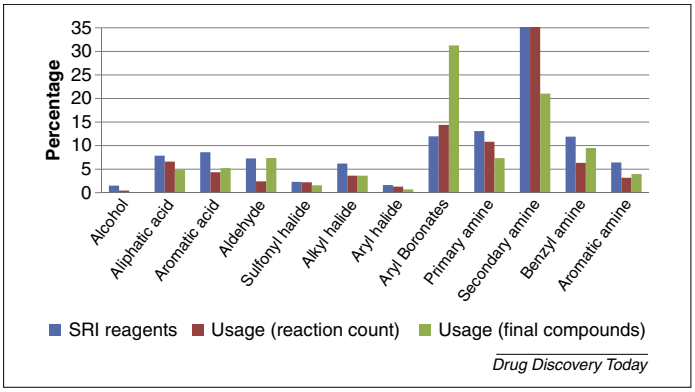

FIGURE 5

Number of strategic reagent initiative (SRI) reagents delivered versus usage for each reagent class. Usage is broken down into screening compounds delivered (substructure matching) and electronic lab notebook (ELN) usages. Available Chemicals Directory reagent usages in the ELN are given as a comparison.

The resulting data by reagent class are given in Fig. 5, where the number of reagents delivered against each reagent class is compared with the resulting screening compounds and reaction count in the ELN. In general, usage matches the supply of the reagent classes. One notable exception is aryl boronates, where the relatively modest delivery of building blocks has resulted in a large number of final screening compounds, reflecting the particular popularity of that reagent class in Suzuki couplings to afford diverse final compounds. The slight disconnect between reaction count usage and final compound usage for aryl boronates suggests that this reagent class has proved to be particularly useful for parallel synthesis.

Although the breakdown by reagent class shown in Fig. 5 is instructive, it is perhaps more useful to look at specific structures that have proved to be unusually popular. Fig. 6 shows some examples (compounds **19**–**22**) of SRI reagents, not found in ACD, that have all been incorporated into >50 final compounds in the AZ corporate collection across more than 10 projects, and have more than ten validated usages in our electronic lab notebook. By this definition, reagents **10**, **11**, **12**, **14**, and **15** shown in Fig. 1 and Fig. 4 can also be classified as 'popular'. In general, these unusually popular reagents are characterised by having relatively simple, low-MW structures that nevertheless offer the potential to provide attractive physical properties as well as providing useful SAR, by incorporating structures that can lower log *D*. The nine examples of 'popular' reagents shown here conform comfortably

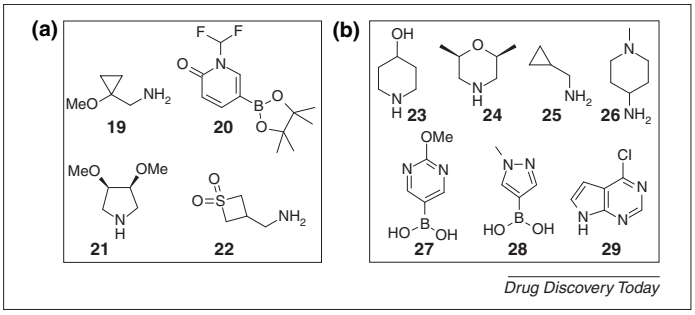

FIGURE 6

Examples of unusually popular strategic reagent initiative (SRI) reagents **(a)** and Available Chemicals Directory reagents **(d)**, as defined by usage statistics.

to the 'rule of 2' criteria; in fact, the MW added by those particular reagents are all <150 and the clogP value increases are all <1. A selection of popular ACD reagents is also shown in Fig. 6 (compounds **23**–**29**), which have all been used in >200 reactions in our ELN. Not surprisingly, the unusually popular ACD reagents are characterised by having simple structures that are desirable to medicinal chemists (e.g. cyclic secondary amines and heterocycles) and are frequently applied to scope basic SAR on a project scaffold. In general, we have observed that the most popular reagents tend to be popular across different target classes and therapeutic areas, and a manual inspection of the examples in Fig. 6 show that to be the case. However, there are exceptions and compound **29** is one example where the frequent use of this building block within AZ might be because of its prevalence in kinase drug discovery [39].

Concluding remarks and future perspectives

Novel, carefully designed, building blocks that can be delivered by custom synthesis have proved to be popular with AZ medicinal chemists, as judged by usage, and such an approach can be considered as a valuable approach to enhancing the quality and speed of drug discovery. This particular initiative at AZ only started in 2009, but we already have examples of SRI reagents being used to generate three candidate drugs and numerous shortlist candidates. One benefit of an extensive, high-quality building block collection is that it enables projects to make serendipitous discoveries that can provide a significant 'step-jump' improvement in profile in one iteration. We have observed several occasions where our designed building blocks enabled projects to take a big step forward, by scanning available building blocks with parallel (library) chemistry. Project data often do not support making such large changes to the molecule, so incremental, iterative design is sometimes not capable of delivering the same results. In addition, a perhaps unexpected extra benefit of the programme was that, by publishing the rationale behind these reagents and discussing the impact in, for example, internal newsletters, and by the nature of the global team that runs it, the programme has facilitated medicinal chemistry learning across projects, sites, and disease areas.

In summary, we believe that this approach of using a global team to design custom synthesis reagents, with appropriate quality control of ideas as discussed above, is a successful strategy to enhance drug discovery programmes and can help to ensure that medicinal chemistry learning is disseminated and captured in our reagent stores.

Our approach has been to invest a lot of thought and rationale into a relatively small set of custom synthesis reagents, and to address the prioritised areas with challenging synthesis if required (more than five synthetic steps and/or challenging synthetic methodologies). However, the general strategy and design principles outlined here could also be applied to an acquisitions strategy from commercial suppliers if larger numbers of reagents were required at low cost. As a possible future development, the same principles could also be applied to an open-innovation approach, where ideas of novel structures to synthesise and physical building blocks could be accessed from diverse external sources, whether from industrial, commercial, or academic laboratories. Another attractive approach would be to consider sharing novel building block collections between organisations, where again the general principles outlined

here could be useful to focus on sharing the building blocks that are most likely to have general utility. Open innovation approaches [40,41], including collection sharing [11], are now well precedented for final screening compounds, but have not to our knowledge been well explored for sharing novel building block collections. Clearly, there would be intellectual property challenges to overcome for such approaches to be successful, as there are with open access and sharing of screening compounds [42]. However, with appropriate agreements in place, this is in principle an attractive approach to increase compound quality in a cost-efficient manner.

Acknowledgements

The authors would like to acknowledge the SRI team and all those who contributed ideas, especially Richard Ward, Yantao Chen, Laurent Knerr, Mike Waring, Iain Simpson, Michael Kossenjans, and Lena Ripa, and also Péter Várkonyi for his assistance with the data analysis.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.drudis.2014.09.023.

References

1 Yusof, I. and Segall, M.D. (2013) Considering the impact drug-like properties have on the chance of success. *Drug Discov. Today* 18, 659–666
2 Bickerton, G.R. *et al.* (2012) Quantifying the chemical beauty of drugs. *Nat. Chem.* 4, 90–98
3 Leeson, P.D. and Springthorpe, B. (2007) The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discov.* 6, 881–890
4 Tarcsay, A. *et al.* (2012) Impact of lipophilic efficiency on compound quality. *J. Med. Chem.* 55, 1252–1260
5 Cumming, J.G. *et al.* (2013) Chemical predictive modelling to improve compound quality. *Nat. Rev. Drug Discov.* 12, 948–962
6 Waring, M.J. (2010) Lipophilicity in drug discovery. *Expert Opin. Drug Discov.* 5, 235–248
7 Ritchie, T.J. and Macdonald, J.F. (2014) How drug-like are 'ugly' drugs: do drug-likeness metrics predict ADME behaviour in human? *Drug Discov. Today* 19, 489–495
8 Stocks, M.J. *et al.* (2009) Practical method for targeted library design balancing lead-like properties with diversity. *Chem. Med. Chem.* 4, 800–808
9 Schamberger, J. *et al.* (2011) Rendezvous in chemical space? Comparing the small molecule compound libraries of Bayer and Schering. *Drug Discov. Today* 16, 636–641
10 Nicholls, A. *et al.* (2010) Molecular shape and medicinal chemistry: a perspective. *J. Med. Chem.* 53, 3862–3886
11 Kogej, T. *et al.* (2013) Big pharma screening collections: more of the same or unique libraries? The AstraZeneca-Bayer Pharma AG case. *Drug Discov. Today* 18, 1014–1024
12 Rosen, J. *et al.* (2009) Novel chemical space exploration via natural products. *J. Med. Chem.* 52, 1953–1962
13 Constantino, L. and Barlocco, D. (2006) Privileged structures as leads in medicinal chemistry. *Curr. Med. Chem.* 13, 65–85
14 Ritchie, T.J. *et al.* (2012) The developability of heteroaromatic and heteroaliphatic rings: do some have a better pedigree as potential drug molecules than others? *Med. Chem. Commun.* 3, 1062–1069
15 St Jean, D.J. and Fotsch, C. (2012) Mitigating heterocycle metabolism in drug discovery. *J. Med. Chem.* 55, 6002–6020
16 Goldberg, F.W. (2014) Design of novel reagent sets to enhance drug discovery. *247th ACS National Meeting*, Dallas, TX, USA, MEDI-29
17 Ward, R.A. and Kettle, J.G. (2011) Systematic enumeration of heteroaromatic ring systems as reagents for use in medicinal chemistry. *J. Med. Chem.* 54, 4670–4677
18 Kettle, J.G. *et al.* (2010) Data-mining patent literature for novel chemical reagents for use in medicinal chemistry design. *Med. Chem. Commun.* 1, 331–338
19 Marino, S.T. *et al.* (2004) Synthesis of chiral building blocks for use in drug discovery. *Molecules* 9, 405–426
20 Lovering, F. *et al.* (2009) Escape from flatland: increasing saturation as an approach to improving clinical success. *J. Med. Chem.* 52, 6752–6756
21 Meanwell, N.A. (2011) Synopsis of some recent tactical application of bioisosteres in drug design. *J. Med. Chem.* 54, 2529–2591
22 Lücking, U. (2013) Sulfoximines: a neglected opportunity in medicinal chemistry. *Angew. Chem. Int. Ed.* 52, 9399–9408
23 Goldberg, F.W. *et al.* (2014) General synthetic strategies towards N-alkyl sulfoximine building blocks for medicinal chemistry and the use of dimethylsulfoximine as a versatile precursor. *Tetrahedron* http://dx.doi.org/10.1016/j.tet.2014.06.120
24 Wang, J. and Hou, T. (2010) Drug and drug candidate building block analysis. *J. Chem. Inf. Model.* 50, 55–67
25 Goldberg, F.W. *et al.* (2012) Free-Wilson and structural approaches to co-optimizing human and rodent isoform potency for 11β-hydroxysteroid dehydrogenase type 1 (11β-HSD1) inhibitors. *J. Med. Chem.* 55, 10652–10661
26 Yngve, U. *et al.* (2013) Triazolopyrimidinones as gamma-secretase modulators: structure-activity relationship, modulator profile, and *in vivo* profiling. *Med. Chem. Commun.* 4, 422–431
27 Lipinski, C.A. *et al.* (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 23, 3–25
28 Oprea, T.I. *et al.* (2001) Is there a difference between leads and drugs? A historical perspective. *J. Chem. Inf. Comput. Sci.* 41, 1308–1315
29 Congreve, M. *et al.* (2003) 'Rule of three' for fragment-based lead discovery? *Drug Discov. Today* 8, 876–877
30 Nadin, A. *et al.* (2012) Lead-oriented synthesis: a new opportunity for synthetic chemistry. *Angew. Chem. Int. Ed.* 51, 1114–1122
31 Liu, K.K. *et al.* (2011) Synthetic approaches to the 2009 new drugs. *Bioorg. Med. Chem.* 19, 1136–1154
32 Liu, K.K. *et al.* (2012) Synthetic approaches to the 2010 new drugs. *Bioorg. Med. Chem.* 20, 1155–1174
33 Ding, H.X. *et al.* (2013) Synthetic approaches to the 2011 new drugs. *Bioorg. Med. Chem.* 21, 2795–2825
34 Ding, H.X. *et al.* (2014) Synthetic approaches to the 2012 new drugs. *Bioorg. Med. Chem.* 22, 2005–2032
35 Carey, J.S. *et al.* (2006) Analysis of the reactions used for the preparation of drug candidate molecules. *Org. Biomol. Chem.* 4, 2337–2347
36 Roughley, S.D. and Jordan, A.M. (2011) The medicinal chemist's toolbox: an analysis of reactions used in the pursuit of drug candidates. *J. Med. Chem.* 54, 3451–3479
37 Cooper, T.W.J. *et al.* (2010) Factors determining the selection of organic reactions by medicinal chemists and the use of these reactions in arrays (small focused libraries). *Angew. Chem. Int. Ed.* 49, 8082–8091
38 Hartenfeller, M. *et al.* (2012) Probing the bioactivity-relevant chemical space of robust reactions and common molecular building blocks. *J. Chem. Inf. Model.* 52, 1167–1178
39 Xing, L. *et al.* (2014) Scaffold mining of kinase hinge binders in crystal structure database. *J. Comput. Aided Mol. Des.* 28, 13–23
40 Schuhmacher, A. *et al.* (2013) Models for open innovation in the pharmaceutical industry. *Drug Discov. Today* 18, 1133–1137
41 Alvim-Gaston, M. *et al.* (2014) Open innovation drug discovery (OIDD): a potential path to novel therapeutic chemical space. *Curr. Top. Med. Chem.* 14, 294–303
42 Judd, D.B. (2013) Open innovation and intellectual property: time for a reboot? *Drug Discov. Today* 18, 907–909