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

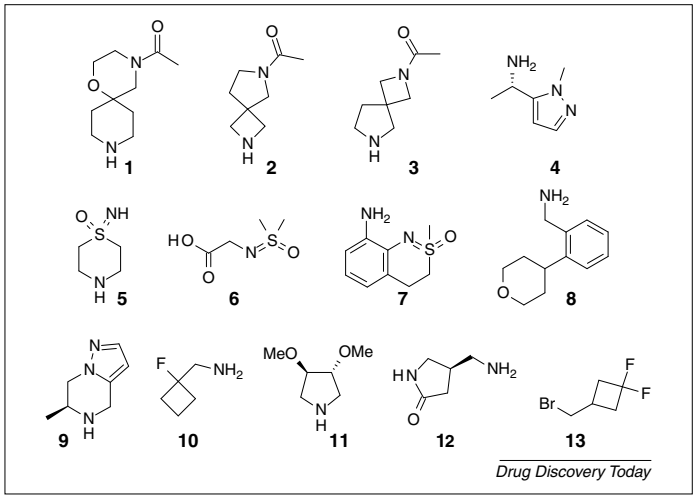

FIGURE 1

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.