ELSEVIER
DRUG DISCOVERY TODAY TECHNOLOGIES
Drug Discovery Today: Technologies
Vol. xxx, No. xx 2017

The evolution of library design: crafting smart compound collections for phenotypic screens☆

Kerry L. Spear[1,*], Scott P. Brown[2]

[1]Blue Oak Pharmaceuticals Inc., Concord, MA, USA
[2]Sunovion Pharmaceuticals Inc., 84 Waterford Dr., Marlborough, MA 01752, USA

The (re)emergence of phenotypic drug discovery has been marked by a growing interest in screening campaigns that utilize phenotypic assays. The key objectives of phenotypic screens are different from those of target-based screens and can require alternate library-design strategies. Designing a library that is appropriate to the selected assay increases the likelihood of identifying better quality hits, which can reduce both timelines and overall cost of the drug-discovery process. Here, we provide an overview of small-molecule library design principles as applied to phenotypic screening.

Introduction

For researchers engaged in target-based drug discovery (TDD), the single most important program decision is selection of a molecular target. If the target is not clinically relevant, there can be no drug. Protein targets are components of complex cellular systems that are both robust and adaptive. Isolation of a single target from a cellular system may facilitate assay design, or simplify data collection, but the relevance of active compounds to a systems context can be lost. This realization has led to a renewed interest in phenotypic drug discovery (PDD). The recognition that diseases can arise from defects in biological systems (rather than defects in molecular-target function) has spurred the shift toward phenotypic assays, which can be more physiologically relevant. With PDD, knowledge of the molecular target is not required. Instead, programs can be driven entirely by assessing functional endpoints and so may be more effective in identifying molecules that engage elements of signaling pathways or key regulatory nodes. This understanding has contributed to the emergence of PDD as an alternative and complimentary approach to TDD [1].

Screening can be an effective mechanism for drug discovery in either a TDD or a PDD setting. One crucial aspect of successful screening is the design of the compound library. The quality of the hits that emerge from screens will influence all subsequent project decisions. Simply put, a well-designed library should contain high-quality compounds, and any molecule with identifiable liabilities should be excluded from further consideration as it represents a resource-consuming distraction. Fewer hit liabilities translate into more efficient

We thank Thomas H. Large for his careful reviews and helpful suggestions during the preparation of the manuscript.
*Corresponding author:: K.L. Spear (kerry.spear@blueoakpharma.com)

1740-6749/$ http://dx.doi.org/10.1016/j.ddtec.2017.05.001

Please cite this article in press as: Spear KL, Brown SP. The evolution of library design: crafting smart compound collections for phenotypic screens, Drug Discov Today: Technol (2017), http://dx.doi.org/10.1016/j.ddtec.2017.05.001

project timelines and greater likelihoods of developing Advanced Leads that have better capabilities of successfully navigating hurdles in preclinical and clinical development.

The goal of this review is to present an overview of small-molecule library design principles as applied to phenotypic screening. The scope is not intended to be comprehensive but rather to provide a summary of recent developments with an emphasis on work published in the last two years.

A selective history of library design: ``What's past is prologue'' [2]

Prior to the early 1990s, screening campaigns relied on libraries that were more likely to be assembled than designed. These were typically idiosyncratic collections of corporate compounds created during the course of drug-discovery programs that could well date back decades. This began to change with the emergence of combinatorial chemistry as well as the sudden commercial availability of often structurally novel compound collections from academic groups in Eastern Europe. High-throughput screening (HTS) methodologies facilitated and accelerated this changing landscape by enabling large collections of molecules to be rapidly and systematically tested. The early exuberance associated with combinatorial chemistry and HTS led to a belief that screening millions of compounds was a desirable goal. Only in retrospect did it become apparent that the added expense did not lead to improvements in overall success rates. Consequently, awareness of the value of library-design principles began to emerge as a means of managing library size and compound quality.

Two important trends in the evolution of library design are illustrated in Fig. 1a. Over time, there has been a general reduction in the number of compounds screened [3], which has been accompanied by overall improvements in compound quality. A consequence of low hit rates from TPP screening campaigns resulted in most early libraries being designed using probabilistic strategies. This encouraged a culture of quantity, resulting in very large library sizes. Eventually, alternate approaches emerged that emphasized compound quality over library size. Efforts to devise more careful compound selection strategies had two complimentary objectives; to improve hit rates [4] or to reduce the attrition rates found during hit triage [4] and beyond [5]. One result was improved effectiveness of the screening process.

Our understanding of what it means to be a quality compound has also evolved (Fig. 1b). Early attempts to refine quality largely focused on improving the structural properties of the compounds. Initially, this meant optimizing structural diversity within a broadly defined chemical space as an alternative to assembling a random collection of compounds. This often relied on the experiences (and biases) of the medicinal chemists leading the process. Later, computational methodologies (e.g., Tanimoto-based similarity comparisons) were introduced to make the selection process more objective. One of the most significant developments in structurally focused design was the concept of the drug-like molecule [6]. This led to implementation of filtering protocols to triage compound collections based on physicochemical properties. It was also quickly extended to include concepts such as lead-like properties [6] and ''bad actor'' motifs [7] (e.g., reactive substructures and pan assay interference compounds). The goal was to increase chemical relevance by enriching libraries with compounds having desirable structural properties, effectively biasing libraries toward specific regions of chemical space.

A design strategy that began to gain traction in the late 1990s explicitly targets parts of biologically relevant chemical space (e.g., gene-family libraries [8]). Biologically relevant

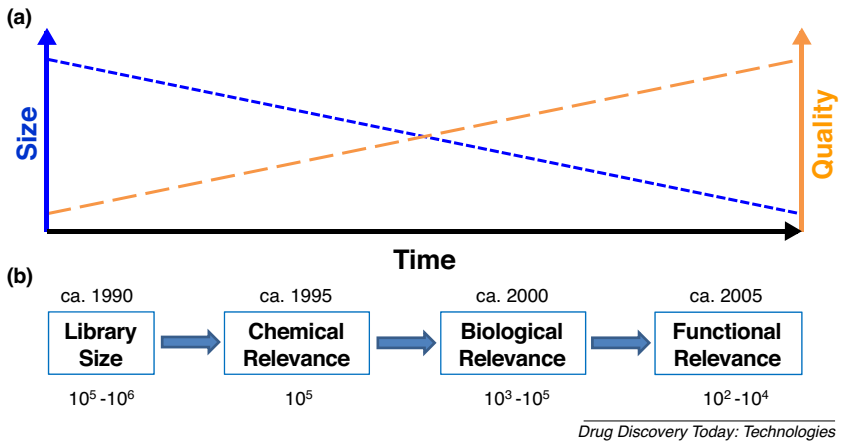

Figure 1. (a) Trends in library design illustrating the evolving role of size and quality over time. (b) Each box indicates a significant milestone in the evolution of design considerations that contributed to the trends exemplified in (a). Above each box we indicate an approximate year of adoption. Typical values for the kinds of library sizes appropriate to each design strategy are indicated below each box.

Please cite this article in press as: Spear KL, Brown SP. The evolution of library design: crafting smart compound collections for phenotypic screens, Drug Discov Today: Technol (2017), http://dx.doi.org/10.1016/j.ddtec.2017.05.001

design is enabled by the realization that chemical diversity may not necessarily overlap biological space. This approach seeks to improve hit rates by defining desirable regions within bioactivity space [9]. It often draws inspiration from the vast selection of biogenic compounds and one version employs a strategy that incorporates properties of natural products (NPs) into the design process [10]. Functionally relevant design is a variation on this theme that selects NPs having biological activity appropriate for project objectives as starting points for library design [11].

An alternate approach to HTS is fragment-based drug discovery (FBDD) [12]. As the name implies, FBDD focuses on screening low molecular weight (MW) compounds (typically 150–300 Da). A key advantage of this approach is that small libraries (often a few thousand members) can more efficiently cover the designated chemical space. The primary disadvantage is that hits tend to exhibit very weak activity in *in vitro* target-based assays. As a consequence, a combination of biochemical and biophysical methods (notably NMR spectroscopy, X-ray crystallography or surface plasmon resonance) are frequently required in order to identify active fragments. These fragments can then be stitched together to afford molecules having improved potency and selectivity for the desired target. This strategy has been used to synthesize novel molecules having superior ligand efficiency.

Library design for phenotypic screening: the current state of the art

Understanding the history of library design is important because the design strategies developed for target-based screens are often applied to phenotypic screens. Many of the same practices are used without modification and it is not uncommon to see TDD libraries repurposed for phenotypic screens. This is particularly true for cell-based phenotypic screening. These assays can be adapted to high-density, multi-well HTS formats, thereby facilitating assaying of large, structurally diverse libraries. For example, a recent HTS of a 1.7 million compound corporate collection identified hits that up-regulated leukemia inhibitory factor [13]. Another example used the Molecular Libraries Small Molecule Repository (>320,000 compounds curated by the NIH based on purity, physicochemical properties, exclusion of reactive groups and optimization of diversity) to screen for cellular lipid deposition [14]. A large library was also recently used for an *in vivo* assay [15]. Here, a probabilistic strategy was applied to screen several combinatorial libraries having sizes ranging up to >4.8 million compounds for analgesic activity in the mouse tail-withdrawal assay. Samples were pooled to facilitate screening and, after deconvolution, compounds from one of the scaffolds afforded the desired antinociceptive response.

Screening smaller, more astutely chosen compound collections can be an effective alternative to large libraries. Small libraries (defined here as <10,000 compounds) are more likely to be designed for a specific function. The use of focused libraries for *in vivo* phenotypic screens can be particularly important when throughput limitations exist. One recent report describes the results of a screening campaign that applied a 1000-compound library to a mouse-based assay designed to identify behaviorally active compounds [16]. The library was designed, in part, by filtering a set of commercially available compounds for physicochemical properties associated with CNS lead-likeness. The final selection was based on optimization of structural diversity. In another example, a library of 80 commercially available compounds selected for structural diversity was applied to a zebrafish anticancer assay [17]. The compound identified from this screen was subsequently found to selectively inhibit PIM3 kinase and activate DAPK1 kinase.

Focused libraries are also used for cell-based phenotypic screening. As might be expected, many of these involve single-celled pathogens. For example, a biased library of 400 compounds, known as the ''Malaria Box,'' was selected from a larger set of 20,000 validated hits identified using a phenotypic antimalarial HTS [18]. The selection process optimized physicochemical properties and diversified scaffold structures. In addition to being an open-source library for the malaria research community, the Malaria Box has also been used in other parasite screens [19]. There is also an increasing number of reports that apply human cell lines to phenotypic screens. In one study, a library of about 300 compounds was tested in two separate high-content (HC) screens, a mast cell degranulation assay and an endocytosis assay [20]. This library was designed to optimize the structural diversity of a subset of amphiphilic compounds selected from a larger compound set. Another group screened a library of 7400 structurally diverse compounds in a proliferation assay using primary human epicardium-derived cells [21]. The compounds in this set were selected either because they were known to act via certain molecular mechanisms or because of ''a reported ability to modulate the phenotype of stem or primary cells.''

New directions for library design: applying lessons from nature

The challenge of repurposing libraries designed for target-based screening (or designing libraries *de novo* based on standard practices developed for target-based screening) for phenotypic assays has been understood for a while now [22]. Since phenotypic screens are designed to interrogate intact biological systems, observed activities may well arise from hits interacting with multiple proteins. Differences in the objectives between target-based and phenotypic-based approaches have encouraged exploration of alternative design strategies. While not all are specific to phenotypic screening, some may well prove to be particularly relevant

Please cite this article in press as: Spear KL, Brown SP. The evolution of library design: crafting smart compound collections for phenotypic screens, Drug Discov Today: Technol (2017), http://dx.doi.org/10.1016/j.ddtec.2017.05.001

to PDD. Several these approaches have been informed by the growing understanding of the role that molecular complexity plays in protein-ligand binding and the consequences to both potency and selectivity. By testing over 15,000 compounds assembled from three sources (commercial, academic and NPs) in a battery of 100 protein binding assays, Clemons et al. identified specific molecular properties that correlate with binding [23]. Notably, both the likelihood of binding to at least one protein as well as hit selectivities were shown to increase with the fraction of sp^3 carbons (Fsp^3C) found in the core scaffolds, as well as with an intermediate proportion of stereocenters. More recently, Nilar et al. expanded the Hann model of molecular complexity to probe protein-ligand binding modes [24]. Based on their analysis, they concluded that low complexity compounds have the highest probability of activity in target-based assays while compounds having an increased degree of complexity improve the likelihood of activity in phenotypic screens. This was rationalized by noting that phenotypic assays simultaneously measure interactions with multiple targets, each having various degrees of complexity. Including ligands with a moderate level of complexity in a screening campaign permits effective binding to the more complex protein active sites in the system.

The benefits of designing and screening molecules based on three-dimensional considerations are becoming clear [25]. However, as described in a recent review [26], a more systematic exploration of this region of chemical space can be hampered by current synthetic methodologies, some of which are limited in scope to nonpolar aromatic substitutions that are often unfunctionalizable. This is unfortunate, since a key requirement of library design is to modulate molecular complexity by synthetic accessibility. A pioneering effort to address this synthetic challenge, diversity-oriented synthesis (DOS), focuses on designing compounds having increased structural complexity. The primary goal of this strategy is to develop robust, modular and efficient synthetic routes to scaffolds that access under-populated regions of chemical space. Furthermore, the strategy explicitly recognizes the importance of incorporating structural complexity into the design process. Relative to synthetic drugs, DOS compounds tend to have an increased Fsp^3C and more stereocenters (examples are shown in Fig. 2a). DOS libraries have been used in several phenotypic screens and applications to the discovery of novel antimicrobials [27] and antimalarials [28] were recently reviewed. Additionally, a large (>100,000 compound) DOS umbrella library was synthesized at the Broad

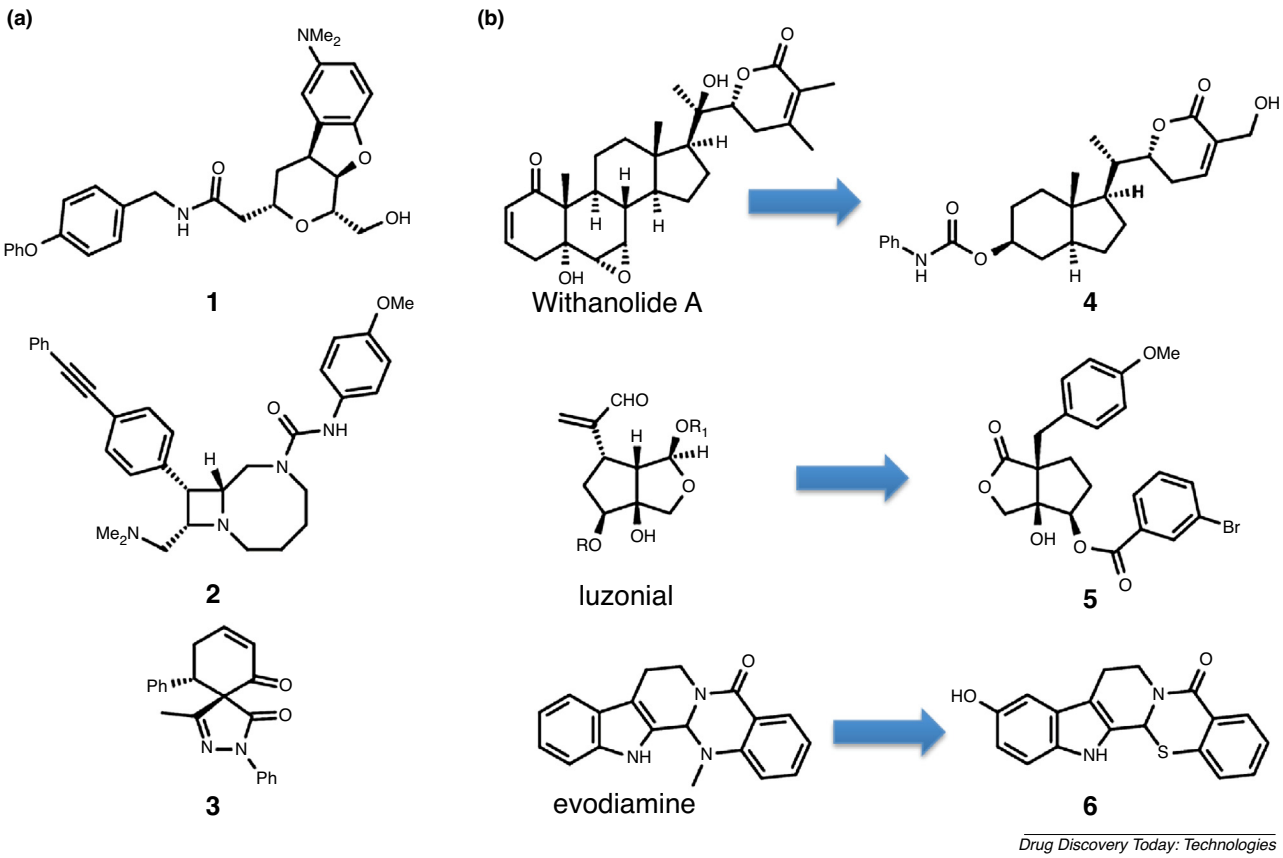

Figure 2. Representative examples of structures synthesized using two different approaches, (a) a DOS strategy and (b) a BIOS strategy.

Please cite this article in press as: Spear KL, Brown SP. The evolution of library design: crafting smart compound collections for phenotypic screens, Drug Discov Today: Technol (2017), http://dx.doi.org/10.1016/j.ddtec.2017.05.001

Institute and has been employed in several phenotypic assays including a HTS that identified modulators of hepatic lipoprotein metabolism (e.g., **1**) [29] and an antiviral screen [30]. A particularly exciting report describes the use of this library to identify structurally and mechanistically novel antimalarial inhibitors (e.g., **2**) [31]. Applications to phenotypic cancer screens have also been reported. For example, a set of 245 compounds selected from 23 novel heterocyclic scaffolds was assayed against 28 leukemia cell lines to identify several active benzothiazepines [32]. In a variation of the DOS strategy, a privileged structure concept was coupled with DOS to create a small library of spiropyrazolones that were screened in solid-tumor cell lines [33] to afford several hits (e.g., **3**).

One of the more intriguing trends to emerge in recent years is the design of libraries that draw inspiration from NPs. The growing awareness of the importance of molecular complexity to biological relevance has led to the identification of scaffolds that are sub-structures of NPs or include structural features of NPs [34]. The concept is based on the recognition that, since evolutionary pressures were responsible for designing these bioactive chemotypes, NPs can be useful vehicles for focusing design efforts onto a space that has already been evolutionarily prevalidated. Based on physicochemical properties, NPs tend to occupy a region of chemical space that is distinct from synthetic drugs (e.g., more stereogenic centers, greater Fsp^3C, fewer N atoms, more O atoms, increased hydrophilicity) [35]. However, when pharmacophoric features of sets of NPs were compared to synthetic drugs by projection onto a generative topographic map, significant overlap between the two sets was found [36]. Lipinski recently published a thoughtful comparison of NPs and synthetic ligands [37]. Additionally, an excellent review of the role of NPs in drug design was also just published [38].

As with the DOS strategy, a key to the success of NP-inspired drug design is the development of short, efficient synthetic routes. The goal here is to identify chemotypes that mimic aspects of NPs. A key application of NP-inspired design, biology-oriented synthesis (BIOS) [39], has led to several screening campaigns (Fig. 2b). For example, the withanolide family of NPs was the inspiration for the synthesis of a small library of 30 bicyclic analogs that was screened in cell-based assays to survey signal transduction pathways and four hits (e.g., **4**) were identified [40]. In a second example, luzonial was used as the starting point for the synthesis of 50 analogs representing two scaffolds [41]. Subsequent screening identified 11 hits (e.g., **5**) that inhibited autophagy in a cell-based assay. Finally, a small library consisting of 11 scaffolds based on the structure of evodiamine was screened against human cancer cell lines to identify a hit (**6**) having robust activities in all assays [42].

When NP-inspired design is coupled with FBDD, the resulting libraries can be enriched with small, synthetically accessible molecules having a surprisingly high degree of molecular complexity. For some therapeutic indications, notably CNS, enriching a library with low MW compounds can be beneficial. This is likely due, at least in part, to the requirement that centrally active molecules cross the blood–brain barrier. Since hits from phenotypic screens must typically transit some biological membrane, enriching libraries with low MW compounds can improve hit rates. As noted above, hits from FBDD screens for TDD applications are usually weakly potent. However, this may not be a deterrent for PDD screening. It is tempting to speculate that the observed phenotypic activity of some hits may arise from weak binding to multiple proteins that interact synergistically within a biological system to afford unexpectedly high efficacy. It may even be possible to advance a fragment hit directly into lead optimization without relying on structure-based drug design or biophysical techniques to first improve potency. A recent report describes the *in vivo* optimization of a 192 Da hit identified from an *in vivo* phenotypic screen to afford an Advanced Lead having excellent efficacy in multiple animal models while exhibiting modest to weak binding to several receptors in a 53-receptor panel [16]. While the screening set used was not fragment-based, the hit has properties normally associated with fragments.

Two publications described interesting methods for identifying and exploiting NP-inspired fragment libraries. In one, Waldmann and co-workers computationally generated fragments of over 180,000 NP structures to identify 2000 clusters of low MW (120–350 Da) natural-product-derived fragments (NPDFs) that are structurally diverse and occupy a part of chemical space that is mostly different from that occupied by commercially available fragments [43] (see Fig. 3a for examples). Using a different strategy, Pascolutti et al. filtered >165,000 NPs to identify a set of ''fragment-sized NPs'' (MW $\leq$250), including 7365 ''non-flat'' molecules [44] (see Fig. 3b for examples). Analysis of these compounds confirmed a high degree of structural diversity. Applications of NPDFs are just beginning to be explored. One recent review describes the use of spirooxindole NPDFs as starting points for the discovery of anticancer drugs [45].

Future trends

As the volume of publically available screening data (both TDD- and PDD-based) continues to expand, efforts to apply these data to develop computational tools that can be used to improve the compound selection process are intensifying. In particular, recent advances in both computing speed and data-mining algorithms have encouraged analysis of screening data associated with large compound sets in order to identify structural properties that correlate with biological activity. For example, in a retrospective analysis, molecular descriptors created using HTS data from >300,000 compounds were used to predict actives from a set of 33 assays [46]. Overall, a 27-fold enrichment (relative to a random

Please cite this article in press as: Spear KL, Brown SP. The evolution of library design: crafting smart compound collections for phenotypic screens, Drug Discov Today: Technol (2017), http://dx.doi.org/10.1016/j.ddtec.2017.05.001

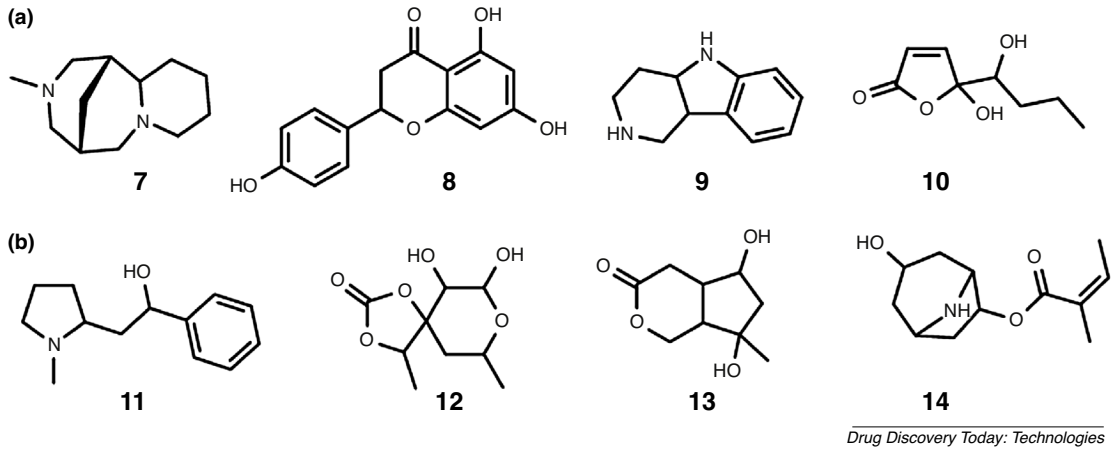

Figure 3. Representative examples of (a) natural-product-derived fragments from Ref. [43] and (b) fragment-sized natural products from Ref. [44].

selection of molecules) was achieved. A similar strategy was applied to a phenotypic assay [47]. Here, a database of 300,000 compounds that had been screened against the parasite that causes Chagas disease was analyzed. Bayesian models were built and used for *in silico* screening of several drug libraries (7569 structures). The 97 top-scoring molecules were then assayed and 11 actives identified, including five having sub-micromolar potencies. The goal of an alternate approach was to identify a more effective way of selecting molecules for cell-based phenotypic screens by building models that did not utilize any structural data [48]. In this work, a strategy of optimizing diversity in phenotypic biological space (as opposed to chemical space) was employed. A set of 31,770 compounds (including 19,164 DOS compounds) was profiled in two HC assays that measured changes in cell morphology and gene expression levels. The performance of sets of actives identified by these assays was then retrospectively evaluated using cell-based HTS. A significant improvement in hit rates was found. This structure-independent approach resulted in identifying a diverse set of actives. The authors suggested that the process could be used as an alternate method of filtering compound collections to optimize biological diversity.

Conclusions

Phenotypic screens are not the same as target-based screens and so will likely require different library design strategies. Adjusting physicochemical property filters and increasing molecular complexity are reasonable first steps. In some cases it may be possible to simultaneously achieve both of these design goals by enriching libraries with NPDFs. Of course, molecular complexity must be balanced with synthetic accessibility. Despite synthetic challenges, there is growing recognition that structures inspired by NPs can be preferred starting points for innovative drug discovery. While the role of concepts such as DOS or BIOS in the design of libraries for phenotypic screening is just now being established, these approaches are already affording structures that are both novel and biologically relevant. Ultimately, as the volume of phenotypic data grows, it seems likely that our understanding of which molecular properties correlate with biological function will continue to evolve. Identifying these properties will require the use of increasingly sophisticated computational tools, but they are likely to lead to a better understanding of how to design high-quality libraries for phenotypic screens.

The importance of good library design should not be under-estimated. Creating a well-designed library can be both time-consuming and expensive. However, it is likely to lead to better quality hits that can be more efficiently, and more effectively, optimized. In the long run, this can reduce both timelines and overall cost of the drug discovery process.

Conflict of interest

The authors have no conflicts of interest to declare.

References

[1] Kell DB. Finding novel pharmaceuticals in the systems biology era using multiple effective drug targets, phenotypic screening and knowledge of transporters: where drug discovery went wrong and how to fix it. FEBS J 2013;280:5957–80.
[2] Shakespeare W. The Tempest. Act 2, Scene1.
[3] Schnur DM. Recent trends in library design: 'Rational design' revisited. Curr Opin Drug Discov Dev 2008;11:375–80.
[4] Paricharak S, Mendez-Lucio O, Ravindranath AC, Bender A, IJzerman AP, van Westen GJP. Data-driven approaches used for compound library design, hit triage and bioactivity modeling in high-throughput screening. Brief Bioinform 2016. http://dx.doi.org/10.1093/bib/bbw105.
[5] Mignani S, Huber S, Tomas H, Rodrigues J, Majoral J-P. Compound high-quality criteria: a new vision to guide the development of drugs, current situation. Drug Discov Today 2016;21:573–84.
[6] Zuegg J, Cooper MA. Drug-likeness and increased hydrophobicity of commercially available compound libraries for drug screening. Curr Top Med Chem 2012;12:1500–13.
[7] Senger MR. Filtering promiscuous compounds in early drug discovery: is it a good idea. Drug Discov Today 2016;21:868–72.

Please cite this article in press as: Spear KL, Brown SP. The evolution of library design: crafting smart compound collections for phenotypic screens, Drug Discov Today: Technol (2017), http://dx.doi.org/10.1016/j.ddtec.2017.05.001

[8] Miller JL. Recent developments in focused library design: targeting gene-families. Curr Top Med Chem 2006;6:19–29.
[9] Petrone PM, Wasserman AM, Lounkine E, Kutchukian P, Simms B, Jenkins J, et al. Biodiversity of small molecules—a new perspective in screening set selection. Drug Discov Today 2013;18:674–80.
[10] Rizzo S, Waldmann H. Development of a natural-product-derived chemical toolbox for modulation of protein function. Chem Rev 2014;114:4621–39.
[11] Wender PA, Quiroz RV, Stevens MC. Function through synthesis-informed design. Acc Chem Res 2015;48:752–60.
[12] Doak BC, Norton RS, Scanlon MJ. The ways and means of fragment-based design. Pharmacol Ther 2016;167:28–37.
[13] Vela L, Caballero I, Fang L, Liu Q, Ramon F, Diez E, et al. Discovery of enhancers of the secretion of leukemia inhibitory factor for the treatment of multiple sclerosis. J Biomol Screen 2016;21:437–45.
[14] Tschapalda K, Zhang Y-Q, Liu L, Golovnina K, Schlemper T, Eighman TO, et al. A class of diacylglycerol acyltransferase1 inhibitors identified by a combination of phenotypic high-throughput screening, genomics and genetics. EBioMedicine 2016;8:49–59.
[15] Houghten RA, Ganno ML, McLaughlin JP, Dooley CT, Eans SO, Santos RG, et al. Direct phenotypic screening in mice: identification of individual, novel antinociceptive compounds from a library of 734821 pyrrolidine bis-piperazines. ACS Comb Sci 2016;18:51–64.
[16] Shao L, Campbell UC, Fang QK, Powell NA, Campbell JE, Jones PG, et al. *In vivo* phenotypic drug discovery: applying a behavioral assay to the discovery and optimization of novel antipsychotic agents. MedChemComm 2016;7:1093–101.
[17] Geldenhuis WJ, Bergeron SA, Mullins JE, Aljammal R, Gaasch BL, Chen W-C, et al. High-content screen using zebrafish (*Danio rerio*) embryos identifies a novel kinase activator and inhibitor. Bioorg Med Chem Lett 2017;27:2029–37.
[18] Van Voorhis WC, Adams JH, Adelfio R, Ahyong V, Akabas MH, Alano P, et al. Open source drug discovery with the malaria box collection for neglected diseases and beyond. PLOS Pathog 2016;12(7):e1005763.
[19] Hostettler I, Muller J, Hemphill A. *In vitro* screening of the open source MMV malaria box reveals novel compounds with profound activities against *Theileria annulata* schizonts. Antimicrob Agents Chemother 2016;60:3301–8.
[20] McShane MP, Friedrichson T, Giner A, Meyenhofer F, Barsacchi R, Bickle M, et al. A combination of screening and computational approaches for the identification of novel compounds that decrease mast cell degranulation. J Biomol Screen 2015;20:720–8.
[21] Paunovic AI, Drowley L, Nordqvist A, Ericson E, Mouchet E, Jonebring A, et al. Phenotypic screen for cardiac regeneration identifies molecules with differential activity in human epicardium-derived cells versus cardiac fibroblasts. ACS Chem Biol 2017;12:132–41.
[22] Keller TH, Shi P-Y, Wang Q-Y. Anti-infectives: can cellular screening deliver? Curr Opin Chem Biol 2011;15:529–33.
[23] Clemons PA, Bodycombe NE, Carrinski HA, Wilson JA, Shamji AF, Wagner BK, et al. Small molecules of different origins have distinct distributions of structural complexity that correlate with protein-binding profiles. Proc Natl Acad Sci U S A 2010;107:18787–92.
[24] Nilar SH, Ma NL, Keller TH. The importance of molecular complexity in the design of screening libraries. J Comput Aided Mol Des 2013;27:783–92.
[25] Meyers J, Carter M, Mok NY, Brown B. On the origins of three-dimensionality in drug-like molecules. Future Med Chem 2016;8:1753–67.
[26] Foley DJ, Nelson A, Marsden SP. Evaluating new chemistry to drive molecular discovery: fit for purpose. Angew Chem Int Ed Engl 2016;55:13650–57.
[27] Comer E, Duvall JR, Lee IV MD. Utilizing diversity-oriented synthesis in antimicrobial research. Future Med Chem 2014;6:1927–42.
[28] Bathula C, Singh S, Sen S. Diversity oriented synthesis for novel anti-malarials. Syst Synth Biol 2015;9:49–53.
[29] Nagiec MM, Skepner AP, Negri J, Eichorn M, Kuperwasser N, Comer E, et al. Modulators of hepatic lipoprotein metabolism identified in a search for small molecule inducers of tribbles pseudokinase 1 expression. PLoS One 2015;10(3):e0120295. http://dx.doi.org/10.1371/journal.pone.0120295.
[30] Duvall JR, VerPlank L, Ludeke B, McLeod SM, Lee IV MD, Vishwanathan K, et al. Novel diversity-oriented synthesis-derived respiratory syncytial virus inhibitors identified via a high throughput replicon-based screen. Antiviral Res 2016;131:19–25.
[31] Kato N, Comer E, Sakata-Kato T, Sharma A, Sharma M, Maetani M, et al. Diversity-oriented synthesis yields novel multistage antimalarial inhibitors. Nature 2016;538:344–9.
[32] Xiang J, Zhang Z, Mu Y, Xu X, Guo S, Liu Y, et al. Discovery of novel tricyclic thiazepine derivatives as anti-drug-resistant cancer agents by combining diversity-oriented synthesis and converging screening approach. ACS Comb Sci 2016;18:230–5.
[33] Wu S, Li Y, Xu G, Chen S, Zhang Y, Liu N, et al. Novel spiropyrazolone antitumor scaffold with potent activity: Design, synthesis and structure-activity relationship. Eur J Med Chem 2016;115:141–7.
[34] Maier ME. Design and synthesis of analogues of natural products. Org Biomol Chem 2015;13:5302–43.
[35] Crane EA, Gademann K. Capturing biological activity in natural product fragments by chemical synthesis. Angew Chem Int Ed Engl 2016;55:3882–902.
[36] Miyao T, Reker D, Schneider P, Funatsu K, Schneider G. Chemography of natural product space. Planta Med 2015;81429–35.
[37] Lipinski CA. Rule of five in 2015 and beyond: target and ligand structural limitations, ligand chemistry structure and drug discovery project decisions. Adv Drug Deliv Rev 2016;101:34–41.
[38] Rodrigues T, Reker D, Schneider P, Schneider G. Counting of natural products for drug design. Nat Chem 2016;8:531–41.
[39] van Hattum H, Waldmann H. Biology-oriented synthesis: harnessing the power of evolution. J Am Chem Soc 2014;136:11853–59.
[40] Svenda J, Sheremet M, Kremer L, Maier L, Bauer JO, Strohmann C, et al. Biology-oriented synthesis of a withanolide-inspired compound collection reveals novel modulators of hedgehog signaling. Angew Chem Int Ed Engl 2015;54:5596–602.
[41] Schroeder P, Bauer JO, Strohmann C, Kumar K, Waldmann H. Synthesis of an iridoid-inspired compound collection and the discovery of autophagy inhibitors. J Org Chem 2016;81:10242–55.
[42] Wang S, Fang K, Dong G, Chen S, Liu N, Miao Z, et al. Scaffold diversity inspired by the natural product evodiamine: discovery of highly potent and multitargeting antitumor agents. J Med Chem 2015;58:6678–96.
[43] Over B, Wetzel S, Grutter C, Nakai Y, Renner S, Rauh D, et al. Natural-product-derived fragments for fragment-based ligand discovery. Nature Chem 2013;5:21–8.
[44] Pascolutti M, Campitelli M, Nguyen B, Pham N, Gorse A-D, Quinn RJ. Capturing nature's diversity. PLoS One 2015;10(4):e0120942. http://dx.doi.org/10.1371/journal.pone.0120942.
[45] Yu B, Zheng Y-C, Shi X-J, Qi P-P, Liu H-M. Natural product-derived spirooxindole fragments serve as privileged substructures for discovery of new anticancer agents. Anticancer Agents Med Chem 2016;16:1–10.
[46] Hela KY, Maciejewski M, Gregori-Puigjane E, Glick M, Wassermann AM. Public domain HTS fingerprints: design and evaluation of compound bioactivity profiles from PubChem's bioassay repository. J Chem Inf Model 2016;56:390–8.
[47] Ekins S, Siqueira-Neto JL, McCall L-I, Sarker M, Yadav M, Ponder EL, et al. Machine learning models and pathway genome data base for *Trypanosoma cruzi* drug discovery. PLoS Negl Trop Dis 2015;9(6):e0003878. http://dx.doi.org/10.1371/journal.pntd.0003878.
[48] Wawer MJ, Li K, Gustafsdottir SM, Ljopsa V, Bodycombe NE, Marton MA, et al. Toward performance-diverse small-molecule libraries for cell-based phenotypic screening using multiplexed high-dimensional profiling. Proc Natl Acad Sci U S A 2014;111:10911–16.

Please cite this article in press as: Spear KL, Brown SP. The evolution of library design: crafting smart compound collections for phenotypic screens, Drug Discov Today: Technol (2017), http://dx.doi.org/10.1016/j.ddtec.2017.05.001