What Do You Get from DNA-Encoded Libraries?

Alexander L. Satz*

Roche Pharma Research and Early Development (pRED), Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd, Grenzacherstrasse 124, CH-4070 Basel, Switzerland

ABSTRACT: Use of DNA-encoded libraries (DELs) in the pharmaceutical industry has rapidly increased. We discuss what to expect when you run a DEL screen and contemplate guidelines for library design. Additionally, we consider some visionary work and extrapolate to the future.
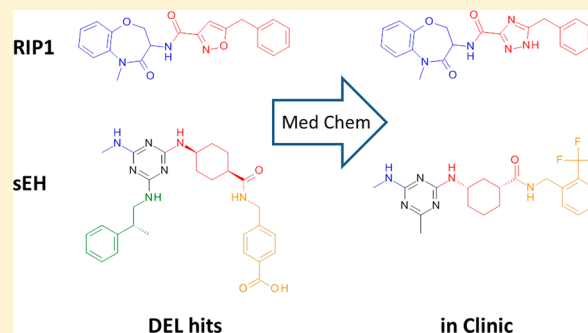
DNA-encoded library (DEL) screening, a modern iteration of combinatorial chemistry made possible by the invention of high-throughput DNA sequencing, has emerged as a robust tool for hit generation in the pharmaceutical industry.[1] Numerous real-world successes of the platform have been published, including progression of DEL-derived hits into the clinic (Figure 1).[2,3] Due to the nature of combinatorial
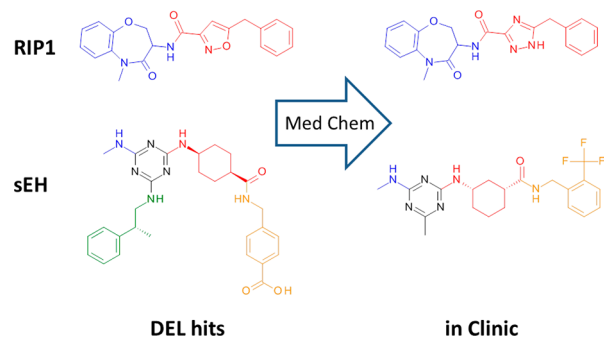

Figure 1. Molecules in the clinic are similar to their original DNA-encoded library (DEL) hits.

chemistry and the requirement of only attomoles of each molecule per screen, DELs succeed in providing a cost-effective method for expanding existing screening collections.[1,4]

DELs are generally synthesized employing split-and-pool combinatorial chemistry. A bifunctional precursor, with functionality for both synthetic chemistry and DNA ligation, is split into thousands of discrete wells.[1] A unique building block and encoding DNA oligomer are sequentially added to each well, ensuring a covalent connection between the chemical structure and the DNA sequence. The discrete reactions are pooled, resplit, and the chemistry-tagging process repeated up to 4 times in total. Assuming split sizes of 10^3 per cycle of

chemistry, two cycles of chemistry would provide a library with numeric size of 10^6. Note that DEL hits often possess fewer cycles of chemistry than the library from which they are derived; thus, each DEL may also be considered to contain numerous truncated (or "nested") libraries.[5]

Numerous DELs, each containing millions of compounds, can be screened against a target by a single scientist in a matter of days;[1,4,5] the low cost of conducting a DEL screen is game-changing, as it allows for iterative experimentation and a low-barrier for hypothesis testing. No sophisticated assay development is needed, as the DEL screening protocol is essentially the same for every target. Various screening conditions may be simultaneously investigated at little additional cost as screens are run in parallel using robotics or multichannel pipettes. Some examples of screening conditions include addition of ATP and analogs, varying pH, and blocking of orthosteric sites with known inhibitors. Additionally, very little protein is required per screening condition (∼0.3 nmol). The screen consists of a single-pot affinity assay where library molecules that bind the target are enriched, while the others are washed away.[1,4] A DEL, once synthesized, should provide enough material for thousands of screens. For comparison, the reagent costs of a DEL screen (not including target protein) are equivalent to the total costs of having a single chemical compound synthesized (a few thousand dollars).

DEL-derived hit molecules possess physical properties similar to those derived from more traditional screening platforms.[5] Still, a goal of the DEL community is to provide more and superior hits. Analysis of DEL-derived hits shows that their physical properties (MW and LogP) fit a normal distribution (Figure 2),[5] and so one strategy for discovering
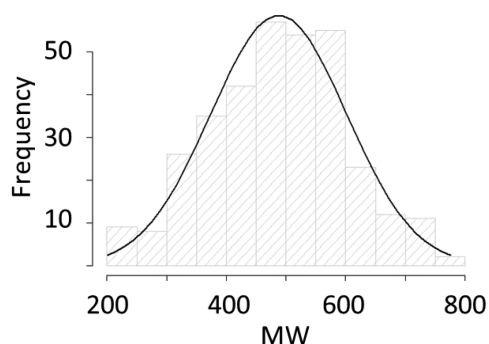
Figure 2. Molecular weights of all DEL-derived Roche hits.

higher quality hits is simply to discover more of them! A complementary strategy is to produce DELs with greater likelihood of producing hits with favorable properties (i.e., shift the distribution). In practice, this means the production of more libraries, containing more compact and polar molecules. Also, it entails the acquisition of more building blocks for library construction and the development of novel DNA-compatible chemistries[7,8] to join the building blocks together into lead-like structures.

Guidelines for library design are provided in the next few paragraphs, and exemplified by two DEL-derived molecules currently in the clinic, inhibitors of receptor interacting protein 1 (RIP1) kinase[2] and soluble epoxide hydrolase (sEH) (Figure 1).[3] We find that libraries synthesized using the most robust chemistries (and fewer chemical steps) are more productive,[5] as low synthetic yields during library production reduce signal-to-noise upon screening, resulting in a higher rate of false negatives.[9,10] The RIP1 hit consists of two building blocks joined together by an amide bond, while the sEH hit results from successive reactions of two amines with cyanuric chloride (a robust reaction with a long history of use in combinatorial chemistry) followed by acylation of a third building block. Note that the requirement of high synthetic yields for detection upon screening is a good thing; otherwise, any minor synthetic byproduct could produce a false positive!

We hypothesize that higher molecular weight (MW) building blocks provide chemical diversity and complexity otherwise lacking in libraries, and that efforts to optimize library design by excluding high MW building blocks are unwise. The RIP1 and sEH hits exemplify a common theme observed upon analysis of DEL screening data; hits often comprise two larger building blocks or three smaller ones.[1−5] The two building blocks that comprise the RIP1 hit have relatively high molecular weights (MW > 200 Da). In contrast, the sEH hit comprises three building blocks (not including the invariant 1,3,5-triazene core), all of which possess relatively low MW (<160 Da).

The discovery of structurally unique and "progressable" hits is achievable by employing simple and robust chemistries such as acylation and reductive amination.[3,5] Comparison of the RIP1 and sEH hits (Figure 1) against the CHEMBL[6] database (1.7 million compounds) reveal no instances of a Tanimoto similarity score >0.7. We reported similar results when comparing a number of our DEL-derived hits with the proprietary Roche compound collection.[5] As the chemistry used for joining building blocks together is often commonplace, we conclude that DEL diversity is highly dependent upon building block diversity. However, we note that there has been an incremental increase in the variety of DNA-compatible

chemistries available,[7] and further progress is expected due to increased academic (and NGO) interest in the technology.[8]

We believe that large split sizes are the reason why DELs today are more successful than combinatorial chemistry efforts of the past. DELs employing thousands of diverse building blocks at each cycle of chemistry are now routine. Employing large split sizes also differentiates DELs from technologies such as phage display. As discussed earlier, high cycle numbers are best avoided with DELs, as they inevitably result in lower synthetic yields and hence lower signal-to-noise upon screening.[9,10] Inspection of published DEL-derived hits suggests that libraries limited to 2−3 cycles of chemistry provide an optimum balance between structural diversity, synthetic yield, and ligand molecular properties.

Retrospective analysis of DEL-derived hits indicates that physical properties such as MW and LogP are impacted by choice of library scaffolding.[5] Again, the RIP1 and sEH hits are excellent exemplars. The scaffolding of RIP1 is essentially an amide bond that connects two building blocks together, yielding a hit with relatively low MW and LogP (377 Da and 2.5, respectively). In contrast, the sEH hit contains an invariant 1,3,5-triazene core, which adds both mass and hydrophobicity to all molecules derived from this DEL. The sEH hit is unsurprisingly heavier and more hydrophobic (517 Da and 5.0 LogP) than the RIP1 hit. An obvious strategy to improve the likelihood of discovering lead-like DEL hits is to employ compact and polar scaffolding.

Despite public claims of trillion-member DELs, there exists no peer-reviewed articles demonstrating the value of such libraries. We doubt there is a perfect numeric library size; however, a maximum of 10^8 seems reasonable. We recently illustrated how larger libraries suffer from lower signal-to-noise and higher rates of false negatives.[10] We also observe no correlation between library size and productivity.[5] Lastly, a literature survey reveals no reported hits from DELs with numeric size >10^9. Again, using the sEH and RIP1 hits as exemplars, these hits (which progressed to the clinic) were derived from truncated libraries containing <10^6 and <10^7 library members, respectively, and not trillion-member DELs. (Note that the library description for the RIP1 hit has not been published; however, its numeric size is likely <10^7, as it is derived from only two building blocks joined together.)

The most common application of the DEL technology remains screening purified soluble protein targets, and DEL screening protocols (as commonly employed in the pharmaceutical industry) appear to remain largely unchanged since the technology was first reduced to practice 15 years ago.[4] Still, some interesting potential advances have been reported including screening of detergent solubilized transmembrane proteins,[11] discovery of covalent inhibitors,[12] and screening of transmembrane targets overexpressed on cell surfaces.[13] However, while interesting, these potential advances also highlight a fundamental limitation of the technology as currently employed; biochemical and cellular screens are not possible.

The basic tools to conduct biochemical and cellular DEL screens already exist, although a screening platform robust enough for the pharmaceutical industry has yet to be reported. For instance, DELs can be synthesized on beads, encapsulated into picoliter aqueous droplets, and compounds are photochemically cleaved from the beads; label-free molecules may then interact freely with protein target or diffuse inside a cell. Active compounds are detected via fluorescent-activated

droplet sorting using a microfluidic circuit.[14] Another method for achieving spatial-separation of one-bead-one-compound DELs are ultralow-volume well arrays, which were employed 20 years ago to investigate cellular activity of bead-based libraries following photochemical cleavage.[15] Ironically, fabricated nanowells are commonly employed by DEL practitioners, but only for the purpose of high-throughput DNA sequencing.[16]

Use of DELs in the pharmaceutical industry has rapidly increased. The technology is often employed to complement high-throughput screening efforts and can be used whenever purified and soluble protein is accessible. DEL screens are especially useful when an assay for high-throughput screening is unavailable. Protein requirements and costs-per-additional DEL screen are minimal, resulting in a low-barrier to initiate target assessment. Low costs also allow for iterative experimentation with various screening conditions, protein constructs, and potentially disease-relevant protein complexes. The pharmaceutical industry requires superior methods to generate and screen chemical matter capable of modulating intracellular protein−protein interactions or bacterial growth. Improving cellular penetration for large molecules or putative antibacterials is difficult, and thus, hits with cellular activity are required for such chemical matter. Current visionary efforts aim for label-free DELs, development of spatially separated screens employing one-bead-one-compound DELs, and the ability to employ DELs in biochemical, bactericidal, and cellular assays.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: alexander_lee.satz@roche.com.

ORCID
Alexander L. Satz: 0000-0003-1284-1977

Notes
The author declares no competing financial interest.

■ REFERENCES

(1) Goodnow, R. A., Jr; Dumelin, C. E.; Keefe, A. D. DNA-encoded chemistry: enabling the deeper sampling of chemical space. *Nat. Rev. Drug Discovery* **2017**, *16*, 131−147.

(2) Harris, P. A.; Berger, S. B.; Jeong, J. U.; Nagilla, R.; Bandyopadhyay, D.; Campobasso, N.; Capriotti, C. A.; Cox, J. A.; Dare, L.; Dong, X.; Eidam, P. M.; Finger, J. N.; Hoffman, S. J.; Kang, J.; Kasparcova, V.; King, B. W.; Lehr, R.; Lan, Y.; Leister, L. K.; Lich, J. D.; MacDonald, T. T.; Miller, N. A.; Ouellette, M. T.; Pao, C. S.; Rahman, A.; Reilly, M. A.; Rendina, A. R.; Rivera, E. J.; Schaeffer, M. C.; Sehon, C. A.; Singhaus, R. R.; Sun, H. H.; Swift, B. A.; Totoritis, R. D.; Vossenkämper, A.; Ward, P.; Wisnoski, D. D.; Zhang, D.; Marquis, R. W.; Gough, P. J.; Bertin, J. Discovery of a First-in-Class Receptor Interacting Protein 1 (RIP1) Kinase Specific Clinical Candidate (GSK2982772) for the Treatment of Inflammatory Diseases. *J. Med. Chem.* **2017**, *60*, 1247−1261.

(3) Belyanskaya, S. L.; Ding, Y.; Callahan, J. F.; Lazaar, A. L.; Israel, D. I. Discovering Drugs with DNA-Encoded Library Technology: From Concept to Clinic with an Inhibitor of Soluble Epoxide Hydrolase. *ChemBioChem* **2017**, *18*, 837−842.

(4) Clark, M. A.; Acharya, R. A.; Arico-Muendel, C. C.; Belyanskaya, S. L.; Benjamin, D. R.; Carlson, N. R.; Centrella, P. A.; Chiu, C. H.; Creaser, S. P.; Cuozzo, J. W.; Davie, C. P.; Ding, Y.; Franklin, G. J.; Franzen, K. D.; Gefter, M. L.; Hale, S. P.; Hansen, N. J. V.; Israel, D. I.; Jiang, J.; Kavarana, M. J.; Kelley, M. S.; Kollmann, C. S.; Li, F.; Lind, K.; Mataruse, S.; Medeiros, P. F.; Messer, J. A.; Myers, P.; O'Keefe, H.; Oliff, M. C.; Rise, C. E.; Satz, A. L.; Skinner, S. R.; Svendsen, J. L.; Tang, L.; van Vloten, K.; Wagner, R. W.; Yao, G.; Zhao, B.; Morgan, B. A. Design, synthesis and selection of DNA-encoded small molecule libraries. *Nat. Chem. Biol.* **2009**, *5*, 647−654.

(5) Eidam, O.; Satz, A. L. Analysis of the Productivity of DNA Encoded Libraries. *MedChemComm* **2016**, *7*, 1323−1331.

(6) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. The ChEMBL bioactivity database: an update. *Nucleic Acids Res.* **2012**, *40*, D1100−1107.

(7) Satz, A. L.; Cai, J.; Chen, Y.; Goodnow, R.; Gruber, F.; Kowalczyk, A.; Petersen, A.; Naderi-Oboodi, G.; Orzechowski, L.; Strebel, Q. DNA compatible multistep synthesis and applications to DNA encoded libraries. *Bioconjugate Chem.* **2015**, *26*, 1623−1632.

(8) The Scripps Research Institute Signs Collaboration Agreement with Pfizer to Advance DNA-Encoded Library Technology. https://www.scripps.edu/news/press/2017/20170110TSRI_Pfizer.html (accessed Feb 23, 2018).

(9) Satz, A. L. DNA encoded library selections and insights provided by computational simulations. *ACS Chem. Biol.* **2015**, *10*, 2237−2245.

(10) Satz, A. L.; Hochstrasser, R.; Petersen, A. C. Analysis of Current DNA Encoded Library Screening Data Indicates Higher False Negative Rates for Numerically Larger Libraries. *ACS Comb. Sci.* **2017**, *19* (4), 234−238.

(11) Ahn, S.; Kahsai, A. W.; Pani, B.; Wang, Q. T.; Zhao, S.; Wall, A. L.; Strachan, R. T.; Staus, D. P.; Wingler, L. M.; Sun, L. D.; Sinnaeve, J.; Choi, M.; Cho, T.; Xu, T. T.; Hansen, G. M.; Burnett, M. B.; Lamerdin, J. E.; Bassoni, D. L.; Gavino, B. J.; Husemoen, G.; Olsen, E. K. Allosteric "beta-blocker" isolated from a DNA-encoded small molecule library. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (7), 1708−1713.

(12) Second Boston Symposium of Encoded Library Platforms. https://bselp.wordpress.com/agenda/ (accessed Feb 23, 2018).

(13) Wu, Z.; Graybill, T. L.; Zeng, X.; Platchek, M.; Zhang, J.; Bodmer, V. Q.; Wisnoski, D. D.; Deng, J.; Coppo, F. T.; Yao, G.; Tamburino, A.; Scavello, G.; Franklin, G. J.; Mataruse, S.; Bedard, K. L.; Ding, Y.; Chai, J.; Summerfield, J.; Centrella, P. A.; Messer, J. A.; Pope, A. J.; Israel, D. I. Cell-based selection expands the utility of DNA-encoded small molecule library technology to cell surface drug targets: Identification of novel antagonists of the NK3 tachykinin receptor. *ACS Comb. Sci.* **2015**, *17* (12), 722−731.

(14) MacConnell, A. B.; Price, A. K.; Paegel, B. M. An Integrated Microfluidic Processor for DNA-Encoded Combinatorial Library Functional Screening. *ACS Comb. Sci.* **2017**, *19*, 181−192.

(15) You, A. J.; Jackman, R. J.; Whitesides, G. M.; Schreiber, S. L. A miniaturized arrayed assay format for detecting small molecule-protein interactions in cells. *Chem. Biol.* **1997**, *4* (12), 969−75.

(16) Patterned Flow Cell Technology-Illumina Video. https://www.illumina.com/company/video-hub/pfZp5Vgsbw0.html (accessed Feb 23, 2018).