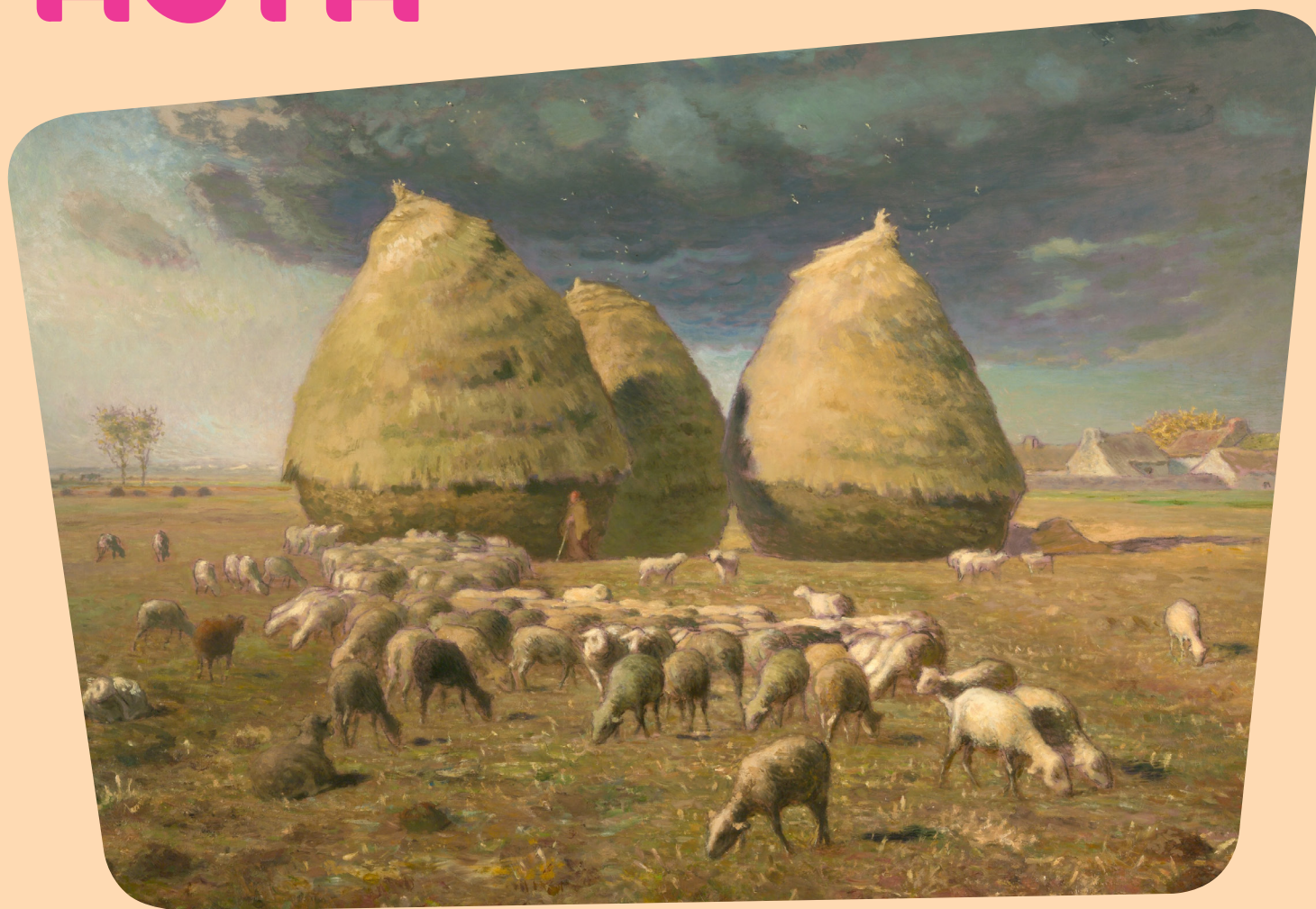**MERCK**

ALDRICHIMICA ACTA

DNA-Encoded Libraries—Finding the Needle in the Haystack

DNA-Encoded Fragment Libraries: Dynamic Assembly,
Single-Molecule Detection, and High-Throughput Hit Validation

DNA-Encoded Library Chemistry: Amplification of Chemical Reaction
Diversity for the Exploration of Chemical Space

DEAR READER:

We, at Merck, value the spirit of discovery. Since 2006, the Bader Award in Synthetic Organic Chemistry has recognized outstanding research from graduate students all over the world. This year, four eminently deserving young scientists were selected to receive the award and were invited to present their winning research at the 2019 Bader Student Chemistry Symposium in Milwaukee, WI, on September 12, 2019.
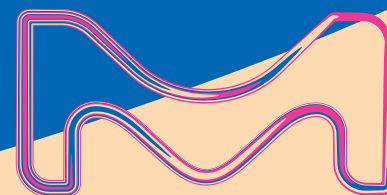
Michael Crocker, of Vanderbilt University, gave a presentation on "The Halo-Amino-Nitro Alkane Functional Group: A Platform for Reaction Discovery". He explained, "From an organic chemistry perspective, the next advancement I expect to fundamentally impact the way we put molecules together is C–H activation. Just as cross-coupling revolutionized synthetic planning, C–H activation will give access to more complex and difficult chemical space with implications in the particularly exciting areas of sustainable energy materials and personalized medicine."

Lucas Hernandez, of the University of Illinois at Urbana-Champaign, stated, "With the advent of machine learning in chemistry, the discovery of novel reactions will become more rapid over the next 100 years." His research on the "Synthesis of Isocarbostyril Alkaloids from Benzene" demonstrated how new compounds can be synthetically discovered and tested, providing molecular solutions to pressing challenges in oncology.

For **Joseph Dennis Jr**, of Massachusetts Institute of Technology, the need to mitigate the adverse environmental and economic effects of climate change on oceans should be paramount in order to protect the livelihood and food supply of millions of people that rely on fisheries. His research on "Breaking the Base Barrier: An Electron-Deficient Catalyst Enables the Use of Common, Soluble Bases in Pd-Catalyzed C–N Coupling" explores possible ways to reverse carbon dioxide absorption that is causing the current acidification of oceans while the world continues to work on eliminating the root causes.

Samantha Green, of Scripps Research, speculated that other variations of dual catalysis would significantly impact chemistry in the next 100 years. Her research on "Quaternary Centers via Dual-Catalytic Alkene Hydroarylation" identified the biocompatibility of MHAT catalysts, which could lead to the development of new synergistic methods for metal–enzyme dual catalysis or artificial metalloenzymes.

In addition to the graduate students, Dr. Joseph R. Clark, of Marquette University, gave the keynote presentation on his research on "Site-Selective Copper-Catalyzed Deuterium Incorporation into Small Molecules".

For more information on this year's event, visit **SigmaAldrich.com/BaderAward.**

"PLEASE BOTHER US."

Dear Fellow Chemists,

Professor Donald A. Watson of the Department of Chemistry and Biochemistry at the University of Delaware kindly suggested that we offer (bathocuproine)NiBr$_2$ (**902985**), which is used as part of a novel and general catalytic system for the C-alkylation of nitroalkanes with unactivated alkyl iodides at 40 °C under mild reaction conditions. The system exhibits excellent functional-group tolerance; works with primary, secondary, and tertiary alkyl iodides; and offers easy access to a diversity of complex nitroalkanes in moderate-to-good yields. The system's effectiveness was demonstrated in a two-step synthesis of the antiviral drug adapromine in 52% yield.

Rezazadeh, S.; Devannah, V.; Watson, D. A. *J. Am. Chem. Soc.* **2017**, *139*, 8110.
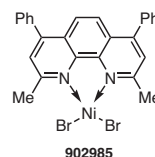
902985	(Bathocuproine)NiBr$_2$	250 mg

We welcome your product ideas. Do you need a product that is not featured on our website? Ask us! For more than 60 years, your research needs and suggestions have shaped our product offering. Email your suggestion to **techserv@sial.com**.

Udit Batra, Ph.D.
CEO, Life Science

TABLE OF CONTENTS

ABOUT OUR COVER

Perhaps it is auspicious to be featuring a field with huge haystacks on the cover of this issue of the *Acta*, which is shining light on the rapidly evolving research field of DNA-Encoded Chemical Libraries (DECLs). Until recently, the search for "hits" in DECLs has seemed like looking for the proverbial needle in the haystack. But, today, researchers, armed with more effective and selective synthetic methods and more sensitive analytical tools, have become more adept at finding promising leads in their DECLs—to the benefit of all.

Jean-François Millet (1814–1875), a member of the Barbizon School of nature painters, finished *Haystacks: Autumn* (oil on canvas, 85.1 x 110.2 cm) ca. 1874. Believed to be one of his last paintings, it was commissioned by a patron as part of a series depicting the four seasons.* The painting depicts an end-of-harvest scene most likely painted in Millet's workshop based on sketches he made outdoors of fields near Barbizon where he was living. The

Detail from *Haystacks: Autumn*. Photo courtesy The Metropolitan Museum of Art, New York, NY.

autumnal colors, the impending storm, and the afternoon sun perhaps foreshadow Millet's death less than a year later. Millet's work was a significant influence on later 19-century European painters such as van Gogh and Seurat.

This painting is a bequest of Lillian S. Timken to the Metropolitan Museum of Art, New York, NY.

** To find out about the other three paintings in the series, visit **SigmaAldrich.com/Acta***
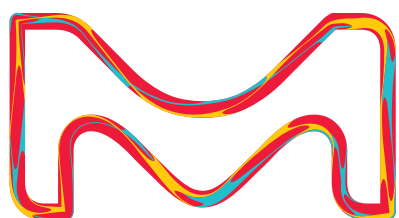
DNA-Encoded Fragment Libraries: Dynamic Assembly, Single-Molecule Detection, and High-Throughput Hit Validation

Dr. F. V. Reddavide

Dr. M. Thompson

Dr. L. Mannocci

Prof. Y. Zhang

Francesco V. Reddavide,[a] Michael Thompson,[a] Luca Mannocci,[b] and Yixin Zhang[*,c]

[a] DyNAbind GmbH, Tatzberg 47, 01307 Dresden, Germany
[b] DECLTech Consulting, Birkenweg 13, 5074 Eiken, Switzerland
[c] B CUBE - Center for Molecular Bioengineering, Technische Universität Dresden, Tatzberg 41, 01307 Dresden, Germany
   Email: yixin.zhang1@tu-dresden.de

Keywords. DNA-encoded chemical library (DECL); self-assembly; screening; selection; fragment-based drug discovery; bidentate; protein; inhibitor; drug discovery; sequencing; DNA origami.

Abstract. An urgent challenge in chemistry and biotechnology is to develop a routine, robust, and cost-effective method for the identification of molecules that specifically bind to a large variety of protein targets. In recent years, an elegant selection method for small-molecule drug discovery, DNA-encoded chemical library (DECL) technology, has been receiving much attention from the pharmaceutical and biotechnology industries. Here, we review the major recent developments in DECL technology, with a focus on the self-assembling dual-display format, which aims to combine the bio-inspired selection process with a fragment-based approach to discover potent binders to protein targets.

Outline

1. Introduction

1.1. Challenges in Drug Discovery

Cancer is not one disease; rather, it is many different diseases that develop in various organs and are associated with a myriad of genes and their mutations, as well as numerous environmental factors. The same is also true for many other diseases such as autoimmune disease or dementia. Specific diagnoses and treatments have been developed for specific diseases, and those have become not only increasingly more efficient, but also more sophisticated, with the ultimate goal of developing truly personalized medicine. Target complexity combined with an ageing population and the looming expiration of blockbuster patents have driven an unprecedented development effort for novel drug discovery technologies to complement existing high-throughput screening (HTS) platforms.[1-2] One such technology is DNA-Encoded Chemical Library (DECL) synthesis and selection technology. Originally proposed by Brenner and Lerner in 1992,[3]

the last 27 years have seen this idea grow from a "back-of-the-envelope scribble" into an increasingly mature and robust tool for hit discovery.[4a]

A DECL is a pooled collection of up to billions of different molecules, each of which is tagged with a unique DNA strand that functions primarily as an identifying barcode. In a standard DECL compound selection, a target protein is immobilized on a solid support and treated with the library, followed by washing to remove non-binding library members. This is followed by elution of binding moieties, which are then amplified via a Polymerase Chain Reaction (PCR) and sequenced by Next Generation Sequencing (NGS) to reveal the identities of the binding compounds by ranking the sequences according to their frequency. The identified hits are then quantitatively validated in follow-up experiments using an orthogonal technology—e.g., enzyme inhibition assay, isothermal titration calorimetry, fluorescence polarization, or surface plasmon resonance—to determine binding affinity and/or biological potency (e.g., K_d, IC_{50}, EC_{50}) and thus prioritize the further development of the hits (**Figure 1**).[4b,c]

Over the past years, advances in DECL technologies have rendered the field too broad for a thorough treatment in only one review article. Therefore, we will focus our discussion in this paper on fragment-based DECL approaches. Generally speaking, fragment-based discovery has become increasingly popular because of the belief that the small size and the potential high ligand efficiency of the molecular fragments can offer advantageous starting points for drugging traditionally difficult pharmaceutical targets, such as protein–protein interactions. Bringing a fragment-based approach to DECL technology can help minimize the potential weaknesses of fragment screening (e.g., solubility), while the use of dual-DNA-display approaches allows the mimicking of evolutionary principles, such as dynamics[5] and recombination principles.[6]

1.2. A Brief History of DECL

The original design of DECL was largely inspired by phage display technology.[3] The first proof-of-principle library was generated through parallel synthesis of peptide and DNA on polystyrene resin in a one-bead one-compound format.[7] While this experiment was crucial to DECL development, the beads utilized were much larger than bacteriophages (up to 10^5–10^6-fold; >1000-fold larger than M13 bacteriophage), setting a practical limit on potential library size, as well as precluding selection experiments with the bio-panning protocol employed in phage display technology. With the bead library, the first proof-of-concept affinity-based selection was performed using a peptide-DNA conjugate library and a fluorescently labelled target protein, and identification of the interaction between the peptide and its antibody was accomplished by Fluorescence Activated Cell Sorting (FACS).[7] Until the beginning of the new millennium, the development and application of DECL technology proceeded very slowly, limited mainly by the lack of sufficiently robust auxiliary technologies. Then, supported by improvements in DNA microarray, DNA conjugation chemistry,[8-12] DNA-templated chemical reactions,[8,13-15] and especially by the introduction of DNA NGS,[16,17] DECL technology experienced a rapid and impressive growth in both academic and industrial settings.[4,18] The introduction of split-and-pool synthesis and enzymatic tag-encoding ushered in the use of a combinatorial assembly approach to quickly prepare libraries of massive sizes, currently on the order of billions or trillions of compounds.[19,20] When DECLs became comparable with phage display libraries in terms of size and biopanning protocol compatibility, their advantages
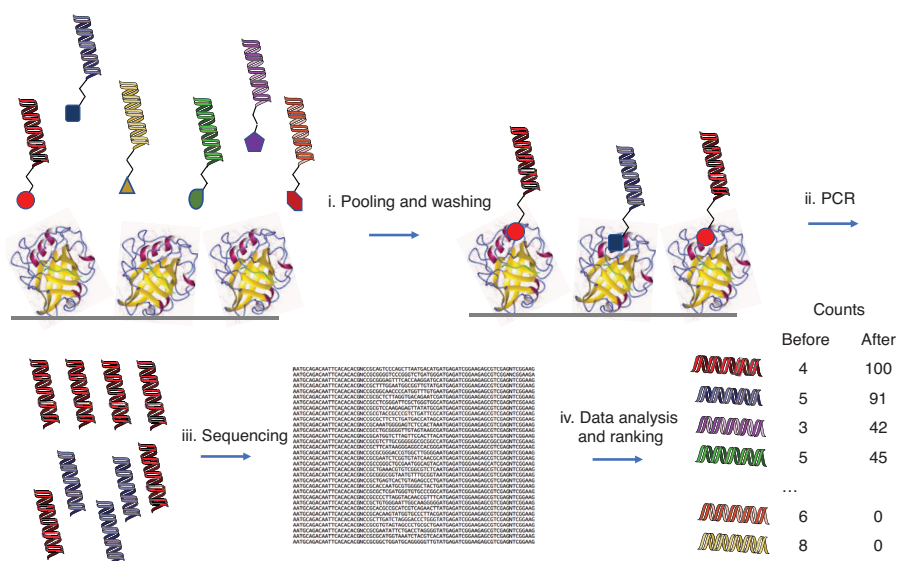

Figure 1. Common Protocol for DECL Compound Selection Experiments. *(Ref. 4b,c)*

became obvious: the run time for a DECL experiment is largely independent of library size, making the process significantly faster than a traditional HTS campaign. Indeed, DECL-based drug discovery methodologies continue to mature and have, in recent years, proven their worth by discovering new drug candidates for a number of diverse targets.[21–24]

2. Selection vs Screening

Four major methods are employed in chemical genomics and drug discovery to identify new pharmaceutically relevant molecular entities (**Figure 2**): (i) *High-throughput screening (HTS)*: This is currently the dominant method for the discovery of small-molecule drug compounds, whereby each chemical compound is assayed one-by-one in well-plate formats.[25–27] (ii) *Array-based method*: In this complementary method to high-throughput screening, chemical compounds are spotted on a planar surface as an addressable library, while the binding of the target protein can be directly detected by using labeled proteins or antibodies.[28,29] (iii) *Selection*: In nature, high-affinity binders, such as antibodies, are generated through selection processes, wherein a number of potential binders are competing simultaneously. The same principle is also used in bioengineering (e.g., protein/peptide display, aptamer technology) to develop macromolecule-based therapeutics.[30,31] (iv) *Structure-based design*: This is a rational approach that commonly focuses on developing potent binders by combining information from fragment screening with protein structural information and structure–activity relationships in multivalent protein–ligand interactions.[32–34]

Each method has its advantages and disadvantages. For example, drug screening can make use of a great variety of chemical structures, but it is expensive and requires assay development for each protein target. Conversely, the array-based method does not need any particular activity-based assay development, but library size is a limitation. Selection technologies allow for the use of larger libraries and can be routinely applied to polypeptides and nucleic acids (e.g., antibodies and aptamers). Therefore, DECL has expanded the scope of the selection approach to small-molecule chemical libraries. However, although the affinity-based assay can lead to high-affinity binders, the hit compounds may not have the desired biological properties or functions. The fragment-based approach is the most rational, but it is also the most time- and labor-intensive method.

HTS, array-based screening, and fragment-based approaches test chemical compounds one-by-one, causing campaign costs to scale with library size. Selection approaches offer the advantage of being relatively independent of library size. Moreover, all compounds compete with each other under the selection conditions, thus reflecting their thermodynamic differences in binding to the target protein. While DECL can be thought of as primarily a selection technology, it borrows concepts from most of the other approaches.[12,35] Similar to screening, the focus in the selection approach is traditionally on small organic molecules or fragments, while recent new developments have also incorporated elements of array technology or structure-based design, as will be discussed in later parts of this review.

The use of selection-based methodologies in the discovery of small-molecule binders was traditionally constrained by the sensitivity of the analytical tools available to identify the selected compounds. Prior to the widespread use of DECL, the best known approach was Novartis's SpeedScreen,[36] which selects binders for a protein target from a mixture of 100–600 compounds using mass spectrometry for the screening
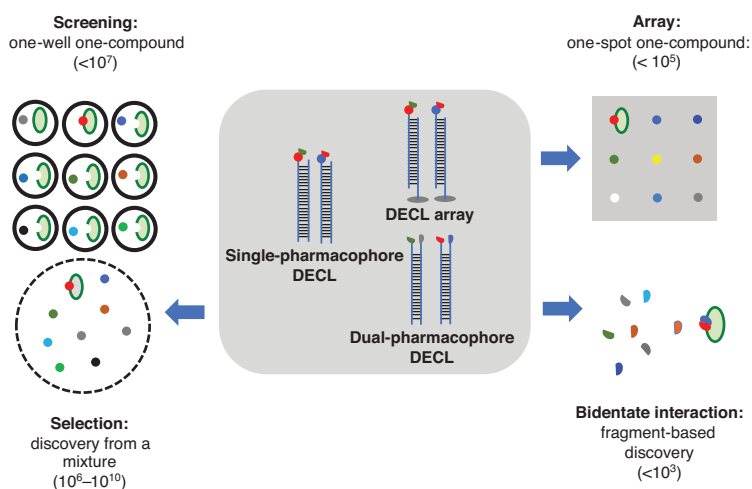
Figure 2. DNA Sequences as Barcodes for Small-Molecule Compounds. DECLs Can Be Designed and Synthesized in Various Forms to Mimic Different Drug Discovery Technologies (Middle). By Conjugating DNA Sequences as Barcodes to Small Organic Molecules, This Technology Allows the Application of Selection Methods to Small-Molecule Ligand Discovery, Which Is Traditionally Done by HTS (Left). DECLs Can Also Be Generated as an Array, and Displayed as Fragment Pairs, Mimicking Small-Molecule Array and Fragment-Based Drug Discovery, Respectively (Right). DECL Selection and DECL-Derived Technology Represent Attractive Alternatives to HTS Technology. *(Ref. 25–34)*

66

DNA-Encoded Fragment Libraries: Dynamic Assembly, Single-Molecule Detection, and High-Throughput Hit Validation
Francesco V. Reddavide, Michael Thompson, Luca Mannocci, and Yixin Zhang*

read-out. Libraries of up to 600,000 compounds (in 96-well plate format, with 400 compounds in each well) have been successfully investigated by SpeedScreen.

Since then, DECLs have been a revolutionary selection technology for small-molecule compounds. Because DNA offers a reasonably broad information space and can be PCR-amplified, the detection limit issues of other analytical chemical methods (e.g., the 600 compounds limit for SpeedScreen using mass spectrometry) can be largely overcome. Today, DECLs of billions of compounds are routinely screened in selection experiments.

2.1. Main Challenges in DECL

With the reports of ever-growing DECL sizes, drug discovery researchers have realized that the major challenges are not only about the number of library members.[37] For example, in order to generate large libraries, increasing the number of synthetic cycles (i.e., the number of combinatorial split-and-pool steps) increases the mass and lipophilicity of the member compound. Even three cycles of combinatorial synthesis with fragment-like building blocks can result in library members moving well outside the bounds of traditional drug-likeness (e.g., Lipinski's rules).[38]

A potentially greater problem, however, is that combinatorial synthesis prevents meaningful quality control or assessment of library contents. While chromatography can be used to clean up the coupling of stage 1 building blocks to DNA, this is no longer feasible once the initial reaction products are pooled. Over additional synthetic cycles, this increases the possible presence of side products or unreacted precursors, resulting in a single DNA tag encoding a number of different compounds and complicating the effort to identify and validate hits.[39] This issue can snowball in very large libraries and manifest itself as false negative results, as the individual copy number of a given molecule can dip below detectable levels.[40] Another limitation of DECL is that many common chemical reactions in the medicinal chemist's toolbox are not compatible with DNA. While progress has recently been made in broadening the synthetic methodology for preparing DECLs[41-43]—especially with the DNA-templated synthesis of small-molecule macrocycles pioneered in the lab of David Liu[8,44]—amide-bond formation nevertheless remains one of the most reliable reactions for DECL construction. Through repetition of coupling steps, relatively high yields with small variations among the building blocks can be achieved.[45]

In contrast to the high-fidelity DNA synthesis by DNA polymerase in the replication process, both the chemical synthesis and sequencing of DNA are error-prone. For the construction of DECL, in addition to chemical library synthesis errors caused by the previously discussed side reactions and low reactivity of some of the building blocks, errors in the coding sequence may also be introduced during the DNA-tagging step. The affinity-based selection protocols with immobilized target proteins on polymer matrices can likewise cause artifacts such as promiscuous interactions and protein misfolding. Eventually, the PCR and NGS decoding steps can similarly introduce further biases.

Data analysis by ranking the detected sequence counts is largely based on the assumption that the binding affinity to a native protein—the amount of DNA-conjugate small molecule captured, the PCR amplification, and the NGS reads—are linearly correlated and mostly error-free. Unfortunately, errors and biases accumulate over the whole DECL workflow, resulting in reduced correlations between the measured binding affinities of the selected compounds and the NGS-decoding counts. While DECLs with more sophisticated structures have been constructed in recent years, this has exacerbated the problem by adding more synthetic and encoding steps. Moreover, another undesired consequence of increasing library size is the inability of NGS to provide sufficient sequencing depth for comprehensive statistical analysis. Together with the challenges associated with individual synthetic and biochemical reactions, statistical requirements must also be taken into consideration in the analysis of the NGS-decoding results in order to statistically identify significantly enriched hit compounds.[46]

Despite the significant progress made in DECL technology in recent years, the main challenges have remained mostly unchanged, and researchers are still working toward two common goals: (i) adding new chemical reactions to the DECL synthesis tool box, not only for increasing library purity and size, but also for generating pools of compounds with higher structural diversity and drug likeness; and (ii) improving library design and selection protocol in order to achieve a superior correlation between the sequence counts and the subsequently measured affinities, thus reducing the occurrence of false positives, and, more importantly, false negatives.

2.2. Optimizing the Selection Conditions, from Surface to Solution

The standard DECL selection protocol uses immobilized target proteins on polymer matrices;[47] this protocol can cause artifacts such as promiscuous interaction and protein misfolding. Most commonly, a protein of interest is either immobilized on cyanogen bromide (CNBr)-activated Sepharose® beads or captured on streptavidin-Sepharose® after protein biotinylation. The use of magnetic beads has enabled automated selections, while non-covalent immobilization allows researchers to evaluate the effect of protein modification on protein activity prior to immobilization.[22] Alternative strategies such as His-tag, FLAG-tag, STREP-TAG®, and GST-fusion can also be employed for immobilization, but biotinylation remains the most favorable method because of the strong and reliable immobilization through streptavidin–biotin interaction which prevents protein dissociation.

In order to minimize the artifacts caused by polymer matrices, a two-step selection protocol has been developed.[48] A DECL is first incubated with the target protein in solution. After the system reaches equilibrium, resin is added to capture the protein and the DECL compounds bound to it. Eventually, the resin is washed and the binding library is recovered either by heat-induced denaturation of the protein or chemical elution. To completely remove the effect associated with the protein immobilization process, Li's group has reported a DECL design

which allows selection against unmodified protein targets in solution.[49] A library of single-strand DECL was hybridized to a DNA carrying a photo-crosslinking group. After target-protein binding and irradiation, the protein–ligand–DNA complex was isolated and submitted to decoding by NGS sequencing.

A number of valuable drug targets are membrane-bound proteins that don't lend themselves to convenient solid-support immobilization. To access this target space, GlaxoSmithKline developed and reported a cell-based selection approach using DECL.[50] Tachykinin receptor neurokinin-3 (NK3), a G-protein coupled receptor, was expressed on HEK293 cells. After optimizing the expression condition, cells expressing approximately 5×10^5 receptors per cell (the highest level that could be achieved with this expression system for NK3) were utilized in the subsequent selection experiments. Remarkably, several hits identified in this study have potency, specificity, and ligand efficiency properties comparable to talnetant and osanetant, two antagonists developed for the same membrane protein through classical screening and medicinal chemistry approaches.[50]

3. Conventional Chemical Library vs Fragment-Based Approach

Similarly to the one-well one-compound format in HTS, most DECLs are in the form of one-DNA one-compound (1_d–1_c), constructed as combinatorial chemical libraries conjugated to a single DNA strand. An alternative approach is to take advantage of DNA hybridization in dual-display technology, where annealed DNA strands are used to present two moieties simultaneously. The two-DNA two-compound (2_d–2_c) approach resembles another drug discovery method: the fragment-based one.[12,35] The (2_d–2_c)-DECL approach was pioneered by Neri's group at ETH Zürich by using a straightforward setup of two DNA strands containing a universally complementary annealing region.[12] After synthesis, the two sub-libraries are mixed, resulting in the random and stable pairing of the molecules. The 2_d–2_c format has an obvious advantage for the construction of large but high-purity libraries: Whereas it is practically impossible to purify every compound in a 1_d-1_c library with 1 million members, it is possible to assemble a 10^6-member library by the pairing of two 10^3-member, DNA-encoded sub-libraries that have been individually HPLC purified and characterized by mass spectrometry.

The majority of existing DECL platforms, however, are still based on a 1_d–1_c architecture. This is at least partially due to the relatively recent development of 2_d–2_c libraries, but one cannot ignore the fact that the two DNA-conjugated moieties must at some point in the process be taken off the DNA and resynthesized as a single compound, which is not a trivial undertaking. On the other hand, the 2_d–2_c format offers a significant, "hidden" increase in library diversity: the fragments in this format possess larger rotational and translational freedoms to interact with a protein, as opposed to the rigidly assembled compounds present in 1_d–1_c DECL or HTS collections. The chemical structures in 2_d–2_c libraries are linked to DNA strands at either the 3' or 5' terminus through flexible linkers, typically

composed of C_3–C_{12} aliphatic chains or short PEGs (polyethylene glycols). This puts fewer restrictions on structure-based drug design and medicinal chemistry to explore different spatial arrangements and conformational spaces. While a systematic test of different linking systems can then be pursued in later stages of hit development, this is also where fragment-based DECLs can fit together nicely with a structure-based design approach by taking advantage of crystallographic information about the target. For the design of a fragment library—provided that binding enthalpy can offset the loss of rigid-body entropy—chemical building blocks are chosen, privileging structures that can form thermodynamically favorable interactions with the target protein.[32,51] Future developments in fragment-based DECLs could also provide additional context for fragment interactions. For example, if different linkers could be employed in generating 2_d–2_c libraries with the structural information of the linkers also encoded in the DNA sequences, it would be possible to get simultaneously both more reliable selection results (e.g., whether same pair of fragments are selected with different linkers) and complementary information regarding the preferred distance between the binding fragments.

The inhibition of protein–protein interactions (PPIs) represents an increasingly prominent goal for drug developers, but also one where traditional methods such as HTS have had only limited success. Fragment-based drug discovery (FBDD) has in the past two decades emerged as an alternative to HTS and, after 20 years of gradual improvement, it is now accepted as a part of mainstream drug discovery. As of 2016, more than 30 drug candidates derived from fragments have entered the clinical phase, with two approved and several more in advanced trials.[51] Research from multiple groups is supporting the idea that the combination of DECL and FBDD approaches is a logical next step for both of these technologies. For example, while weakly binding fragments can be missed in classical fragment screening experiments due to assay sensitivity, their chances of successful detection are improved in 2_d–2_c DECLs. This is due to the bidentate interaction in which a pair of weak binders assembled by a DNA duplex will exhibit a higher cooperative affinity than when displayed as individual fragments.[22] Furthermore, a common problem with FBDD—the often limited solubility—is overcome by conjugation with highly soluble DNA.

3.1. Developments in Fragment-Based DECL

As discussed in the previous section, some challenges with 2_d–2_c libraries arise because multiple moieties are presented for target binding. Selection with a 2_d–2_c fragment library does not *directly* lead to a list of fragment pairs revealed by their corresponding DNA codes, because in a 2_d–2_c library both chemical fragments and their corresponding DNA codes are separated on two different DNA strands. Therefore, in the early days of fragment DEL development, the fragments identified from sequencing had to then be combinatorially assembled and tested to reveal ideal pairings.[52] To circumvent such a cumbersome workflow, most 2_d–2_c fragment libraries were designed in an affinity-maturation format,[12,22] in which one

68

DNA-Encoded Fragment Libraries: Dynamic Assembly, Single-Molecule Detection, and High-Throughput Hit Validation
Francesco V. Reddavide, Michael Thompson, Luca Mannocci, and Yixin Zhang*

of the two displayed strands was exclusively conjugated to a known but relatively weak binder to the target protein. In 2015, Neri, Scheuermann, and co-workers disclosed an elegant inter-strand code-transfer method that results in a 2_d–2_c library in which one of the two DNA strands is able to host the coding information of both pairing fragments (**Figure 3**).[53]

As shown in Figure 3, sub-library B was constructed in a way that allowed hybridization with sub-library A and, at the same time, the transfer of B coding sequences onto the corresponding A strand by means of a fill-in reaction. With this strategy, the A strand for each library member would contain sequence information that would unambiguously identify the A–B fragment pair, while the B strand would not be amplifiable by PCR. Using this approach, the authors have reported the identification of a low-micromolar binder to alpha-1-acid glycoprotein (AGP) and the affinity maturation of a ligand to carbonic anhydrase IX, a well-known marker of carcinoma of kidney cells.[53]

4. Self-Assembled Dynamic DECL

A particularly clever innovation in drug discovery screening is Dynamic Combinatorial Chemistry (DCC). A library using the DCC approach consists of a pool of small-molecule building blocks that can interconnect via reversible reactions that form transient adducts in a thermodynamic equilibrium (e.g., disulfide bond and Schiff base). Each building block possesses one or more functional groups that allow dynamic combinatorial interactions with other members of the library. In this pool, new chemical species are continuously formed and dissociated following the reversible equilibrium of the DCC reaction. When a target protein is introduced into the system, the equilibrium can be shifted toward the formation of higher-affinity binders. The protein thus acts as a template for the synthesis in situ of the binders through thermodynamic stabilization of the otherwise unstable adducts.
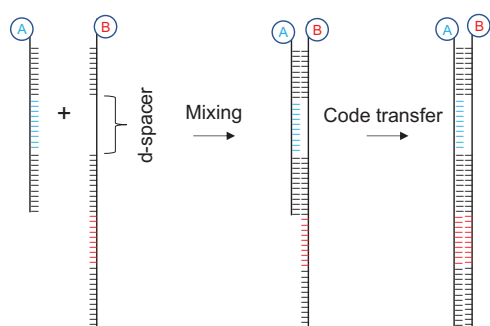

Figure 3. Library Construction with Code-Transferring Mechanism. The Syntheses of Sub-Libraries A (Light Blue) and B (Red) Were Carried Out Separately. A DNA Polymerase Assisted Fill-in Reaction Allowed the B Code to Be Transferred onto the Sub-Library A Strand. The Final Library Was Used for Selection Experiments against Target Proteins of Choice. Subsequent PCR Amplified Only the A Strand, and High-Throughput DNA Sequencing Simultaneously Revealed Binding Fragment Pairs. *(Ref. 53)*

The development of DCC technology was limited by the available analytical tools, such as mass spectrometry and NMR, which can analyze only mixtures of compounds with very low complexity. In contrast, techniques capable of analyzing complex mixtures of oligonucleotides, like microarray technologies and the more recent NGS, are much more powerful. These novel analytical technologies can in principle resolve a complex mix of thousands or millions of different oligonucleotide sequences in a single analysis in a few hours.[18,54,55] We thus saw an opportunity to fuse the efficient and simple self-assembling and "self-screening" features of DCC with the extremely sensitive hit-identification methods of DECL, allowing the generation of very large dynamic libraries.

As the utility of DCC in drug discovery also represents a fragment-based approach, it was relatively straightforward to design a DCC-based DECL in the 2_d–2_c format.[53] The first "DNA-Encoded Dynamic Combinatorial Chemical Library" (EDCCL) was constructed in a dynamic 2_d–2_c format including a hybridization domain of 4 to 8 nucleotides. These short domains are naturally unstable, driving only transient pairing of the chemical moieties conjugated to the DNA construct, in analogy to the reverse reactions in DCC. All the chemical moieties from one sub-library are free to pair with all the other members of the other sub-library. Fragment-pair binding to a target protein can overcome the weak pairing of the hybridization domain, stabilizing higher affinity pairs and shifting the binding equilibrium. One of the advantages of a dynamic library over its static counterpart is that different chemical species are continuously synthesized in the selection mixture, while only high-affinity binders are "locked" on the protein. If we think of a pair of high-affinity fragments as the "right combination" of the "right fragments", then most "right fragments" in a static library are in the "wrong combinations", i.e., paired with nonbinding fragments. In a dynamic library, the right fragments in wrong combinations can always be unpaired and thus free to search for another fragment for pairing, and so on until stabilized by the protein after forming the "right combination". In this way, a high signal-to-noise ratio can be achieved by sorting out a number of nonbinding fragments. EDCCL can also be deployed for ligand optimization (similarly to the previously discussed "affinity maturation"). In this approach, one of the sub-libraries is composed by a single member, a known binder. When the known binder is in low amounts with respect to the pairing (sub)-library, it is "forced" to form the "right combination" with the "right partner", thus enhancing the signal-to-noise ratio of the "right" to "wrong" partner. While this has obvious applications for improving binding strength, it can just as effectively be employed for improving target specificity.

The 2_d-2_c dynamic format pays an energy cost for de-stabilizing the tertiary complex of two DNA strands and one protein, as the weak interaction between the two strands can weaken the chelation effect. Among the different hybridization domains tested, the 4 and 5 base-pair (bp) domains are not sufficiently stable to guarantee the enrichment of high-affinity binders, while the 6 and 8 bp domains show a much higher enrichment of the high-affinity pairs compared to a static counterpart (e.g., 18 bp

hybridization domain, with no compound reshuffling possible). To reduce the energy barrier associated with short dynamic hybridization domains, "Heat Induced-EDCCL" technology (hi-EDCCL) has been developed.[56] Here, the dynamic equilibrium is established by exposing the unbound library fraction to a temperature higher than the hybridization domain melting point, while the protein-loaded matrix is kept at relatively low temperature (achieved via physical separation) to prevent unfolding of the protein. Heat denaturation followed by a DNA re-annealing step cause pair reshuffling and the rearrangement of the library into a collection of new pairs, some of which are high-affinity binders. The reshuffled library is then reapplied onto the immobilized protein to allow the newly generated pairs to bind to the target. Shuffling and selection can be repeated several times to increase the signal-to-noise ratio and stringency of selection. In order to overcome reproducibility issues and the critical loss of material during the manual handling of liquids at the microliter scale, a dedicated hi-EDCCL automated microfluidic device has been developed. This device is capable of controlling the continuous liquid transfer and the temperature in different sections of the pathway.[57]

More recently, we have developed a second-generation EDCCL, which utilizes a Y-shaped DNA construct for the creation of a third DNA strand—at the end of the selection process—containing the codes of both binding fragments to reveal binder-pairing information (**Figure 4**).[58] A challenge in developing this technology was that Y-shaped constructs exhibit different melting behavior than linear duplexes. Linear constructs begin to melt from the extremities, while Y-shaped constructs have the tri-junction as the point of instability. For this reason, larger hybridization domains, 11 and 13 bp, were optimal for building a Y-shaped EDCCL. At the end of the selection process, the two coding regions of the binding fragments captured on the immobilized protein were linked by means of DNA ligation. Second-generation Y-shaped EDCCL are also capable of minimizing the signal noise caused by unpaired single-fragment binding. Indeed, when a fragment binds to the target protein without previously forming a pair, it can neither be ligated to a partner fragment nor PCR-amplified later, thus becoming "invisible" to the NGS decoding.

Li and co-workers have developed another format of EDCCL that featues two hybridization domains of 6 and 7 base pairs, one of which also carries a psoralen photo-crosslinker.[59] Following incubation of the dynamic library with a soluble target protein, UV irradiation triggers the photo-crosslinking of the two DNA strands, thus "freezing" the dynamic exchange at the binding
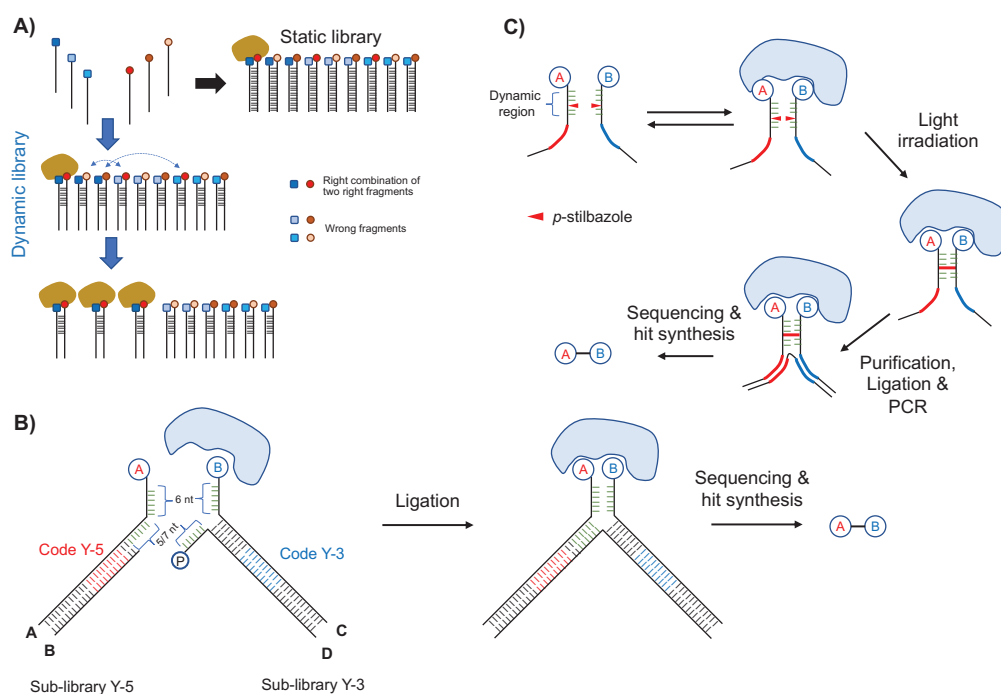
Figure 4. (A) Principle of Dynamic DECL. For 2d-2c Fragment Libraries with n x n Combinations, When the Library Is in a Static Form, a Certain "Right Combination" of Two "Right Fragments" Is of $1/n^2$ of the Entire Library, Independently of the Presence or Absence of a Protein Target. When the Library Is in a Dynamic Form, the Thermodynamic Equilibrium Will Shift the "Right Combination" from $1/n^2$ to $1/n$ of the Entire Library in the Presence of the Target Protein, through Shuffling Between the Different Combinations. **(B) Construction of Dynamic DECL with Code Ligation Mechanism.** Strands A and B Are Assembled to Form Sub-Library Y-5. Strands C and D Make Up Sub-Library Y-3. Upon Binding to the Target Protein, Phosphorylated Strand D Can Be Ligated to Strand B. **(C) Dynamic DECL with Photo-Crosslinking.** Two Sets of DNA-Encoded Small Molecules with a Short Complementary Dynamic Region Form the DNA-Encoded Dynamic Library. Target Addition Shifts the Equilibrium, Promoting the Formation of High-Affinity Duplexes. UV Irradiation Locks the Shifted Equilibrium through *p*-Stilbazole (B-Base)-Mediated Photo-Crosslinking. Crosslinked Duplexes Are Isolated, and the Sequences at the Encoding Sites Are Decoded for Hit Identification. *(Ref. 58).*

70

DNA-Encoded Fragment Libraries: Dynamic Assembly, Single-Molecule Detection, and High-Throughput Hit Validation
Francesco V. Reddavide, Michael Thompson, Luca Mannocci, and Yixin Zhang*

equilibrium. Finally, the photo-crosslinked oligonucleotides are purified via gel electrophoresis, and then sequenced. In 2018, the same laboratory reported further development of this concept.[60] To improve the photo-crosslinking efficacy, the psoralen photo-crosslinker was substituted by two facing *p*-stilbazoles photo-crosslinkers in the center of the dynamic hybridization domain. *p*-Stilbazole does not disturb the DNA base pairing when located in the middle of the oligonucleotide sequence. Moreover, the ability to crosslink the oligonucleotide in the center of the sequence permitted the joining of the two codes of the binding fragments into one sequence. Since a DNA polymerase cannot go through the unnatural linkage of the oligonucleotides, a relay-primer bypass strategy was adopted. This strategy features the use of a primer to mask the unnatural linking region, while a DNA polymerase and ligase assemble the final amplicon, which eventually can be PCR-amplified and subjected to NGS.

5. Hit Validation

While a number of research groups have made significant advances in areas related to library architecture and DNA-compatible chemistry, relatively little progress has been achieved in downstream hit validation. The most common hit-validation approach has been to immediately resynthesize the compounds off-DNA for follow-up testing. It is worth noting however, that modern DECL's increasing library sizes and potentially large hit lists would require a substantial synthetic effort and cause this step to be a bottleneck in the discovery process.

Hit validation can be carried out with a conventional protein-activity assay in solution. However, since DECL discovery works well even for targets where only limited knowledge is available, a functional assay (e.g., enzyme inhibition) may not be available. In this case, a number of kinetic or thermodynamic binding assays can be utilized. For example, a fluorescently labelled off-DNA compound can be synthesized, and the binding constant can be measured with such techniques as fluorescence polarization or microscale thermophoresis.[61] While microcalorimetry offers thermodynamic characterization of protein–ligand interactions, avoiding artifacts associated with labeling and surface immobilization, the consumption of materials and the labor effort involved are high compared to other methods.[62]

One of the most appealing methods for characterizing protein-ligand interactions is biosensor-based technology, which is able to simultaneously provide thermodynamic information about the dissociation constants (K_d) as well as the characteristic kinetic parameter values K_{on} and K_{off} that are associated with the binding process. Unfortunately, this approach is challenging for high-throughput hit validation of small-molecule compounds, since small-molecule covalent immobilization on solid support is time-consuming and cost-inefficient. Moreover, utilizing small molecules as binders in the mobile phase frequently leads to very weak signals and poor data quality. Neri, Scheuermann, and co-workers have reported the validation of on-DNA resynthesized hits by fluorescence polarization, Alphascreen® technology, and

microscale thermophoresis.[61] For dual-display approaches, hit validation is doubly problematic. To prevent false negatives from incorrect linker selection, off-DNA synthetic efforts need to include a number of different linking groups, again presenting an intimidating synthetic task for any hit list containing more than a small number of molecular structures.[22,52,58]

To increase the throughput of hit validation, Lin et al. have described the use of a regenerable biosensor chip for characterizing on-DNA small-molecule compounds (**Figure 5**).[63] The concept is simple: A DNA handle for binding library members is covalently attached to the biosensor surface. On-DNA hit compounds are then loaded by annealing onto this handle, subjected to protein binding measurement, and then chemically stripped to regenerate the DNA handle for characterizing the next round of hit compounds. The loading and regeneration procedure can be reiterated over 20 cycles without losing the signal intensity. This permits the automated measurement of kinetic profiles of protein–ligand interactions for DECL hits. Moreover, this method can be adapted to characterize pairs of compounds, greatly alleviating the concerns over the synthetic effort required for dual-fragment hit validation. While the most promising hit pairs must still be resynthesized and linked off-DNA, this approach allows an initial triage in the same chemical context in which the hits were originally registered, reducing the risk of false negative results.[63]

6. Application of Nanotechnology in DECL

As discussed previously, the synthesis of structurally diversified high–purity, drug-like libraries and the development of new selection procedures with superior signal-to-noise ratios that minimize the rates of false positives or false negatives represent the two major challenges for DECL. While library size is currently not considered as the limiting factor for DECLs, the development of new chemical and biochemical methods for DECL will inevitably lead to the production of larger libraries. Indeed, NGS has revolutionized the field of DECL. However, to cope with libraries of growing sizes, sequencing depth (i.e., the ratio between the number of reads and library size) must be similarly improved. Therefore, sequencing power may become soon a new bottleneck. Because of the resolution limitation associated with the fluorescence-based detection mechanism, as well as the need to prevent over-clustering,[64] NGS will soon meet its limitation of 1 sequence/μm^2 (or $10^8/cm^2$). Therefore, the ultimate limitation of DECL in the current setup is neither the chemical space nor the DNA sequences, but rather the optical resolution.

As many modern imaging techniques (e.g., atomic force microscopy (AFM), super-resolution fluorescence microscopy, and scanning tunneling microscopy) are not limited by optical resolution, Kielar et al. made use of the available DNA nanotechnology tool box to construct nanoscale pharmacophore arrays (**Figure 6**).[65] Several different pharmacophores were displayed with nanometer precision on DNA origami substrates either as individual ligands or as fragment pairs, and their binding to different model proteins was evaluated by AFM as a

single-molecule detection method. This work demonstrated the successful detection of several different binding events including strong binding, weak binding, symmetric bidentate binding, and asymmetric bidentate binding. This method was further applied to the discovery of bidentate trypsin binders based

on the pairing of the weak trypsin inhibitor benzamidine with different aromatic fragments.[65] The combination of benzamidine with the fluorescent dye carboxytetramethylrhodamine (TAMRA) results in about a 10-fold enhancement of the trypsin binding yield compared to that from benzamidine alone. Since AFM can
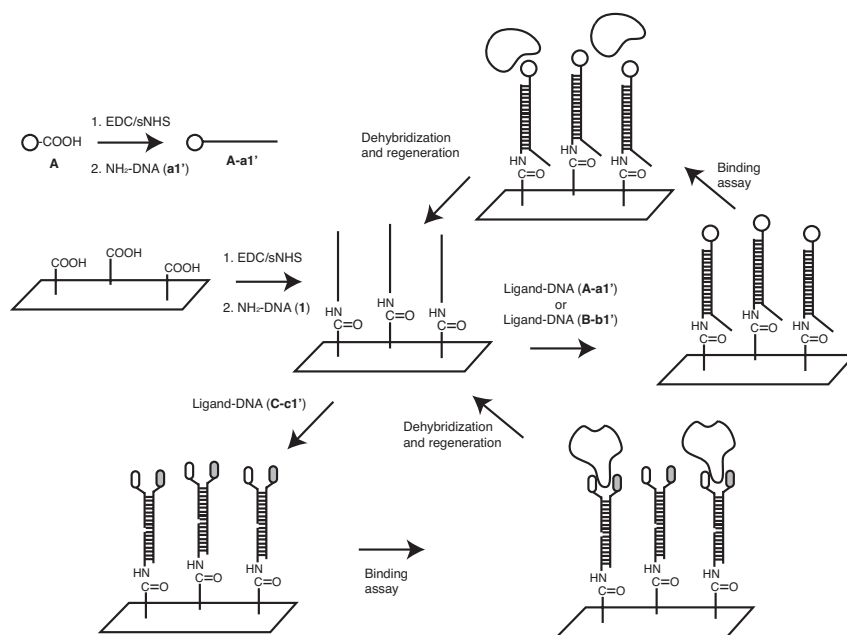

Figure 5. Principle of Affinity Measurement on a Regenerable Biosensor Chip. The Carboxylic Acid Groups on the Chip Were Coupled with the Amino Groups of 5'-Amino-DNA (1) (Middle Cycle). A Small-Molecule Ligand (A) Was Conjugated to the Amino Group of 5'-Amino-DNA (a1') That Is Complementary to Sequence 1. After Annealing A-a1' to 1 on the Chip, the Target Protein Is Injected at Different Concentrations to Assess the Binding Affinity. A-a1' Can Be Washed off the Chip under Dehybridization/Regeneration Conditions. Another DNA-Ligand Conjugate, B-b1', Can Then Be Immobilized on the Regenerated Chip, and Its Binding Affinity to the Same or a Different Protein Can Be Evaluated (Upper Cycle). This Method Can Also Be Applied to a Double-Strand DNA, C-c1', Displaying Two Ligands (Lower Cycle). *(Ref. 63)*
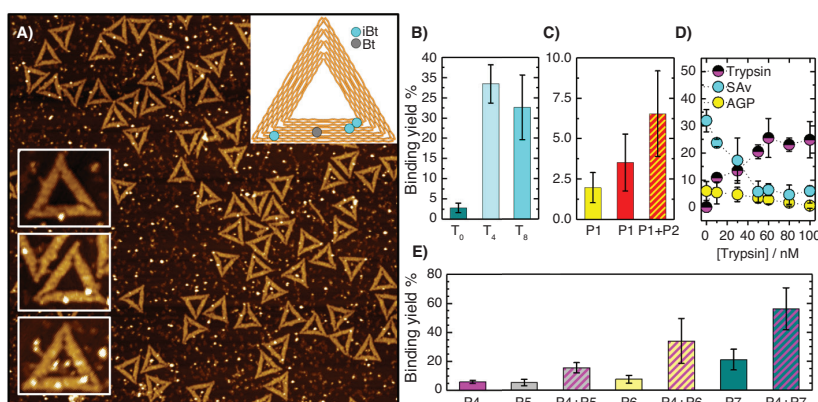

Figure 6. DNA-Origami-Based Nanoarrays for Studying Protein–Ligand Interactions. (A) Representative AFM Image of Iminobiotin (Ibt) Nanoarrays Taken after Incubation with Streptavidin (SAv). The Image Size and Height Scale Are 2 x 2 μm^2 and 4.2 nm, Respectively. The Upper Right Inset Depicts the Positions of the Modified Staple Strands. A Single Biotin (Bt) Modification Is Used to Distinguish Between the Two Different iBt Sites. The Insets on the Left Are Zoom-ins of Single DNA-Origami Substrates Exhibiting (i) No SAv Bound to iBt, (ii) Bidentate Binding of SAv to iBt, and (iii) Monodentate and Bidentate Binding of SAv to iBt. **(B)** Bidentate SAv-iBt Binding Yields for Different Thymine Spacer Lengths. **(C)** Binding Yields for Mono- and Bidentate Binding of AGP to Pharmacophore Ligands P1 (Propenamide Derivative) and P2 (Acetamide Derivative). **(D)** Binding Yields for Bidentate SAv-iBt Binding, Monodentate AGP-P1 Binding, and Bidentate Trypsin Binding to Ligands P3 (3-Iodophenyl Isothiocyanate) + P4 (4-Aminobenzamidine) in a Single Nanoarray as a Function of Trypsin Concentration. **(E)** Binding Yields for Mono- and Bidentate Trypsin Binding to Fragments P4 to P7 (TAMRA). *(Ref. 65)*

72

DNA-Encoded Fragment Libraries: Dynamic Assembly, Single-Molecule Detection, and High-Throughput Hit Validation
Francesco V. Reddavide, Michael Thompson, Luca Mannocci, and Yixin Zhang*

distinguish two DNA-origami-bound proteins with a distance of about 10 nm, these pharmacophore nanoarrays could achieve spatial resolutions 5–7 orders of magnitude higher than the conventional microarray technologies on glass slides, which have feature sizes of 10–100 µm. Given that no modifications or immobilization of the target proteins is required, the method can detect the binding of pharmacophores in their native forms. Moreover, the DNA-origami substrates provide a platform to display DNA-modified ligands with variable spatial arrangements, and thus represent a versatile tool for fragment-based lead discovery research.

The chemical library on DNA origami has only been demonstrated in proof-of-principle experiments, while many challenges still remain for its real application in drug discovery. However, with the ever-increasing library sizes, single-molecule-based detection methods will inevitably be needed in the future. In addition to ultra-high-resolution imaging techniques such as AFM,[66] the emerging nanopore sequencing[67] is a very attractive technology: Using nanopore sequencing, a single molecule of DNA or RNA can be sequenced without PCR amplification in a label-free fashion. Although nanopore sequencing has higher error rates than NGS based on chromatin immunoprecipitation sequencing (ChIP-seq) technology, DNA codes can be designed to possess high error tolerance,[68] which would make nanopore sequencing particularly suitable for large DECLs.

7. Conclusion and Outlook

The continued, rapid advances in library synthesis methodology and library size beg the question: Is there a limit to how large a library can be? Answering this question literally, one could calculate that, in a 1 mL DECL solution with a total compound concentration of 1 µM, there would be room for more than 10^{15} individual molecules—a number that our current synthetic methods won't get us anywhere close to. A more prudent question might be: Is there a limit to how large a useful library can be? To answer these questions, Samain and co-workers have recently reported the quantitative evaluation of affinity selection performance using both quantitative PCR (qPCR) and NGS techniques.[40] Interestingly, for compounds with sub-µM dissociation constants, selection performance drops if 10^4 copies per library member are used as the input. This result implies a size limit of DECL in the range of 10 to 100 billion different molecules, to ensure that each potential hit has more than 10^4 copies. Whether the limit of 10^4 copies can be circumvented remains to be tested in the future, probably through implementation of single-molecule techniques to assess the interactions at a single-molecule level, or by employing new generations of sequencing methods with higher sequencing depth, or by applying advanced statistical tools in data analysis.[37,46,69–71]

Dynamic DECL technology represents another possibility to circumvent the size limit. In a 1 mL solution of dynamic DECL with two sub-libraries, each with a total compound concentration of 0.5 µM, there exist about 10^{29} different possible combinations. When the 10^4-copies rule is applied to both sub-libraries, the theoretical size limit for a dynamic DECL is 10^{21}-member

compound pairs. While in the absence of protein, each of these pairs has a near-zero representation, the presence of a protein will shift the thermal dynamic equilibrium, generating these high-affinity pairs on demand. The triplex library with three pharmacophores (3_d–3_c) that was suggested by Neri and co-workers during the early development of self-assembled DECL has so far not been practically implemented.[12] Although such a library would present challenges in later hit linkage and re-synthesis, if the 3_d–3_c library could be realized in a dynamic format, the number of possible structure combinations would be astronomical.

As we've aimed to convey in this review, DECL research also focuses on questions more complex than just library size. For standard libraries in the 1_d–1_c format, improving the library quality, as well as developing new DNA-compatible chemical reactions for library synthesis, represent two major challenges for chemistry. In combination with novel biochemical assays, e.g., in-solution or on-cell selection, the goal has been to achieve a good correlation between the sequencing outcome and the measured binding affinity of different hit compounds. For fragment libraries in the 2_d–2_c format, DECL can become a very powerful tool for fragment-based drug discovery, which has a great potential to overcome the shortage of screening-based methods for tackling difficult drug targets such as protein–protein interactions. However, similarly to all fragment-based approaches, the question of how to link the fragments and generate high-affinity binders remains the most daunting challenge. Because of the large size of the compound library, DECL technology allows us to collect unprecedented amounts of information. When the binding of fragments presented on a DNA duplex to target a protein can be elucidated in crystallographic analysis, valuable information can be provided for structure-based drug design. Our ability to effectively use this information, to obtain new insights into the structure–activity relationship, and design potent drug compounds will benefit from new technologies and concepts such as structural biology and artificial intelligence.

8. References

(1)　Paul, S. M.; Mytelka, D. S.; Dunwiddie, C. T.; Persinger, C. C.; Munos, B. H.; Lindborg, S. R.; Schacht, A. L. *Nat. Rev. Drug Discovery* **2010**, *9*, 203.

(2)　Khanna, I. *Drug Discovery Today* **2012**, *17*, 1088.

(3)　Brenner, S.; Lerner, R. A. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 5381.

(4)　(a) Goodnow, R. A., Jr.; Dumelin, C. E.; Keefe, A. D. *Nat. Rev. Drug Discovery* **2017**, *16*, 131. (b) Zhang, Y. Hit Identification and Hit Follow-Up. In *A Handbook for DNA-Encoded Chemistry: Theory and Applications for Exploring Chemical Space and Drug Discovery;* Goodnow, R. A., Jr., Ed.; Wiley: Hoboken, NJ, 2014; pp 357-376. (c) Goodnow, R. A., Jr., Ed. *A Handbook for DNA-Encoded Chemistry: Theory and Applications for Exploring Chemical Space and Drug Discovery*; Wiley: Hoboken, NJ, 2014.

(5)　Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *Drug Discovery Today* **2002**, 7, 117.

(6) Market, E.; Papavasiliou, F. N. *PLoS Biol.* **2003**, *1*, 24.

(7) Needels, M. C.; Jones, D. G.; Tate, E. H.; Heinkel, G. L.; Kochersperger, L. M.; Dower, W. J.; Barrett, R. W.; Gallop, M. A. *Proc. Nat. Acad. Sci. U. S. A.* **1993**, *90*, 10700.

(8) Gartner, Z. J.; Tse, B. N.; Grubina, R.; Doyon, J. B.; Snyder, T. M.; Liu, D. R. *Science* **2004**, *305*, 1601.

(9) Halpin, D. R.; Harbury, P. B. *PLoS Biol.* **2004**, *2*, 1015.

(10) Halpin, D. R.; Harbury, P. B. *PLoS Biol.* **2004**, *2*, 1022.

(11) Halpin, D. R.; Lee, J. A.; Wrenn, S. J.; Harbury, P. B. *PLoS Biol.* **2004**, *2*, 1031.

(12) Melkko, S.; Scheuermann, J.; Dumelin, C. E.; Neri, D. *Nat. Biotech.* **2004**, *22*, 568.

(13) Gartner, Z. J.; Liu, D. R. *J. Am. Chem. Soc.* **2001**, *123*, 6961.

(14) Calderone, C. T.; Puckett, J. W.; Gartner, Z. J.; Liu, D. R. *Angew. Chem., Int. Ed.* **2002**, *41*, 4104.

(15) Gartner, Z. J.; Kanan, M. W.; Liu, D. R. *Angew. Chem., Int. Ed.* **2002**, *41*, 1796.

(16) Margulies, M.; Egholm, M.; Altman, W. E.; Attiya, S.; Bader, J. S.; Bemben, L. A.; Berka, J.; Braverman, M. S.; Chen, Y.-J.; Chen, Z.; Dewell, S. B.; Du, L.; Fierro, J. M.; Gomes, X. V.; Godwin, B. C.; He, W.; Helgesen, S.; Ho, C. H.; Irzyk, G. P.; Jando, S. C.; Alenquer, M. L. I.; Jarvie, T. P.; Jirage, K. B.; Kim, J.-B.; Knight, J. R.; Lanza, J. R.; Leamon, J. H.; Lefkowitz, S. M.; Lei, M.; Li, J.; Lohman, K. L.; Lu, H.; Makhijani, V. B.; McDade, K. E.; McKenna, M. P.; Myers, E. W.; Nickerson, E.; Nobile, J. R.; Plant, R.; Puc, B. P.; Ronan, M. T.; Roth, G. T.; Sarkis, G. J.; Simons, J. F.; Simpson, J. W.; Srinivasan, M.; Tartaro, K. R.; Tomasz, A.; Vogt, K. A.; Volkmer, G. A.; Wang, S. H.; Wang, Y.; Weiner, M. P.; Yu, P.; Begley, R. F.; Rothberg, J. M. *Nature* **2005**, *437*, 376.

(17) Mannocci, L.; Zhang, Y.; Scheuermann, J.; Leimbacher, M.; De Bellis, G.; Rizzi, E.; Dumelin, C.; Melkko, S.; Neri, D. *Proc. Nat. Acad. Sci. U. S. A.* **2008**, *105*, 17670.

(18) Neri, D.; Lerner, R. A. *Annu. Rev. Biochem.* **2018**, *87*, 479.

(19) Kinoshita, Y.; Nishigaki, K. *Nucleic Acids Symp. Ser.* **1995**, *34*, 201.

(20) Buller, F.; Mannocci, L.; Zhang, Y.; Dumelin, C. E.; Scheuermann, J.; Neri, D. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5926.

(21) 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. *Nat. Chem. Biol.* **2009**, *5*, 647.

(22) Melkko, S.; Zhang, Y.; Dumelin, C. E.; Scheuermann, J.; Neri, D. *Angew. Chem., Int. Ed.* **2007**, *46*, 4671.

(23) Buller, F.; Steiner, M.; Frey, K.; Mircsof, D.; Scheuermann, J.; Kalisch, M.; Bühlmann, P.; Supuran, C. T.; Neri, D. *ACS Chem. Biol.* **2011**, *6*, 336.

(24) Clark, M. A. *Curr. Opin. Chem. Biol.* **2010**, *14*, 396.

(25) Hann, M. M.; Oprea, T. I. *Curr. Opin. Chem. Biol.* **2004**, *8*, 255.

(26) Inglese, J.; Johnson, R. L.; Simeonov, A.; Xia, M.; Zheng, W.; Austin, C. P.; Auld, D. S. *Nat. Chem. Biol.* **2007**, *3*, 466.

(27) Macarron, R.; Banks, M. N.; Bojanic, D.; Burns, D. J.; Cirovic, D. A.; Garyantes, T.; Green, D. V. S.; Hertzberg, R. P.; Janzen, W. P.; Paslay, J. W.; Schopfer, U.; Sittampalam, G. S. *Nat. Rev. Drug Discovery* **2011**, *10*, 188.

(28) Novoa, A.; Machida, T.; Barluenga, S.; Imberty, A.; Winssinger, N. *ChemBioChem* **2014**, *15*, 2058.

(29) Winkler, D. F.; Campbell, W. D. *Methods. Mol. Biol.* **2008**, *494*, 47.

(30) Smith, G. P.; Petrenko, V. A. *Chem. Rev.* **1997**, *97*, 391.

(31) Ellington, A. D.; Szostak, J. W. *Nature* **1990**, *346*, 818.

(32) Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. *Drug Discovery Today* **2003**, *8*, 876.

(33) Hajduk, P. J.; Greer, J. *Nat. Rev. Drug Discovery* **2007**, *6*, 211.

(34) Joseph-McCarthy, D.; Campbell, A. J.; Kern, G.; Moustakas, D. *J. Chem. Inf. Model.* **2014**, *54*, 693.

(35) Daguer, J.-P.; Zambaldo, C.; Ciobanu, M.; Morieux, P.; Barluenga, S.; Winssinger, N. *Chem. Sci.* **2015**, *6*, 739.

(36) Muckenschnabel, I.; Falchetto, R.; Mayr, L. M.; Filipuzzi, I. *Anal. Biochem.* **2004**, *324*, 241.

(37) Satz, A. L.; Hochstrasser, R.; Petersen, A. C. *ACS Comb. Sci.* **2017**, *19*, 234.

(38) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Delivery Rev.* **2001**, *46*, 3.

(39) Eidam, O.; Satz, A. L. *MedChemComm* **2016**, *7*, 1323.

(40) Sannino, A.; Gabriele, E.; Bigatti, M.; Mulatto, S.; Piazzi, J.; Scheuermann, J.; Neri, D.; Donckele, E. J.; Samain, F. *ChemBioChem* **2019**, *20*, 955.

(41) Škopić, M. K.; Willems, S.; Wagner, B.; Schieven, J.; Krause, N.; Brunschweiger, A. *Org. Biomol. Chem.* **2017**, *15*, 8648.

(42) Ding, Y.; Franklin, G. J.; DeLorey, J. L.; Centrella, P. A.; Mataruse, S.; Clark, M. A.; Skinner, S. R.; Belyanskaya, S. *ACS Comb. Sci.* **2016**, *18*, 625.

(43) Satz, A. L.; Cai, J.; Chen, Y.; Goodnow, R. A., Jr.; Gruber, F.; Kowalczyk, A.; Petersen, A.; Naderi-Oboodi, G.; Orzechowski, L.; Strebel, Q. *Bioconjugate Chem.* **2015**, *26*, 1623.

(44) Usanov, D. L.; Chan, A. I.; Maianti, J. P.; Liu, D. R. *Nat. Chem.* **2018**, *10*, 704.

(45) Leimbacher, M.; Zhang, Y.; Mannocci, L.; Stravs, M.; Geppert, T.; Scheuermann, J.; Schneider, G.; Neri, D. *Chem.—Eur. J.* **2012**, *18*, 7729.

(46) Buller, F.; Zhang, Y.; Scheuermann, J.; Schäfer, J.; Bühlmann, P.; Neri, D. *Chem. Biol.* **2009**, *16*, 1075.

(47) Decurtins, W.; Wichert, M.; Franzini, R. M.; Buller, F.; Stravs, M. A.; Zhang, Y.; Neri, D.; Scheuermann, J. *Nat. Protoc.* **2016**, *11*, 764.

(48) Cuozzo, J. W.; Centrella, P. A.; Gikunju, D.; Habeshian, S.; Hupp, C. D.; Keefe, A. D.; Sigel, E. A.; Soutter, H. H.; Thomson, H. A.; Zhang, Y.; Clark, M. A. *ChemBioChem* **2017**, *18*, 864.

(49) Zhao, P.; Chen, Z.; Li, Y.; Sun, D.; Gao, Y.; Huang, Y.; Li, X. *Angew. Chem., Int. Ed.* **2014**, *53*, 10056.

(50) 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. *ACS Comb. Sci.* **2015**, *17*, 722.

(51) Erlanson, D. A.; Fesik, S. W.; Hubbard, R. E.; Jahnke, W.; Jhoti,

H. *Nat. Rev. Drug Discovery* **2016**, *15*, 605.

(52) Scheuermann, J.; Dumelin, C. E.; Melkko, S.; Zhang, Y.; Mannocci, L.; Jaggi, M.; Sobek, J.; Neri, D. *Bioconjugate Chem.* **2008**, *19*, 778.

(53) Wichert, M.; Krall, N.; Decurtins, W.; Franzini, R. M.; Pretto, F.; Schneider, P.; Neri, D.; Scheuermann, J. *Nat. Chem.* **2015**, *7*, 241.

(54) Goodnow, R. A., Jr. *SLAS Discovery* **2018**, *23*, 385.

(55) Ottl, J.; Leder, L.; Schaefer, J. V.; Dumelin, C. E. *Molecules* **2019**, *24*, 1629.

(56) Reddavide, F. V.; Lin, W.; Lehnert, S.; Zhang, Y. *Angew. Chem., Int. Ed.* **2015**, *54*, 7924.

(57) Grünzner, S.; Reddavide, F. V.; Steinfelder, C.; Cui, M.; Busek, M.; Klotzbach, U.; Zhang, Y.; Sonntag, F. Lab-on-a-Chip Platform for High Throughput Drug Discovery with DNA-Encoded Chemical Libraries. In *Microfluidics, BioMEMS, and Medical Microsystems XV*, Proceedings of the SPIE BiOS Conference, San Francisco, CA, January 28–February 2, 2017; Gray, B. L., Becker, H., Eds.; Society of Photo-Optical Instrumentation Engineers (SPIE), February 28, 2017; Volume 10061, Paper 1006111 (https://doi.org/10.1117/12.2253840).

(58) Reddavide, F. V.; Cui, M.; Lin, W.; Fu, N.; Heiden, S.; Andrade, H.; Thompson, M.; Zhang, Y. *Chem. Commun*. **2019**, *55*, 3753.

(59) Li, G.; Zheng, W.; Chen, Z.; Zhou, Y.; Liu, Y.; Yang, J.; Huang, Y.; Li, X. *Chem. Sci.* **2015**, *6*, 7097.

(60) Zhou, Y.; Li, C.; Peng, J.; Xie, L.; Meng, L.; Li, Q.; Zhang, J.; Li, X. D.; Li, X.; Huang, X.; Li, X. *J. Am. Chem. Soc.* **2018**, *140*, 15859.

(61) Zimmermann, G.; Li, Y.; Rieder, U.; Mattarella, M.; Neri, D.; Scheuermann, J. *ChemBioChem* **2017**, *18*, 853.

(62) Freyer, M. W.; Lewis, E. A. *Methods Cell Biol*. **2008**, *84*, 79.

(63) Lin, W.; Reddavide, F. V.; Uzunova, V.; Gur, F. N.; Zhang, Y. *Anal. Chem.* **2015**, *87*, 864.

(64) Ma, Z.; Lee, R. W.; Li, B.; Kenney, P.; Wang, Y.; Erikson, J.; Goyal, S.; Lao, K. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 14320.

(65) Kielar, C.; Reddavide, F. V.; Tubbenhauer, S.; Cui, M.; Xu, X.; Grundmeier, G.; Zhang, Y.; Keller, A. *Angew. Chem., Int. Ed.* **2018**, *57*, 14873.

(66) Piontek, M. C.; Roos, W. H. *Methods Mol. Biol.* **2018**, *1665*, 243.

(67) Van Dijk, E. L.; Jaszczyszyn, Y.; Naquin, D.; Thermes, C. *Trends Genet.* **2018**, *34*, 666.

(68) Zhang, Y.; Herrmann, J.; Wieduwild, R.; Boden, A. Method for the Deconvolution of Nucleic Acid Containing Substance Mixtures. World Patent WO2015/104411A1, July 16, 2015.

(69) Satz, A. L. *ACS Chem. Biol.* **2015**, *10*, 2237.

(70) Satz, A. L. *ACS Comb. Sci.* **2016**, *18*, 415.

(71) Kuai, L.; O'Keeffe, T.; Arico-Muendel, C. *SLAS Discovery* **2018**, *23*, 405.

Trademarks. Alphascreen® (PerkinElmer, Inc.); Sepharose® (GE Healthcare Bio-Sciences AB); STREP-TAG® (IBA GMBH).

About the Authors

Francesco Reddavide is the Director of R&D at DyNAbind GmbH. He graduated in 2011 from the University of Bari (Italy) with an M.Sc. degree in Medical and Pharmaceutical Biotechnology. From 2012 to 2016, Francesco pursued his doctoral studies in the lab of Professor Yixin Zhang at TU Dresden, where he participated in the development of a regenerable chip for kinetic hit validation of DNA-encoded compounds. He is co-author of multiple patents and peer-reviewed publications about novel designs and applications of DELs and auxiliary technologies. In 2015, he joined DyNAbind's EXIST-Forschungstransfer development team to lead development efforts aimed at taking DyNAbind's technology from proof-of-principle to market-ready status.

Mike Thompson grew up in Florida and studied biochemistry at the University of Florida, where he worked on organic ligand synthesis for organometallic clusters in the laboratory of Professor George Christou. In 2011, he moved to Germany for his doctoral work in the lab of Professor Yixin Zhang. After completing his Ph.D. degree in 2015, Mike led a team to secure EXIST-Forschungstransfer funding and scale their DECL technology for industrial applications, leading to the founding of DyNAbind GmbH in 2017, where he now serves as CEO.

Luca Mannocci was born in 1979 in Pisa, Italy, and received his B.S. degree in chemistry from the University of Pisa. In 2008, Luca earned his Ph.D. degree in chemistry under the supervision of Professor Dario Neri at the Institute of Pharmaceutical Sciences of ETH Zürich (Switzerland). After leading a drug discovery unit at Philochem AG, Luca has worked as independent consultant on the development and use of DECLs for drug and ligand discovery applications for major pharmaceutical, life science, and biotech companies as well as academic groups. In addition to over 15 years of experience in DECL technologies, Luca has co-founded DyNAbind GmbH and authored several key publications on DECL-technology in peer-reviewed journals and in drug discovery and DECL-dedicated books.

Yixin Zhang is a professor at Technische Universität Dresden (TU Dresden), Germany. He studied organic chemistry in Shanghai, China, and did his doctorate studies on immunosuppressive drugs and cis–trans isomerase with Gunter Fischer in Halle, Germany. After postdoctoral work with Professor Dario Neri in ETH Zürich, he joined B CUBE, Center for Molecular Bioengineering, in 2009 as group leader, and became a professor of biomolecular interaction at TU Dresden in 2016. His research interests include (i) self-assembled biomaterials, (ii) drug screening with DELs, (iii) drug design, and (iv) peptide bond cis–trans isomerase.

DNA-Encoded Library Chemistry: Amplification of Chemical Reaction Diversity for the Exploration of Chemical Space

Ms. Y. Huang

Ms. O. Savych

Dr. Y. Moroz

Prof. Y. Chen

Dr. R. A. Goodnow, Jr.

Ying Huang,[a] Olena Savych,[b] Yuii Moroz,[b,c] Yiyun Chen,*,[a] and Robert A. Goodnow, Jr.*,[d]

[a] Shanghai Institute of Organic Chemistry, University of Chinese Academy of Sciences, 200032 Shanghai, China,
   Email: yiyunchen@sioc.ac.cn
[b] Chemspace, ilukstes iela 38-5, Riga, LV-1082, Latvia
[c] National Taras Shevchenko University of Kyiv, 61 Volodymyrska Street, 01601 Kyiv, Ukraine
[d] Pharmaron USA Inc., Boston, MA 02451, USA, Email: robert.goodnow@pharmaron.com

Keywords. DNA-encoding; DNA-compatible chemistry; chemical diversity; hit generation; affinity selection; building blocks; commercial availability; drug-like properties; library design.

Abstract. The use of DNA for encoding and decoding in small-molecule synthesis for lead identification continues to gain widespread attention and application—more than a quarter century after its first disclosure. Successful execution of a diverse, drug-like library usually requires hundreds to thousands of commonly functionalized building blocks of relatively similar reactivity profiles. Aqueous and DNA-compatible organic reactions that utilize a large number of functionalized building blocks are perhaps among the most obvious and often discussed aspects of the successful application of this chemistry. This review highlights recent (since ~2015), relevant, new, and potentially highly useful such chemical transformations. Thereafter follows a discussion of the properties, requirements, costs, and diversity of building blocks that are currently available and may be useful for the construction of DNA-encoded libraries.

Outline
1. What Is DELT and Why Does It Continue in to Be Important?
2. General DEL Design Principles: Balancing Diversity, Size, and Drug-Like Properties against the Requirements of DNA-Compatible Reactions
3. Recently Developed Reactions That Have Been Applied on DNA or Have the Potential to Be Applied on DNA
    3.1. Photochemical and Radical Reactions
    3.2. Transition-Metal-Catalyzed Reactions
    3.3. Nucleophilic Reactions
    3.4. Enzymatic Reactions
4. Commonly Employed Reactions in DELT
5. Analysis of Commonly Used, Commercially Available Building Block Sets
6. Conclusion and Outlook
7. Acknowledgments
8. References and Notes

1. What Is DELT and Why Does It Continue to Be Important?

DNA-Encoded Library Technology (DELT) is a combination of several technologies for the creation of large mixtures of small molecules that are then used in affinity selection methods for the identification of small molecules having high binding affinity for a particular biological target, usually proteinaceous. In a DEL, each small molecule is covalently attached to a strand

76

DNA-Encoded Library Chemistry: Amplification of Chemical Reaction Diversity for the Exploration of Chemical Space
Ying Huang, Olena Savych, Yuii Moroz, Yiyun Chen,* and Robert A. Goodnow, Jr.*

of single- or double-strand DNA as a means of encoding its synthetic history. Population enrichment of high affinity binders, followed by amplification of the encoding DNA with a polymerase chain reaction (PCR) and subsequent sequencing of the resulting amplicons by next generation sequencing (NGS) make possible the identification of such higher affinity hit molecules. The chemistry for creating DELs has been useful for developing mixtures of hundreds of millions to billions of small molecules on a small scale. The concept for this innovation was first proposed in 1992 by Brenner and Lerner,[1] and was promptly followed by its exemplification.[2] During the intervening quarter century, DELT has grown from an interesting concept for the encoding of short peptides synthesized on a solid phase, to a well-established method for small-molecule hit identification, increasingly commonplace in pharma, biotech, and academic drug discovery research. This approach has been thoroughly reviewed.[3] It has also been described in multiple publications that are focused on synthetic chemistry as applied to this technology and hit identification[4] for biological targets of interest as well as clinical drug candidates.[5] Innovation in DELT is appealing to many scientists given its multidisciplinary nature that includes small-molecule organic chemistry, oligonucleotide chemistry, affinity chromatography, informatics, high–throughput (HT) DNA sequencing, and process design.

This review highlights newly disclosed DNA-compatible synthetic organic reactions that have been employed in the preparation of DELs or have the potential of being used for this purpose. Since a critical component for success in DELT is ready access to hundreds to thousands of building blocks (BBs), their availability, cost, chemical classification, and properties may seem obvious and simplistic. However, upon further examination, and given the number of BBs in question, it becomes clear that rigorous attention to these BB aspects is critical for implementing a successful DELT process. To this end, key BB aspects are explored in this review, thereby providing readers with, not only an update of the latest organic chemistry that might be applied in a DELT, but also an understanding of the general availability of BBs needed to demonstrate new molecular designs.

2. General DEL Design Principles: Balancing Diversity, Size, and Drug-Like Properties against the Requirements of DNA-Compatible Reactions

Whereas aspects of cost and feasibility of HTS decks limit screening campaigns to 10^6–10^7 compounds,[6,7] the advent of DELT now allows libraries to comprise typically 10^7–10^{10} compounds.[8,9] While this remains a tiny sampling of the theoretically accessible chemical space (10^{18}–10^{200} molecules, depending on the method of calculation),[10] DELT remains an attractive choice for hit generation. This is because DEL compound libraries result from split-and-pool synthesis, which has been known for many years,[11] even though the details of encoding and decoding millions of small molecules have come about only with the advent of NGS and the perfection of DELT.[1] The power of DELT's combinatorial synthesis to produce large numbers of compounds is strikingly demonstrated in **Table 1**.

However, well known to drug discovery scientists is Lipinski's "rule of 5"[12] that guides the design of small molecules with respect to oral bioavailability according to upper limits on molecular weight (MW), CLogP, and the count of hydrogen bond donors and acceptors (i.e., 500 Da, 5, 5, and 10, respectively). Without proper and rigorous attention to building block selection, multicycle combinatorial libraries often create molecules that surpass these limits. High affinity hits may be found, but their transformations to useful, orally bioavailable, drug-like substances may be challenging or impossible. Vividly illustrating this obvious point, is a seminal DELT paper by Clark et al. which describes the synthesis of 3- and 4-cycle DELs that are based on the triazine scaffold.[13] The DELs were then probed by affinity selection against p38α, resulting in the identification of high–affinity, off-DNA inhibitors of p38 MAP kinase (**Figure 1**).[13] In **Table 2**, entries 2 and 5 show the final molecular weights that would result from fairly low-molecular-weight building blocks, whereas entries 3 and 6 show the same from heavier building blocks. The highest molecular weight building blocks in these libraries are in excess of 275 Da for several of the inputs. It thus becomes clear that for 3-cycle DEL-A, the building block and core MWs cannot exceed 125 Da in order to produce molecules of less than 500 Da; for DEL-B the limit is 105 Da. Based on this simple analysis, it should be obvious that building block selection for DELs is quite constrained by molecular weight. As a rule of thumb, DEL building blocks and cores of molecular weights of 125 Da or less are generally the most useful if one expects all members of a DEL library to have less than 500 Da molecular weights. Often DELT chemists design libraries with building blocks that are available to them at the time and thus can end up with final compounds having MWs far in excess of 500 Da. The thinking then is that the DEL library will have some representation of molecular weights less than 500 Da, while larger MWs can be useful for structure–activity relationship (SAR) studies. Indeed, while diversity exploration of higher-molecular-weight products is useful for selection, their potential utility as starting points for small-molecule hit-to-lead drug discovery is invariably negatively assessed against these stringent, hard-wrought design criteria.

Before optimizing the drug-like properties of a DEL for use in hit generation, two key aspects need to be considered: (i) a good understanding of new and established chemical reactions that are compatible with a DNA encoding format, and (ii) the

Table 1. The Number of Compounds That May Result from 3- and 4-Cycle Combinatorial Libraries

No. of Cycles	Input Count for				No. of Library Members
	Cycle I	Cycle II	Cycle III	Cycle IV	
3	96	96	96	–	884,736
3	96	96	96	96	84,934,656
4	1,000	1,000	1,000	–	1×10^9
4	1,000	1,000	1,000	1,000	1×10^{12}

availability of suitable building blocks for the modeling of such reactions. Large pharmaceutical companies, having thousands of small-molecule building blocks in their compound inventories, have a significant advantage in terms of availability of small, diverse, and novel molecular inputs at already sunk costs. For DELT chemists who do not have access to such "hidden treasure", commercially available BBs are the next best option. To this end, the analysis in Section 5 of commercially available BB sets is presented as a means to understand the numbers of commonly used building blocks that are readily commercially available. Whereas an exhaustive analysis of relevant BBs may be more meaningful to a particular design and the preforming organization, it is still useful to the practicing DELT chemist to have an overview of what is available.

3. Recently Developed Reactions That Have Been Applied on DNA or Have the Potential to Be Applied on DNA
3.1. Photochemical and Radical Reactions
Liu and co-workers reported in 2011 the first visible-light-induced photoredox reaction performed on DNA, which was run in pH 7.4 aqueous buffer without precautions against air or oxygen (**eq 1**).[14] The substrate scope expansion and functional group compatibility of the azide reduction reaction were demonstrated with small molecules in organic solvents. A variety of protic functional groups; including free indoles, acids, and alcohols; were compatible with the reaction. Functional groups that are sensitive to hydrogenation; including alkenes, alkynes, and aryl halides; were not affected. In addition, functional groups that are sensitive to nucleophiles; including alkyl halides, alkyl mesylates, and aldehydes; were stable under the reaction conditions. An alkyl azide was also reduced in 24 hours under modified conditions, giving rise to the corresponding amine in 72% yield.

A "photoclick" cycloaddition between a diaryltetrazole tethered to a single-strand or a double-strand DNA, **3**, and the electron-deficient maleimide double bond in sulfo-Cy3 dye **4** was induced by UV irradiation at 365 nm in a temperature-controlled device (**eq 2**).[15] It was observed that the yield of the purified product DNA, **5**, was higher (34%) on the double-strand DNA relative to the single-strand one (8%). Remarkably, the measured rate constant was found to be high relative to those of other copper-free bioorthogonal transformations and was similar to those of the traditional copper-catalyzed "click" reaction. Two other alkene dipolarophiles, *N*-methylmaleimide and methylmethacrylate, were also reacted with **3** to evaluate the scope of the reaction and they led to 42% and 23% yield, respectively, of the corresponding product DNA.

The first visible-light-induced oxidative coupling reaction with the potential for DNA modification and DNA-encoded library synthesis was disclosed by Chen and co-workers (**eq 3**).[16] The reaction worked well with readily available primary, secondary, and tertiary alkylboronic acids or trifluoroborates to generate aryl-, silyl-, and alkyl-substituted alkynes. Various functional groups including alkenes, alkynes, aldehydes, ketones, esters, nitriles, azides, aryl halides, alkyl halides, alcohols, and

indoles were tolerated in this deboronative alkynylation in organic solvents. In addition, compatibility of the reaction with amino acids, oligosaccharides, proteins, and cell lysates was demonstrated. The reaction could be run in pH 7.4 aqueous buffer without excluding air or oxygen.

The same laboratory has also reported the first visible-light-induced reductive coupling reactions with DNA-compatibility, and the reaction conditions did not affect the enzyme activity of a protein enzyme (**eq 4**).[17] This C_{sp3}–C_{sp} bond coupling reaction worked with primary, secondary, tertiary, and α-heteroatom-substituted alkyl *N*-acyloxyphthalimides to construct aryl-, alkyl-,
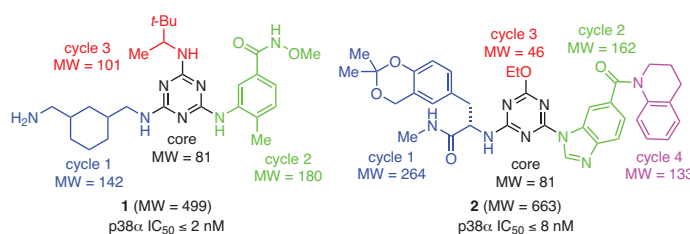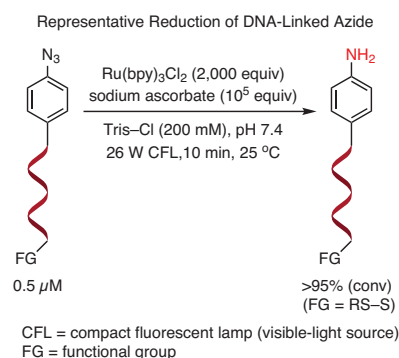

Figure 1. DEL-A and DEL-B, 3- and 4-Cycle DNA-Encoded Libraries Were Employed to Identify High-Affinity Hits against p38α. *(Ref. 13)*

Table 2. Final Product Molecular Weight (MW) That Results from Building Block Inputs for 3- and 4-Cycle DELs[a]

Entry	Building Block MW Input for				Final Product	MW of Final Product
	Cycle I	Cycle II	Cycle III	Cycle IV		
1	142	180	101	–	DEL-A hit **1**	499
2	75	75	75	–	DEL-A low	301
3	200	200	200	–	DEL-A high	676
4	264	162	46	133	DEL-B hit **2**	663
5	75	75	75	75	DEL-B low	357
6	200	200	200	200	DEL-B high	875

[a] Core MW = 81

Representative Reduction of DNA-Linked Azide

CFL = compact fluorescent lamp (visible-light source)
FG = functional group

eq 1 *(Ref. 14)*

78

DNA-Encoded Library Chemistry: Amplification of Chemical Reaction Diversity for the Exploration of Chemical Space
Ying Huang, Olena Savych, Yurii Moroz, Yiyun Chen,* and Robert A. Goodnow, Jr.*
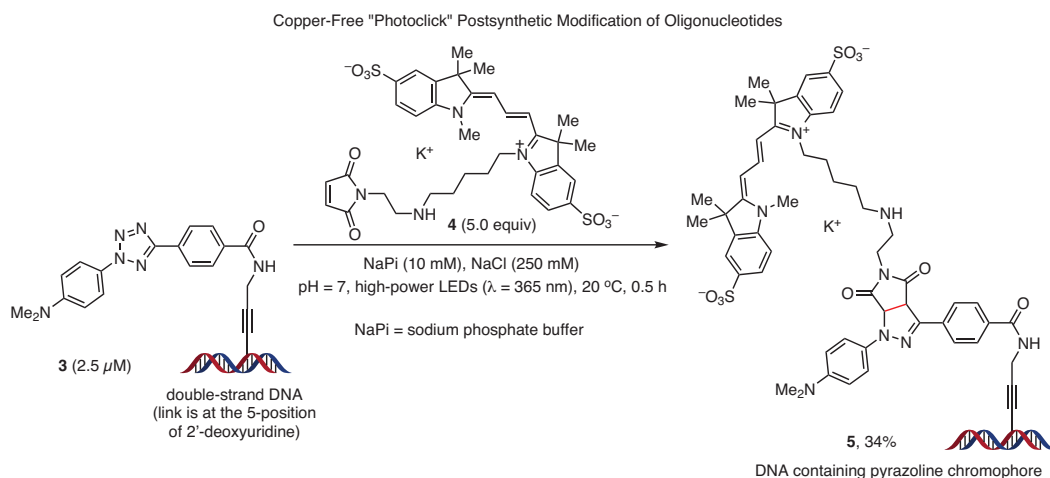
and silyl-substituted alkynes. The reaction was chemoselective when substrates were used that contained sensitive functional groups such as alkenes, alkynes, aldehydes, ketones, esters, amides, azides, aryl/alkyl halides, alcohols, phenols, carboxylic acids, and indoles. This rapid and mild reaction is compatible with biomolecules and thus has the potential to be applied to DNA modification and DNA-encoded library synthesis.

Flanagan and co-workers have reported a first proof of concept that a photoredox-mediated, 1,4-radical-addition reaction can take place under mild, aqueous conditions that are suitable for the preparation of DELs of around 75 million compounds (**eq 5**).[18] This C_{sp3}–C_{sp3} coupling reaction is compatible with a range of structurally diverse radical precursors a well as Michael acceptors and styrene derivatives. Moreover, to achieve high coupling yields, 1,000 equivalents of the protected α-amino acids need to be used. Alcohol, carboxylic acid, carboxamide,

and guanidine functional groups are tolerated; however, unprotected aliphatic α-amines, such as H-Pro-OH, afford the coupling product in significantly diminished yield. Unlike the other transformations presented thus far in this review, oxygen has a detrimental effect on the reaction necessitating that the glass reaction vial be degassed with nitrogen for 5 minutes before irradiation with blue LED light.

A fast, chemoselective, and general *anti*-Markovnikov hydrothiolation of alkenes and alkynes under mild conditions has been disclosed by Glorius, Guldi, and co-workers (**eq 6**).[19] The transformation relies on a biocompatible disulfide–ene reaction that proceeds by triplet–triplet energy transfer (TTEnT) sensitization of disulfides by the visible-light photocatalyst, and tolerates a wide range of functional groups including amides, sulfonamides, nitriles, alcohols, epoxides, aldehydes, ketones, esters, and heterocycles. The electron-rich, sterically

Copper-Free "Photoclick" Postsynthetic Modification of Oligonucleotides

eq 2 (Ref. 15)

Visible-Light-Induced Oxidative Coupling Reaction

6; >95% (conv), 76% (^{1}H NMR yield)
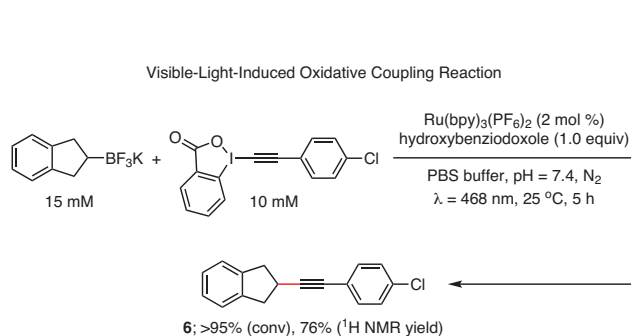
In the presence of L-tyrosine, L-cysteine, L-methionine, guanosine, naringin, ssDNA, bovine serum albumin, and bacterial cell lysates: >95% (conv) and 68–86% (^{1}H NMR) yield of **6** in 5–12 h

eq 3 (Ref. 16)

Visible-Light-Induced Reductive Coupling

naringin-*N*-acyloxyphthalimide (0.1 mM)
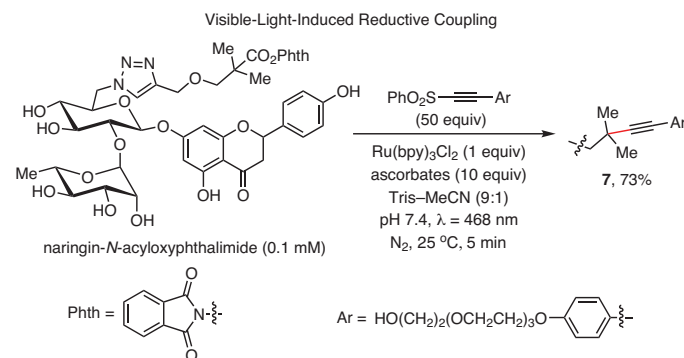
The reaction was compatible with stoichiometric amounts of amines, amino acids, proteins, oligosaccharides, nucleic acids, nucleosides, and cell lysates: Yield of **7**: 73–84%

The enzymatic activity of Human Carbonic Anhydrase II (HCA II) was unaffected by the reaction conditions
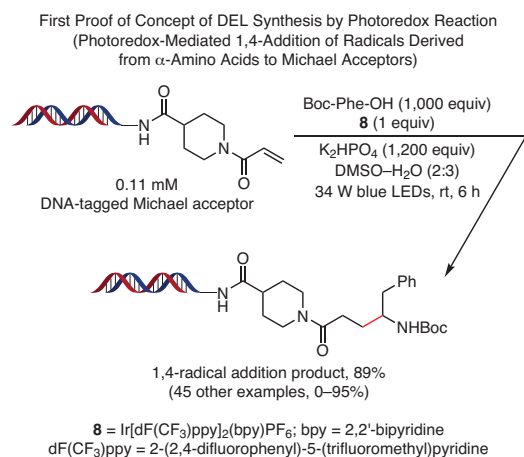
eq 4 (Ref. 17)

accessible alkene functionality reacts exclusively, while sterically demanding disulfides such as *tert*-butyl disulfide are unreactive. In aqueous buffer (pH 7.4), the yield of the reaction of carvone with dimethyl disulfide was not affected by the presence of 20 biomolecules such as saccharides, amino acids, nucleosides, nucleic acids, and human cell lysates. A minor modification of the reaction conditions permitted diaryl disulfides to be utilized for the synthesis of arylthioethers.

A team of researchers from Scripps, Pfizer, and Asymchem led by Baran and Blackmond have reported the first decarboxylative one-electron alkylation of a DNA-bound molecule using zinc nanopowder as the reductant (**eq 7**).[20] The reaction works with amino-containing alkyl carboxylic acids and even a dipeptide, and tolerates functional groups, such as thioethers and aryl iodides, that are potentially sensitive to reductive conditions or nonphysiological pH. Highly hindered C_{sp3}–C_{sp3} bond linkages, such as those forming quaternary carbons, could also be constructed on DNA through this cross-coupling reaction with electron-deficient alkenes.
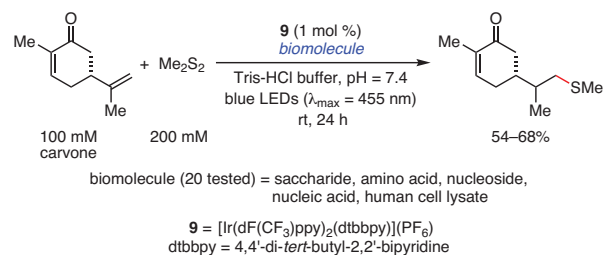
The direct, visible-light photoredox-catalyzed deoxygenative ketone synthesis has also been achieved under mild conditions in aqueous medium.[21] This formal hydroacylation of alkenes is compatible with a broad range of alkenes and para-substituted aromatic carboxylic acids and is achieved in moderate-to-good yields. It is proposed to occur via an acyl radical species [RC(O)•] that adds to the alkene with Ph_3P acting as an oxygen-transfer agent and water supplying a proton in the final step of the reaction. Both pyridyl-substituted alkenes and styrene derivatives were good alkene partners, and the reaction was not affected by the presence of biomolecules such as amino acids, oligosaccharides, nucleosides, nucleic acids, and proteins.

Very recently, two open-air, on-DNA synthetic protocols have been developed by Molander and co-workers (**Scheme 1**).[22] The first involves a Ni/photoredox dual catalytic C_{sp2}–C_{sp3} cross-coupling between a halogenated arene and an alkane. The
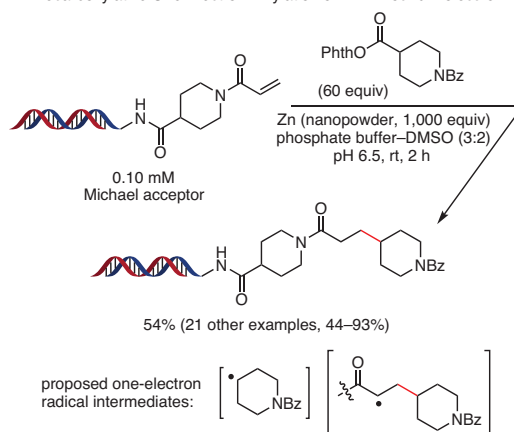

Biocompatibility of the Visible-Light Induced Disulfide–Ene Reaction

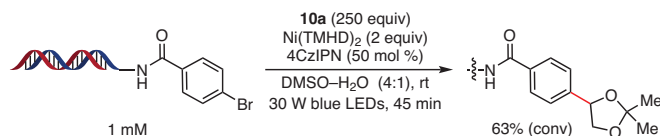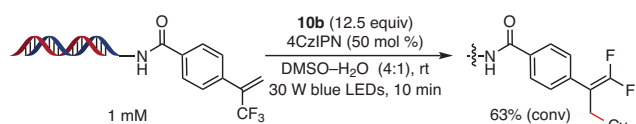biomolecule (20 tested) = saccharide, amino acid, nucleoside, nucleic acid, human cell lysate

9 = [Ir(dF(CF₃)ppy)₂(dtbbpy)](PF₆)
dtbbpy = 4,4'-di-*tert*-butyl-2,2'-bipyridine

eq 6 *(Ref. 19)*

Decarboxylative One-Electron Alkylation of DNA-Bound Molecule

54% (21 other examples, 44–93%)

proposed one-electron radical intermediates:

eq 7 *(Ref. 20)*

First Proof of Concept of DEL Synthesis by Photoredox Reaction
(Photoredox-Mediated 1,4-Addition of Radicals Derived from α-Amino Acids to Michael Acceptors)

1,4-radical addition product, 89%
(45 other examples, 0–95%)

8 = Ir[dF(CF₃)ppy]₂(bpy)PF₆; bpy = 2,2'-bipyridine
dF(CF₃)ppy = 2-(2,4-difluorophenyl)-5-(trifluoromethyl)pyridine

eq 5 *(Ref. 18)*

(a) Ni/Photoredox Dual Catalytic C_{sp2}–C_{sp3} Cross-Coupling

10a (250 equiv)
Ni(TMHD)₂ (2 equiv)
4CzIPN (50 mol %)
DMSO–H₂O (4:1), rt
30 W blue LEDs, 45 min

63% (conv)

(b) Photoredox-Mediated Crossover Defluorinative Alkylation

10b (12.5 equiv)
4CzIPN (50 mol %)
DMSO–H₂O (4:1), rt
30 W blue LEDs, 10 min

63% (conv)

The *gem*-difluoroalkenyl group is considered a metabolically stable isostere of the carbonyl group

4CzIPN = 1,2,3,5-tetrakis(carbazol-9-yl)-4,6-dicyanobenzene
TMHD = bis(2,2,6,6-tetramethyl-3,5-heptanedionate)

10a, R = (2,2-dimethyl-[1,3]-dioxolan-4-yl)
10b, R = cyclohexyl (Cy)
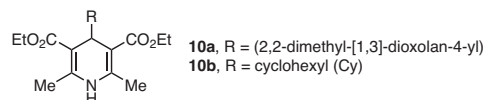
Scheme 1. Photoredox-Mediated, on-DNA Alkylations in Ambient Air and Aqueous Media under Mild Conditions. *(Ref. 22)*

80

DNA-Encoded Library Chemistry: Amplification of Chemical Reaction Diversity for the Exploration of Chemical Space
Ying Huang, Olena Savych, Yuii Moroz, Yiyun Chen,* and Robert A. Goodnow, Jr.*

second is the first example of a photoredox-mediated crossover defluorinative alkylation of a trifluoromethylalkene to generate a *gem*-difluoroalkene, which is proposed as a metabolically stable isostere of the carbonyl group. In the dual catalytic process, the water-compatible alkyl radical precursors included alkyl silicates, alkyl dihydropyridines, and amino acids which reacted well with variously substituted aryl iodides and with heteroaryl bromides bearing electron-deficient substituents. Functional groups that could be further elaborated in later cycles of DEL synthesis—e.g., epoxides, esters, alkenes, and *N*-Boc amines—were compatible with the reaction conditions. In the second protocol, the defluoroalkylation of trifluoromethyl alkenes, a DNA compatibility test was done by DNA misreads frequency sequencing, which was meaningful for the DEL synthesis, and such assessment of DNA integrity had often been overlooked before.

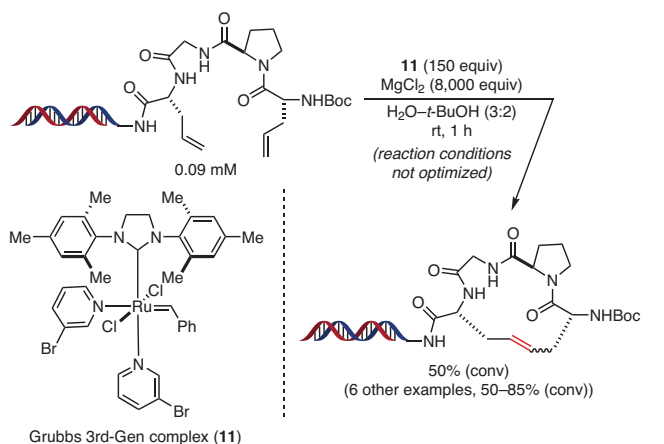### 3.2. Transition-Metal-Catalyzed Reactions

The first on-DNA ring-closing metathesis (RCM) and cross-metathesis (CM) were achieved on a double-strand DNA in water–*tert*-butyl alcohol solvent system by Lu et al. (**eq 8**).[23] The unoptimized RCM was promoted by Grubbs third-generation ruthenium complex **11** (150 equivalents), and required a large excess of $MgCl_2$ (8,000 equivalents) to be utilized to protect the DNA backbone from ruthenium-induced decomposition. The substrate scope for the on-DNA RCM was investigated with the

purpose of forming saturated small (5-, 6-, 7-, and 8-membered) and large (14- and 16-membered) rings in synthetically useful conversions. The on-DNA CM was tested with a representative on-DNA alkene and allyl alcohol as the other olefin under similar reaction conditions; here, too, $MgCl_2$ was required to achieve a 50% conversion.
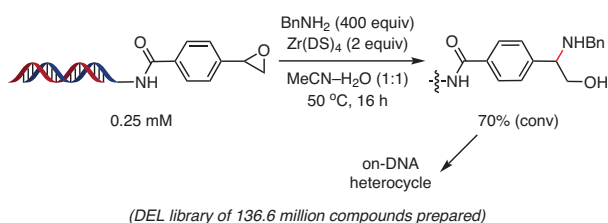
Fan and Davie reported in 2017 the first water-compatible, Zr(IV)-catalyzed aminolysis of epoxides on DNA to afford β-amino alcohols (**eq 9**).[24] Both aliphatic and aromatic primary amines gave moderate-to-excellent yields of the amino alcohols, while sterically hindered amines reacted more sluggishly to form mostly diols. The researchers took advantage of this protocol to build a library of 136.6 million on-DNA β-amino alcohols and their cyclization products, which led to the identification of multiple hits for a number of targets.

The acid- and gold-mediated on-DNA syntheses of hexathymidine-DNA-heterocycle chimeras have been disclosed by a team of researchers from Dortmund University (**Scheme 2**).[25] Diversely substituted β-carbolines—core scaffolds of drugs and pharmacologically active natural products—were accessed by a Brønsted acid catalyzed Pictet–Spengler reaction, while hexathymidine-DNA (hexT) conjugates of highly substituted pyrazolines and pyrazoles were prepared by a Au(I)-mediated cascade reaction. The hexT-heterocycle conjugates could then be ligated to the coding DNA sequences by T4 DNA ligation to produce encoded screening libraries that were informed by drug structures.

The same laboratory has also reported a similar Au(I)-mediated three-component reaction between alternately DNA-tethered aldehydes, hydrazides, and alkynols.[26] The reaction was compatible with thymine-, cytosine-, and adenine-
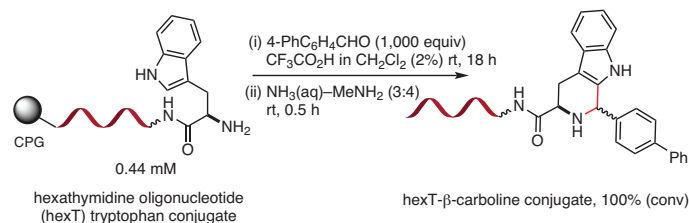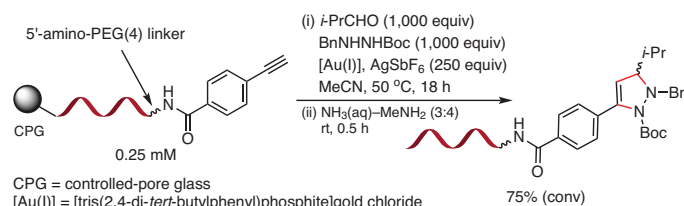

Grubbs 3rd-Gen complex (**11**)

11 (150 equiv)
$MgCl_2$ (8,000 equiv)
H_2O–*t*-BuOH (3:2)
rt, 1 h
(reaction conditions not optimized)

50% (conv)
(6 other examples, 50–85% (conv))

eq 8 *(Ref. 23)*

$BnNH_2$ (400 equiv)
Zr(DS)₄ (2 equiv)
MeCN–H_2O (1:1)
50 °C, 16 h

70% (conv)

on-DNA heterocycle

(DEL library of 136.6 million compounds prepared)

eq 9 *(Ref. 24)*

(a) Brønsted Acid Mediated Pictet–Spengler Reaction

(i) 4-PhC₆H₄CHO (1,000 equiv)
CF_3CO_2H in CH_2Cl_2 (2%) rt, 18 h
(ii) NH_3(aq)–$MeNH_2$ (3:4)
rt, 0.5 h

hexathymidine oligonucleotide (hexT) tryptophan conjugate

hexT-β-carboline conjugate, 100% (conv)

(b) Au(I)-Mediated Three-Component Reaction Forming DNA-Tagged Heterocycles

(i) *i*-PrCHO (1,000 equiv)
BnNHNHBoc (1,000 equiv)
[Au(I)], AgSbF₆ (250 equiv)
MeCN, 50 °C, 18 h
(ii) NH_3(aq)–$MeNH_2$ (3:4)
rt, 0.5 h

CPG = controlled-pore glass
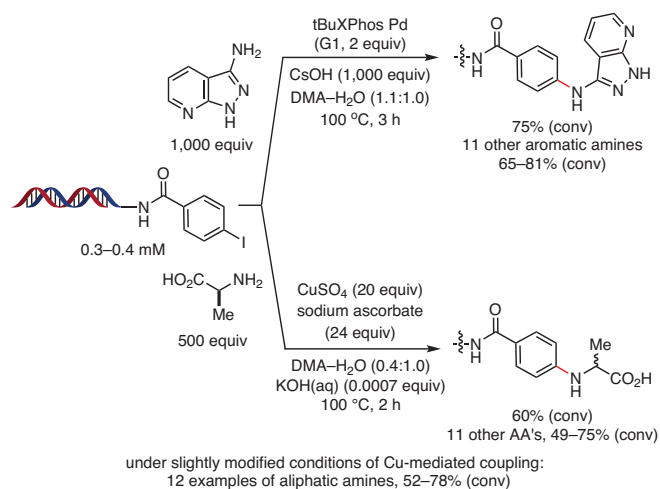[Au(I)] = [tris(2,4-di-*tert*-butylphenyl)phosphite]gold chloride

75% (conv)

Scheme 2. TFA- and Au(I)-Mediated on-DNA Syntheses of Substituted β-Carbolines, Pyrazolines, and Pyrazoles. *(Ref. 25)*

containing DNA, while guanine-containing DNA strands were degraded under the reaction conditions. The DNA-coupled aldehyde starting material resulted in the best yields of DNA-tagged substituted spiroheterocycles, whereas the alkynol- and hydrazide DNA-conjugates gave complex product mixtures, which precludes their use for parallel synthesis.

The first Pd- or Cu-promoted C–N cross-coupling between DNA-conjugated aryl iodides and primary amines has been achieved (**Scheme 3**).[27] The reaction works well for primary aromatic amines, aliphatic amines, and amino acids, which included over five hundred amino acids. The amino acids not only serve as Cu(I) ligands to promote the C–N coupling, but also prevent DNA decomposition as a result of DNA coordination to Cu(I). Using this technology, the authors succeeded in generating DNA-encoded libraries of 30.4 and 177 million compounds, of which a number of small-molecule hits were identified for biological targets of interest. An analogous copper-promoted amination of DNA-conjugated aryl iodides was reported by Ruff and Berst.[28] In this variant, the novel ligand 2-((2,6-dimethoxyphenyl)amino)-2-oxoacetic acid was found to improve conversions, and both primary and secondary amines—even some sterically congested secondary amines—were competent coupling partners. It is worth noting that this protocol could be carried out in air at low temperature, worked well for a large number of amine building blocks, and was compatible with DMSO as a water-miscible organic cosolvent.

More recently, a palladium-mediated C–N cross-coupling between DNA-conjugated aryl bromides and aromatic amines was disclosed by Torrado and collaborators.[29] The reaction worked for diverse heteroaromatic amines such as thiazoles, pyrazoles, pyrazines, pyridines, benzoxazoles, benzimidazoles, and oxadiazoles as well as anilines containing aliphatic alcohols, sulfonami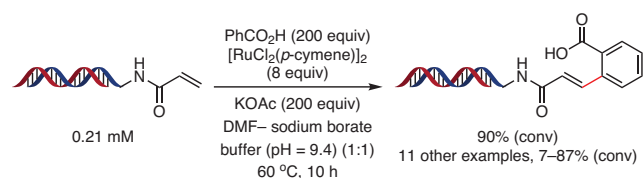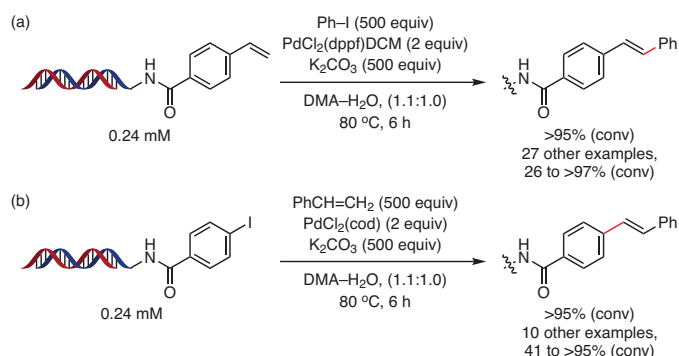des, and ester functional groups. Not surprisingly, the coupling of the aryl bromides required a higher temperature (60 °C) to achieve useful conversions than the coupling of the aryl iodides (30 °C).

Lu, Zhong, and co-workers have disclosed the first ruthenium-promoted, on-DNA C_{sp2}–C_{sp2} coupling between acrylamides and aromatic carboxylic acids via a C−H activation reaction (**eq 10**).[30] Under the optimized conditions, good-to-excellent conversions were obtained regardless of whether the acrylamide or the carboxylic acid was conjugated to the DNA. Because aromatic carboxylic acids are more readily commercially available, the reaction scope was investigated with respect to the acid partner. The reaction tolerated bromo, iodo, chloro, amino, hydroxyl, and carboxyl substituents in the carboxylic acid partner, whereas aldehydo, vinyl, and alkoxycarbonyl substituents resulted in low conversions. The value of this approach is that the aromatic carboxylic acid partner serves as a bifunctional building block, whereby the carboxylic acid group can, after the coupling step, be further elaborated to introduce more diversity in a DEL that could range in size from millions to billions after three cycles.

The palladium-promoted Heck coupling has been successfully extended to the reaction of double-strand and single-strand DNA-conjugated acrylamides, styrenes, and aryl iodides with aryl iodides and bromides, aromatic borates, and styrenes (**Scheme 4**).[31] Moderate-to-excellent conversions were observed, and the reaction was compatible with cyano, hydroxyl, amino, and halide functional groups, which could conceivably be elaborated further. The applicability of this method to library diversification was demonstrated with the on-DNA Heck reaction as the cycle 2 step in a three-cycle DEL synthesis.

eq 10 (Ref. 30)

Scheme 3. Pd- or Cu-Promoted C–N Cross-Coupling between DNA-Conjugated Aryl Iodides and Primary Amines or Amino Acids. (Ref. 27)

Scheme 4. Palladium-Promoted Heck Reaction on the DNA. (Ref. 31)

82

DNA-Encoded Library Chemistry: Amplification of Chemical Reaction Diversity for the Exploration of Chemical Space
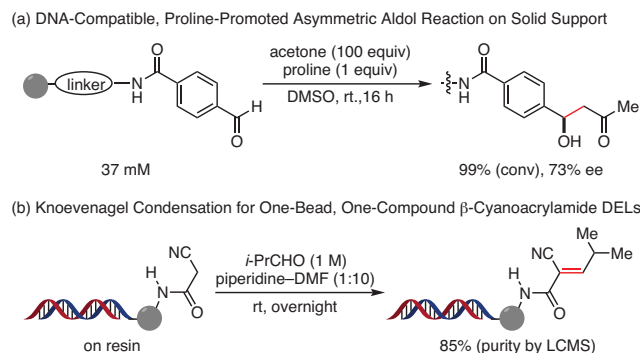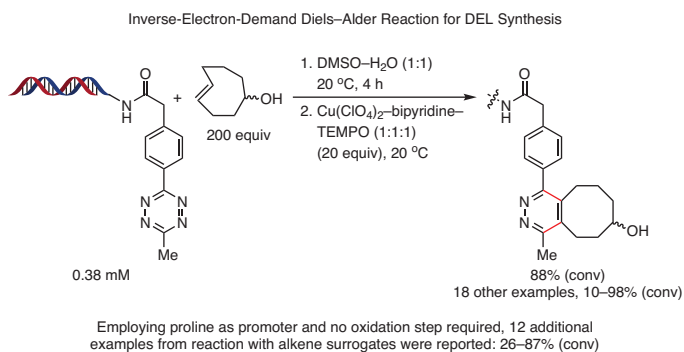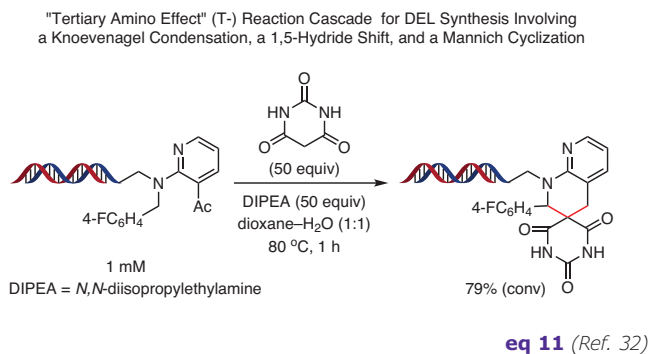Ying Huang, Olena Savych, Yuii Moroz, Yiyun Chen,* and Robert A. Goodnow, Jr.*

3.3. Nucleophilic Reactions

One of very few examples of an on-DNA reaction sequence that forms two carbon–carbon bonds was reported by Tian et al. (**eq 11**).[32] In this process, termed "T-reaction" for "tertiary amino effect reaction", *ortho*-dialkylaminoaryl aldehydes undergo a reaction cascade consisting of a Knoevenagel condensation, a 1,5-hydride shift, and a Mannich cyclization to form diversely substituted spirocycles. Various benzylic amines were applicable, less activated alkyl amines were tolerated, and differentially substituted secondary amines were also tolerated. Heterocyclic and structurally complex substrates performed equally well as simple piperidines and pyrrolidines. The reaction worked for activated cyclic ketones, amides and esters, which gave higher yields than their acyclic counterparts. Various remote functional groups such as thiocarbonyls and esters did not hinder the reaction.

Neri and co-workers employed an on-DNA Diels–Alder cycloaddition protocol to generate a 4,000-compound DEL in good purities and yields. The DNA fragments functioned as amplifiable bar codes for identification purposes, and the DEL was compatible with decoding strategies that are based on ultra-HTS techniques,[33] In a similar vein, Dai and co-workers demonstrated the utility and versatility of the inverse-electron-demand Diels–Alder reaction for the synthesis of DNA-tagged

pyridazines (**eq 12**).[34] The authors reported two complementary protocols for reacting DNA-conjugated 1,2,4,5-tetrazines with olefins and alkene surrogates such as cyclic ketones to yield the desired pyridazines with moderate-to-excellent conversions. Select functionalized pyridazines were shown to undergo Suzuki–Miyaura coupling, acylation, and S_NAr substitution, respectively, underscoring the potential utility of the pyridine scaffold for future DEL syntheses.

In a series of papers,[35–37] Kodadek and his research teams have reported the development of DNA-compatible variants of the asymmetric aldol reaction (**Scheme 5**, Part (a)), the asymmetric Mannich reaction, and the Knoevenagel condensation (Scheme 5, Part (b)) all on solid support. In the aldol variant, proline promoted the asymmetric synthesis of β-hydroxy ketones from immobilized aldehydes and soluble ketones. Heteroaromatic aldehydes were compatible with the reaction and the reaction conditions did not adversely affect the polymerase chain amplification of DNA; however, the enantioselectivities observed were generally modest (ee = 54%−79%). Similarly, the same laboratory disclosed a DNA-compatible, solid-phase synthesis of β-amino ketones through a proline-mediated asymmetric Mannich reaction, again between immobilized aldehydes and soluble ketones.[36] Anilines substituted at the meta or para position worked well, in contrast to ortho-substituted anilines which did not form the desired products. Moreover, the syn product was favored, and the ee of the syn product varied from 54% to 96%. Extending the applicability of the DNA-compatible, combinatorial synthesis on solid support approach, another of Kodadek's teams reported the synthesis of β-cyanoacrylamides by a Knoevenagel condensation between immobilized α-cyano amides and soluble aldehydes.[37] The cyanoacrylamides underwent Michael additions with thiol and phosphine nucleophiles, and thus could serve as a source of reversibly covalent protein ligands that are cysteine-reactive. The Knoevenagel reaction worked well for aliphatic and aromatic aldehydes including those with branched alkyl groups; however ketones were unreactive.

"Tertiary Amino Effect" (T-) Reaction Cascade for DEL Synthesis Involving a Knoevenagel Condensation, a 1,5-Hydride Shift, and a Mannich Cyclization

eq 11 *(Ref. 32)*

Inverse-Electron-Demand Diels–Alder Reaction for DEL Synthesis

Employing proline as promoter and no oxidation step required, 12 additional examples from reaction with alkene surrogates were reported: 26–87% (conv)

eq 12 *(Ref. 34)*

(a) DNA-Compatible, Proline-Promoted Asymmetric Aldol Reaction on Solid Support

(b) Knoevenagel Condensation for One-Bead, One-Compound β-Cyanoacrylamide DELs

Scheme 5. DNA-Compatible Aldol Reaction and Knoevenagel Condensation on Solid Support. *(Ref. 35,37)*

A method for synthesizing DELs that contain members possessing the less well studied polycyclic isoxazolidine moiety was developed by a team led by Schreiber.[38] The method relies on an intramolecular and DNA-compatible nitrone–alkene [3 + 2] cycloaddition that forms the DNA-linked isoxazolidine, a structural feature found in a number of alkaloids and bioactive compounds (**eq 13**). The nitrone is formed in situ from reaction of the corresponding *N*-alkylhydroxylamine with the carbonyl group of the substrate, and the subsequent intramolecular cycloaddition generates a minimum of two stereogenic centers. The method allows for appendage diversification and does not cause DNA damage or preclude subsequent amplification and ligation steps.

Very recently, Sharpless and co-workers reported the first iminosulfur oxydifluoride (R–N=SOF$_2$) reaction with primary amines and phenols on single-strand DNA with the aim of demonstrating the usefulness of the SuFEx click chemistry for bioconjugation applications (**Scheme 6**).[39] The off-DNA model reactions took place under mild conditions in aqueous buffer to give sulfamides (from primary amines), sulfuramidimidoyl fluorides (from secondary amines), and sulfurofluoridoimidates (from phenols) in up to 99% isolated yields (29 examples).

3.4. Enzymatic Reactions
The first on-DNA chemo-enzymatic synthesis of a small, carbohydrate-based library was reported by Thomas et al. (**Scheme 7**).[40] This approach relied on either an enzymatic galactosylation or an enzymatic and site-specific oxidation to generate the DNA-glycoconjugate library members. In the galactosylation protocol, bovine β1,4-galactosyltransferase (β1,4-GalT) catalyzed the selective transfer of Gal-β1,4 from uridine diphosphate galactose (UDP-Gal) to an acceptor GlcNAc in the presence of MnCl$_2$ under mild conditions and leading to excellent conversions. In the site-specific oxidation protocol, galactose oxidase (GOase) variants M$_1$ and F$_2$ generated aldehydes from hexoses by a highly selective oxidation of the C6 hydroxyl group in the galactose moiety. The aldehydes were then elaborated further by hydrazone/oxime ligation or reductive amination.
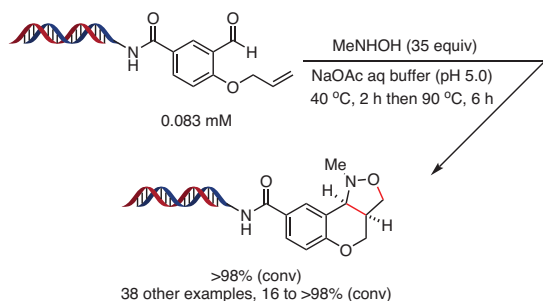

eq 13 *(Ref. 38)*

4. Commonly Employed Reactions in DELT
DEL chemistry is at its core combinatorial chemistry. Accordingly, it should not be a surprise that its practice has evolved to date using chemical transformations that are robust, high yielding, and for which the necessary building blocks are readily available. DELT-compatible chemistries must not only satisfy the three criteria just listed, but must also operate under DNA-compatible conditions: dilute (0.1−1 mM) aqueous media at pH between 4 and 14 and at temperatures between 25 and 90 °C, and the reactions must not alter or degrade the DNA encoding tags. In Section 3, we have presented "recently developed reactions", which we arbitrarily define as those
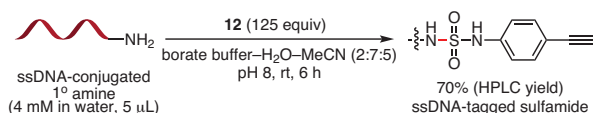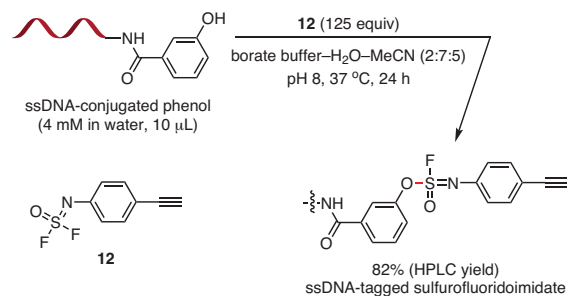
Scheme 6. First Iminosulfur Oxydifluoride on-DNA Reaction with Primary Amines and Phenols. *(Ref. 39)*

β1,4-GalT = bovine β1,4-galactosyltransferase; GOase = galactose oxidase
UDP-Gal = uridine diphosphate galactose; HRP = horseradish peroxidase
MES = 2-(*N*-morpholino)ethanesulfonic acid; NaPi = sodium phosphate

Scheme 7. Carbohydrate DEL Synthesis via Enzymatic Catalysis. *(Ref. 40)*

DNA-Encoded Library Chemistry: Amplification of Chemical Reaction Diversity for the Exploration of Chemical Space
Ying Huang, Olena Savych, Yuii Moroz, Yiyun Chen,* and Robert A. Goodnow, Jr.*

having been reported since 2015, and which have included many interesting and potentially very useful transformations. For the purposes of this review, we term "commonly employed reactions" in DELT as those already reported and tabulated in various reviews and communications.[41,42] Such reactions include amide formation, reductive aminations, amine capping reactions, Suzuki and Sonogashira couplings, some condensations and some heterocycle-forming cyclizations.

5. Analysis of Commonly Used, Commercially Available Building Block Sets

Presently, published reports of methods developed specifically for assembling DEL libraries are growing rapidly. However, many newly described DELT-applied reactions are often exotic and lacking sufficient validation with a wide variety of reagents. For some of these reactions, reagents are simply not available in any practical numbers (hundreds) to create a DEL. At the same time, large and diverse DELs reported to date have been synthesized using what one would mostly call classical combinatorial chemistry transformations and producing libraries that have been used in affinity selection campaigns to identify useful hits for drug discovery. While such classical and well-known reactions are modest in number compared to standard organic chemistry in solution, an abundance of building blocks for these reactions (e.g., amines, carboxylic acids, and aryl halides) has enabled the creation

of DELs of sufficient size, quantity, and diversity. Therefore, while more advanced approaches will continue to be reported and building blocks for them will become available, the use of commonly applied reactions will likely remain the mainstay of DEL production for the time being. Thus, an understanding of the availability and diversity of building blocks commercially available and applicable to DELT becomes critical to creating the most economical, diverse, drug-like, and ultimately the most useful DELs for drug discovery.

The number of modifiable functional groups in the DELT cores is also limited because of the necessity of orthogonal reactivity. For this analysis, we selected several functional groups that have many readily available reaction partners in well-established and DELT-compatible protocols. The list includes classical functional groups such as amines, aldehydes, and carboxylic acids, as well as more recently introduced ones such as terminal alkynes, azides, aryl halides, and boronic acids/esters.

We have grouped the modifiers in three categories: (1) *Amine Modifiers* (aldehydes, aryl halides, carboxylic acids, and sulfonyl halides but not alkyl halides due to concerns about reaction selectivity); (2) *Carbo Modifiers (*primary and secondary amines)*,* and (3) *Couplers* (alkynes, aryl halides, azides, and boronic acids or esters). Indeed, depending on the library chemistry in question, one might require a different classification of functional groups. Furthermore, the exclusion of various chemical substituents will likely reduce these numbers. Nevertheless, the intent of this analysis is to give an understanding of building blocks that exist physically and fall into acceptable categories of cost and molecular weight.

To analyze the availability of the modifiers, we selected 12 vendors with more than 8,000 "off-the-shelf" compounds.[43] This cut-off identifies vendors that could immediately ship large numbers of building blocks as a single source. As such, this helps to reduce costs, logistical complexity, and lead time, as special formatting is typically required. The distribution of the modifiers from each supplier is shown in **Figure 2**, and the percentages among them of so-called *small* compounds and *inexpensive* compounds are shown in **Table 3**. The criteria used for selecting *small* and *inexpensive* are: (i) A *small* modifier for two-cycle libraries should have a molecular weight (MW) ≤150 Da, while a *small* core building block's MW should not exceed 200 Da. (ii) An *inexpensive* compound is one costing not more
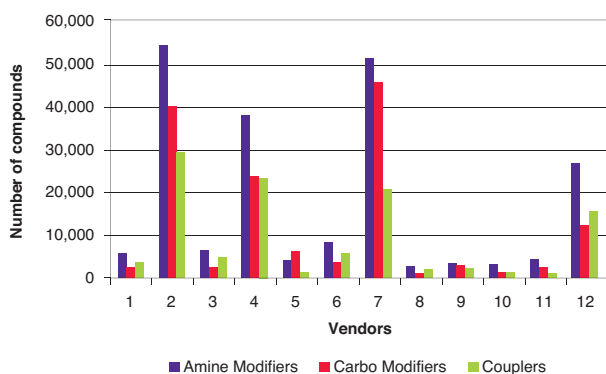

Figure 2. Distribution of Modifiers from the Ten Vendors Selected. *(Ref. 43)*

Table 3. Commercially Available Modifier Compounds: Count and % Small / % Inexpensive. *(Ref. 43)*

Vendor No	1	2	3	4	5	6	7	8	9	10	11	12
Amine Modifiers[a]	**5,665**	**54,083**	**6,328**	**37,892**	**4,169**	**8,200**	**51,163**	**2,476**	**3,409**	**2,897**	**4,300**	**26,571**
	65/42	24/24	60/40	22/29	46/25	60/42	15/22	64/53	32/42	0/22	21/14	34/30
Carbo Modifiers[b]	**2,400**	**40,342**	**2,338**	**23,575**	**5,891**	**3,528**	**45,602**	**1,047**	**2,856**	**1,241**	**2,202**	**12,124**
	55/26	15/14	59/25	14/16	48/14	51/22	10/13	57/34	20/25	0/14	20/11	27/19
Couplers[c]	**3,608**	**29,323**	**4,590**	**23,497**	**1,396**	**5,757**	**20,516**	**1,752**	**2,354**	**1,161**	**1,282**	**15,408**
	67/57	25/37	58/52	22/42	50/46	60/55	18/37	66/67	32/51	0/37	25/25	33/46

[a] Aldehydes, aryl halides, carboxylic acids, and sulfonyl halides. [b] Primary and secondary amines. [c] Alkynes, aryl halides, azides, and boronics.

than US $100 per gram. To add flavor to these numbers, we estimate the size of a possible 2-cycle DEL assembled from each vendor's collection. For this simulation, we selected an alkyne-containing N-protected amino acid as the core molecule (**Scheme 8**). The chemistry for decorating these cores appears classical from one side while challenging from the other: *amine modifiers* being the largest group of decorators, while *azides* being the smallest.

The sizes of the resultant libraries are obtained by multiplying the number of decorators by the number of cores (**Table 4**). As the number of cores, we took all available from selected 12 vendors core compounds. Column "All" shows the libraries assembled with all decorators and cores; columns "*Small*" and "*Inexpensive*" show the libraries assembled with small or inexpensive modifiers, respectively. For the "*Inexpensive*" libraries, we took all available numbers of cores, while for "*Small*" libraries, the cores had MW ≤ 200 Da. The number of cores is indicated in the header.

In some cases, modifiers that meet the small or inexpensive criteria do not exist and, thus, the library size is zero. Clearly, it is generally possible to enumerate many compounds. This is a good start from which to exclude compounds according to various design criteria (e.g., size, molecular diversity, undesired functional groups, predicted target affinity, etc.). This simulation also illustrates the numbers of building blocks that will have to be available in order for recently reported DNA-compatible reactions to have the impact for hit identification as the more classical reactions have already had.

6. Conclusion and Outlook

The past decades have witnessed excellent and widely used applications of transition-metal catalysis and nucleophilic addition reactions in the synthesis of DNA-encoded libraries. In the years to come, it is expected that the recently disclosed photochemical, radical, and enzymatic reactions—with their excellent chemoselectivity, functional group compatibility, and mild reaction conditions—will have a profound impact on DNA-encoded library chemistry by enabling greater chemical reaction diversity amplification and more extensive chemical space exploration. At the same time, business drivers will make available ever larger numbers of building blocks to create large DEL sets based on these reactions.

7. Acknowledgments

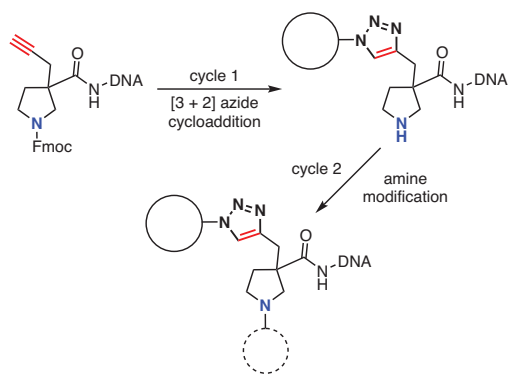## 8. References and Notes

(1) Brenner, S.; Lerner, R. A. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 5381.

(2) Nielsen, J.; Brenner, S.; Janda, K. D. *J. Am. Chem. Soc.* **1993**, *115*, 9812.

(3) Goodnow, R. A., Jr.; Dumelin, C. E.; Keefe, A. D. *Nat. Rev. Drug Discovery* **2017**, *16*, 131.

(4) Franzini, R. M.; Randolph, C. *J. Med. Chem.* **2016**, *59*, 6629.

(5) 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,

Scheme 8. Decoration of Alkyne-Containing N-Protected Amino Acid for a Simulated DEL.

Table 4. Estimated Size of a Possible 2-Cycle DEL Assembled from each Vendor's Collection.[a]

Vendor No	All Cores & Modifiers	"Small" Cores & Modifiers	All Cores & "Inexpensive" Modifiers
1	291,195	0	0
2	169,523,700	1,789,452	5,453,700
3	1,635,300	732,144	0
4	34,957,725	917,084	836,925
5	908,460	141,470	95,130
6	1,688,400	394,730	52,575
7	715,456,170	2,085,552	35,033,625
8	42,810	0	0
9	17,902,500	1,382,346	4,120,830
10	524,205	0	19,650
11	7,628,310	93,296	54,540
12	31,920,480	2,971,458	854,595

[a] Cores: 15 total, 14 "small".

86

DNA-Encoded Library Chemistry: Amplification of Chemical Reaction Diversity for the Exploration of Chemical Space
Ying Huang, Olena Savych, Yuii Moroz, Yiyun Chen,* and Robert A. Goodnow, Jr.*

J. *J. Med. Chem.* **2017**, *60*, 1247.

(6) Mayr, L. M.; Bojanic, D. *Curr. Opin. Pharmacol.* **2009**, *9*, 580.

(7) Gong, Z.; Hu, G.; Liu, Z.; Wang, F.; Zhang, X.; Xiong, J.; Li, P.; Xu, Y.; Ma, R.; Chen, S.; Li, J. *Curr. Drug Discovery Technol.* **2017**, *14*, 216.

(8) Dickson, P.; Kodadek, T. *Org. Biomol. Chem.* **2019**, *17*, 4676.

(9) Favalli, N.; Bassi, G.; Scheuermann, J.; Neri, D. *FEBS Lett.* **2018**, *592*, 2168.

(10) Follmann, M.; Briem, H.; Steinmeyer, A.; Hillisch, A.; Schmitt, M. H.; Haning, H.; Meier, H. *Drug Discovery Today* **2019**, *24*, 668.

(11) Furka, Á.; Sebestyén, F.; Asgedom, M.; Dibó, G. *Abstracts of Papers, 10th International Symposium on Medicinal Chemistry*, Budapest, Hungary, 1988; p 288.

(12) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Delivery Rev.* **1997**, *23*, 3.

(13) 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. *Nat. Chem. Biol.* **2009**, *5*, 647.

(14) Chen, Y.; Kamlet, A. S.; Steinman, J. B.; Liu, D. R. *Nat. Chem.* **2011**, *3*, 146.

(15) Arndt, S.; Wagenknecht, H. A. *Angew. Chem., Int. Ed.* **2014**, *53*, 14580.

(16) Huang, H.; Zhang, G.; Gong, L.; Zhang, S.; Chen, Y. *J. Am. Chem. Soc.* **2014**, *136*, 2280.

(17) Yang, J.; Zhang, J.; Qi, L.; Hu, C.; Chen, Y. *Chem. Commun.* **2015**, *51*, 5275.

(18) Kölmel, D. K.; Loach, R. P.; Knauber, T.; Flanagan, M. E. *ChemMedChem* **2018**, *13*, 2159.

(19) Teders, M.; Henkel, C.; Anhäuser, L.; Strieth-Kalthoff, F.; Gomez-Suarez, A.; Kleinmans, R.; Kahnt, A.; Rentmeister, A.; Guldi, D.; Glorius, F. *Nat. Chem.* **2018**, *10*, 981.

(20) Wang, J.; Lundberg, H.; Asai, S.; Martin-Acosta, P.; Chen, J. S.; Brown, S.; Farrell, W.; Dushin, R. G.; O'Donnell, C. J.; Ratnayake, A. S.; Richardson, P.; Liu, Z.; Qin, T.; Blackmond, D. G.; Baran, P. S. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E6404.

(21) Zhang, M.; Xie, J.; Zhu, C. *Nat. Commun.* **2018**, *9*, 3517.

(22) Phelan, J. P.; Lang, S. B.; Sim, J.; Berritt, S.; Peat, A. J.; Billings, K.; Fan, L.; Molander, G. A. *J. Am. Chem. Soc.* **2019**, *141*, 3723.

(23) Lu, X.; Fan, L.; Phelps, C. B.; Davie, C. P.; Donahue, C. P. *Bioconjugate Chem.* **2017**, *28*, 1625.

(24) Fan, L.; Davie, C. P. *ChemBioChem* **2017**, *18*, 843.

(25) Škopić, M. K.; Salamon, H.; Bugain, O.; Jung, K.; Gohla, A.; Doetsch, L. J.; dos Santos, D.; Bhat, A.; Wagner, B.; Brunschweiger, A. *Chem. Sci.* **2017**, *8*, 3356.

(26) Škopić, M. K.; Willems, S.; Wagner, B.; Schieven, J.; Krause, N.; Brunschweiger, A. *Org. Biomol. Chem.* **2017**, *15*, 8648.

(27) Lu, X.; Roberts, S. E.; Franklin, G. J.; Davie, C. P. *MedChemComm* **2017**, *8*, 1614.

(28) Ruff, Y.; Berst, F. *MedChemComm* **2018**, *9*, 1188.

(29) De Pedro Beato, E.; Priego, J.; Gironda-Martinez, A.; González, F.; Benavides, J.; Blas, J.; Martin-Ortega, M. D.; Toledo, M. Á.; Ezquerra, J.; Torrado, A. *ACS Comb. Sci.* **2019**, *21*, 69.

(30) Wang, X.; Sun, H.; Liu, J.; Dai, D.; Zhang, M.; Zhou, H.; Zhong, W.; Lu, X. *Org. Lett.* **2018**, *20*, 4764.

(31) Wang, X.; Sun, H.; Liu, J.; Zhong, W.; Zhang, M.; Zhou, H.; Dai, D.; Lu, X. *Org. Lett.* **2019**, *21*, 719.

(32) Tian, X.; Basarab, G. S.; Selmi, N.; Kogej, T.; Zhang, Y.; Clark, M.; Goodnow, R. A., Jr. *Med. Chem. Commun.* **2016**, *7*, 1316.

(33) Buller, F.; Mannocci, L.; Zhang, Y.; Dumelin, C. E.; Scheuermann, J.; Neri, D. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5926.

(34) Li, H.; Sun, Z.; Wu, W.; Wang, X.; Zhang, M.; Lu, X.; Zhong, W.; Dai, D. *Org. Lett.* **2018**, *20*, 7186.

(35) Shu, K.; Kodadek, T. *ACS Comb. Sci.* **2018**, *20*, 277.

(36) Tran-Hoang, N.; Kodadek, T. *ACS Comb. Sci.* **2018**, *20*, 55.

(37) Pels, K.; Dickson, P.; An, H.; Kodadek, T. *ACS Comb. Sci.* **2018**, *20*, 61.

(38) Gerry, C. J.; Yang, Z.; Stasi, M.; Schreiber, S. L. *Org. Lett.* **2019**, *21*, 1325.

(39) Liu, F.; Wang, H.; Li, S.; Bare, G. A. L.; Chen, X.; Wang, C.; Moses, J. E.; Wu, P.; Sharpless, K. B. *Angew. Chem., Int. Ed.* **2019**, *58*, 8029.

(40) Thomas, B.; Lu, X.; Birmingham, W. R.; Huang, K.; Both, P.; Reyes Martinez, J. E.; Young, R. J.; Davie, C. P.; Flitsch, S. L. *ChemBioChem* **2017**, *18*, 858.

(41) Satz, A. L.; Cai, J.; Chen, Y.; Goodnow, R.; Gruber, F.; Kowalczyk, A.; Petersen, A.; Naderi-Oboodi, G.; Orzechowski, L.; Strebel, Q. *Bioconjugate Chem.* **2015**, *26*, 1623.

(42) Luk, K.-C.; Satz, A. L. DNA-Compatible Chemistry. In *A Handbook for DNA-Encoded Chemistry: Theory and Applications for Exploring Chemical Space and Drug Discovery;* Goodnow, R. A., Jr., Ed.; Wiley: Hoboken, NJ, 2014; pp 67–98.

(43) (a) Vendor: Accela ChemBio (1), Angene Chemical (2), Apollo Scientific (3), AstaTech (4), ChemBridge (5), ChemScene (6), Enamine (7), Manchester Organics (8), Pharmablock (9), Specs (10), Toronto Research Chemicals (11), and WuXi LabNetwork (12). (b) Analysis of the vendors' catalogs and pricing as well as the classification of compounds and calculation of molecular weights (using RDKit software tools) were performed by Chemspace (chem-space.com).

About the Authors

Ying Huang received her B.S. degree in chemistry in 2015 from Central China Normal University. The same year, she moved to the Shanghai Institute of Organic Chemistry, University of Chinese Academy of Sciences, to pursue a master's degree under the direction of Professor Yiyun Chen. She is currently a doctorate student in Professor Yiyun Chen's group, and is focusing on developing novel and biocompatible light-induced chemical reactions for the post-synthetic modification of nucleic acids.

Olena Savych received her M.Sci. degree in organic chemistry in 2017 from the National Taras Shevchenko University of Kyiv. Her thesis research project focused on developing new

technologies for parallel chemistry applications. She is currently a data curation specialist at Chemspace.

Yurii Moroz received his Ph.D. degree in inorganic chemistry in 2010 from the National Taras Shevchenko University of Kyiv. His doctoral dissertation was on the topic of polymetal coordination compounds. Yurii then did postdoctoral research at Tufts University, working with organometallic compounds, and at Syracuse University, working with peptides and proteins. After returning to the Ukraine, Yurii has focused on approaches to generate synthetically feasible small molecules. He joined Chemspace as CEO in 2017.

Yiyun Chen is a professor at the Shanghai Institute of Organic Chemistry and a joint professor at ShanghaiTech University. Yiyun received his B.S. degree in chemistry with honors in 2002 from Peking University, and earned his Ph.D. degree in organic chemistry in 2007 with Professor Chulbom Lee at Princeton University. During his postdoctoral studies with Professor David R. Liu at Harvard University and Howard Hughes Medical Institute, Yiyun used the DNA-encoded reaction-discovery system to discover a new biocompatible visible-light-induced azide reduction reaction. In 2011, Yiyun was appointed Principal Investigator at the Shanghai Institute of

Organic Chemistry, where he is developing novel biocompatible light-induced chemical reactions and new optochemical tools for studying biological systems.

Robert A. Goodnow, Jr. obtained his B.S. degree in chemistry from Georgetown University and his Ph.D. degree (in natural products organic and bioorganic chemistry) from Columbia University. He then carried out research on oligosaccharide synthesis at Princeton University as an NIH Postdoctoral Fellow. Rob began his industrial career at Hoffmann-La Roche, and his research interest has since focused on medicinal chemistry and the implementation of several chemistry platform technologies such as small-molecule targeted delivery of siRNA, analytical chemistry, and DNA-encoded chemistry. In 2013, he joined AstraZeneca as Executive Director of the Discovery Sciences Chemistry Innovation Centre, where he led research projects in chemical biology, fragment-based lead generation, and computational chemistry. Since October 2016, Rob has been Vice President of Chemistry Innovation at Pharmaron, Inc. (Boston, MA), and is responsible for shared risk research programs. He has edited the book, *A Handbook for DNA-Encoded Chemistry: Theory and Applications for Exploring Chemical Space and Drug Discovery* (Wiley, 2014). M

product highlight

Enrich Your Chemical Toolbox

The fluoroalkylation toolbox has now been expanded beyond the standard Togni Reagents to include novel hypervalent iodine perfluoroalkylation reagents as well as fluoroalkyl bromides, silanes, carboxylates, and sulfonyl fluorides.

The installation of highly fluorinated groups into drug and pesticide candidates is a powerful strategy to modulate their properties. An incorporated fluorinated sidechain can tune the acidobasic behavior and lipophilicity, impart a dipole moment, lock a favorable conformation, and mitigate undesirable metabolic degradation of the parent compound. Fluorinated groups have been typically limited to a single fluorine or trifluoromethyl groups because they are easier to access through synthesis. More elaborate fluoroalkylation is now possible with the development of a new suite of reagents— including hypervalent iodine perfluoroalkylation reagents as well as fluoroalkyl bromides, silanes, carboxylates, and sulfonyl fluorides—that allow late-stage fluoroalkylation of a variety of functional groups through different reactivities.

To learn about the entire fluoroalkylation toolbox, visit
**SigmaAldrich.com/fluoroalkylation**

CF0035

CF0014

CF0031

CF0021

CF0012

CF0002
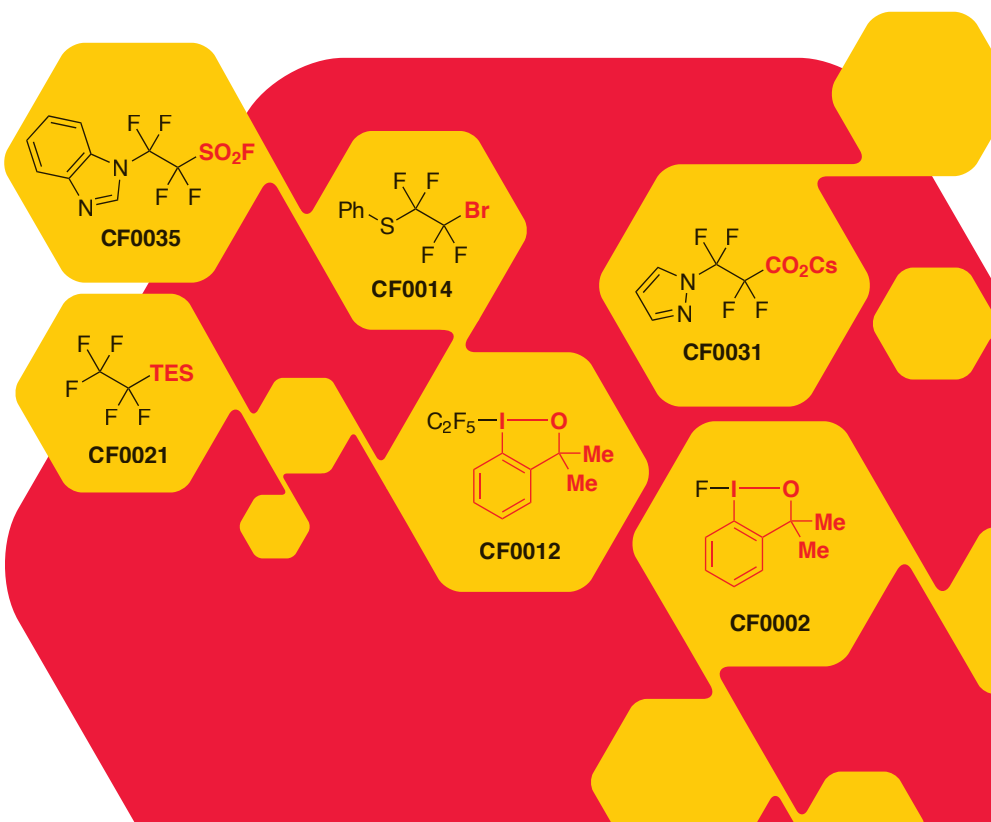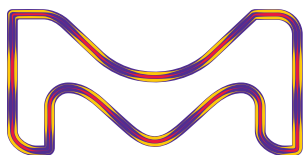
DESIGN THE PERFECT PROTAC®

Our degrader building blocks are the easiest way to generate libraries of protein degraders that can be screened for effective degradation of your target from cells.

What will be your perfect combination?

See our growing collection at
SigmaAldrich.com/TargetDegradation
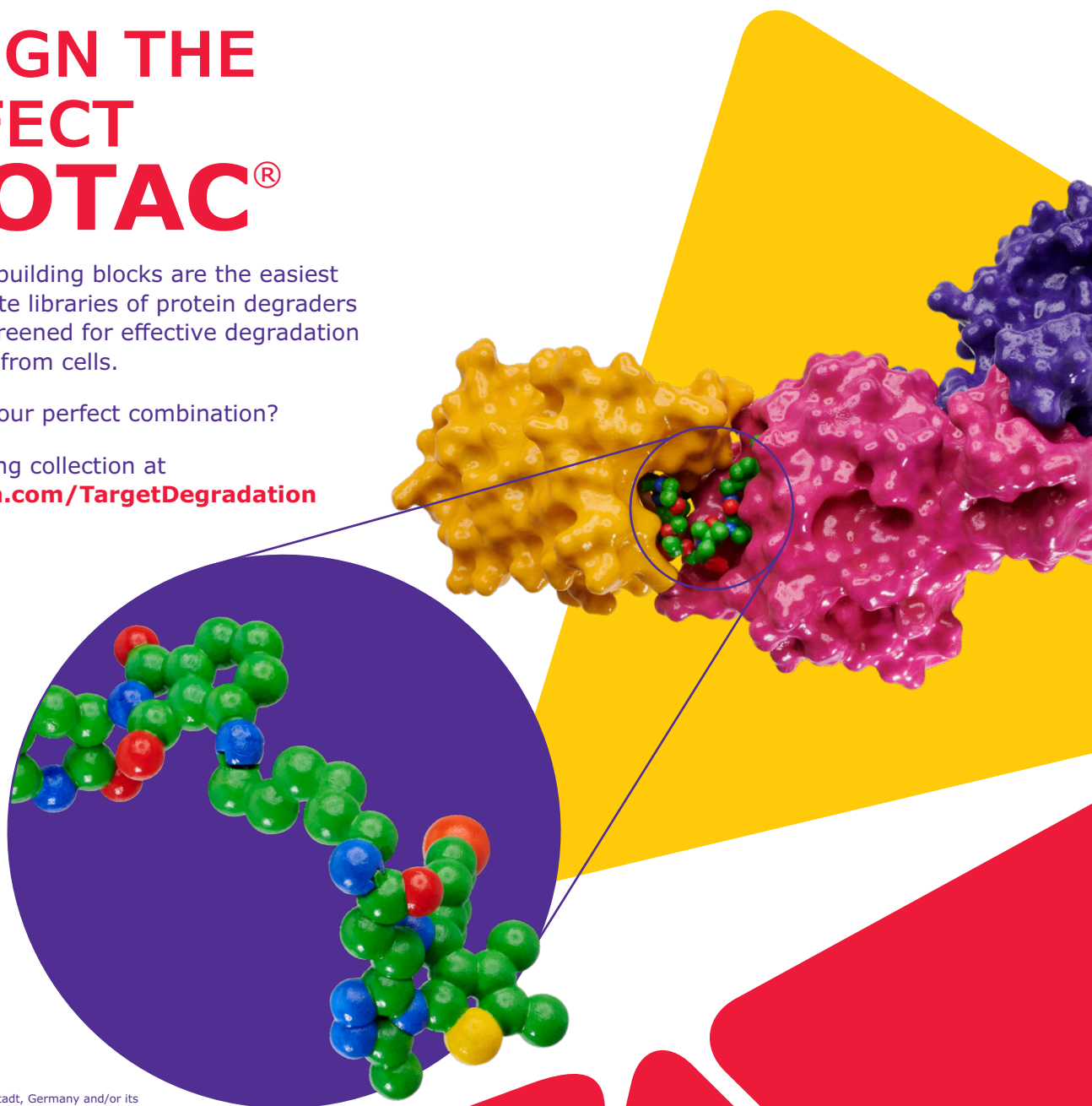
The life science business of Merck operates as MilliporeSigma in the U.S. and Canada.
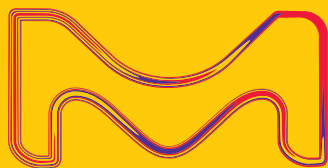
Sigma-Aldrich®
Lab & Production Materials

Get connected

Get ChemNews

Get current news and information about chemistry with our free monthly *ChemNews* email newsletter. Learn new techniques, find out about late-breaking innovations from our collaborators, access useful technology spotlights, and share practical tips to keep your lab at the fore.

For more information, visit
SigmaAldrich.com/ChemNews

The life science business of Merck operates as MilliporeSigma in the U.S. and Canada.

Sigma–Aldrich®

Lab & Production Materials