ACS Editors' Choice

ACCOUNTS of chemical research

Article

pubs.acs.org/accounts

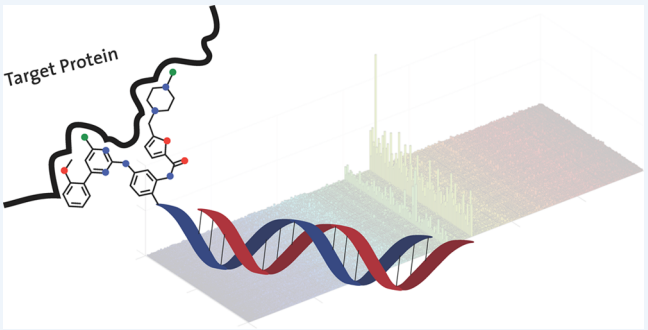
Terms of Use

DNA-Encoded Chemical Libraries: Advancing beyond Conventional Small-Molecule Libraries

Raphael M. Franzini, Dario Neri,* and Jörg Scheuermann*

Institute of Pharmaceutical Sciences, Department of Chemistry and Applied Biosciences, ETH Zürich, Wolfgang-Pauli-Strasse 10, 8093 Zürich, Switzerland

CONSPECTUS: DNA-encoded chemical libraries (DECLs) represent a promising tool in drug discovery. DECL technology allows the synthesis and screening of chemical libraries of unprecedented size at moderate costs. In analogy to phage-display technology, where large antibody libraries are displayed on the surface of filamentous phage and are genetically encoded in the phage genome, DECLs feature the display of individual small organic chemical moieties on DNA fragments serving as amplifiable identification barcodes. The DNA-tag facilitates the synthesis and allows the simultaneous screening of very large sets of compounds (up to billions of molecules), because the hit compounds can easily be identified and quantified by PCR-amplification of the DNA-barcode followed by high-throughput DNA sequencing. Several approaches have been used to generate DECLs, differing both in the methods used for library encoding and for the combinatorial assembly of chemical moieties. For example, DECLs can be used for fragment-based drug discovery, displaying a single molecule on DNA or two chemical moieties at the extremities of complementary DNA strands.



DECLs can vary substantially in the chemical structures and the library size. While ultralarge libraries containing billions of compounds have been reported containing four or more sets of building blocks, also smaller libraries have been shown to be efficient for ligand discovery. In general, it has been found that the overall library size is a poor predictor for library performance and that the number and diversity of the building blocks are rather important indicators. Smaller libraries consisting of two to three sets of building blocks better fulfill the criteria of drug-likeness and often have higher quality. In this Account, we present advances in the DECL field from proof-of-principle studies to practical applications for drug discovery, both in industry and in academia.

DECL technology can yield specific binders to a variety of target proteins and is likely to become a standard tool for pharmaceutical hit discovery, lead expansion, and Chemical Biology research. The introduction of new methodologies for library encoding and for compound synthesis in the presence of DNA is an exciting research field and will crucially contribute to the performance and the propagation of the technology.

DNA-encoded chemical libraries (DECLs) are collections of organic compounds, individually coupled to oligonucleotides or DNA fragments, serving as amplifiable identification barcodes. The embodiment of a "phenotype" (i.e., an organic small-molecule binder to a target protein of interest) and a "genotype" (i.e., a DNA sequence that permits the identification of the corresponding phenotypic molecule) is reminiscent of display technologies (e.g., phage display,[1,2] yeast display,[3] mRNA display,[4] ribosome display,[5] SELEX[6,7]), which have been used to generate large combinatorial libraries of biomacromolecules (e.g., peptides, proteins, antibodies, nucleic acids) and to isolate binders against a variety of target proteins[8] [Figure 1].

In the display technologies mentioned, the nucleic acids code for the biosynthesis of the corresponding polypeptides. These technologies have been expanded to allow for the incorporation of non-natural amino acids,[9−11] for the formation of cyclic structures,[12,13] and for modification with chemical moieties.[14] For example, the reaction of a small molecule scaffold containing two or three reactive groups with cysteine residues in the polypeptide chain has been applied to generate libraries of peptide bicycles.[15] By contrast, it is not the transcription/translation machinery that links the DNA sequence of DECLs to the synthesis of the organic molecule; instead the DNA merely acts as an unambiguous and amplifiable identification tag or contains information for hybridization steps during library synthesis. It is thus possible to construct libraries containing molecules with a substantially lower molecular weight compared with peptide libraries.

DECL technology was originally proposed by Lerner and Brenner to allow the construction of peptide libraries on beads that contain oligonucleotide barcodes by means of mutually compatible synthetic routes.[16] However, it was later shown that the technology could be implemented without the need for

Received: November 27, 2013
Published: March 28, 2014

   dx.doi.org/10.1021/ar400284t | Acc. Chem. Res. 2014, 47, 1247−1255

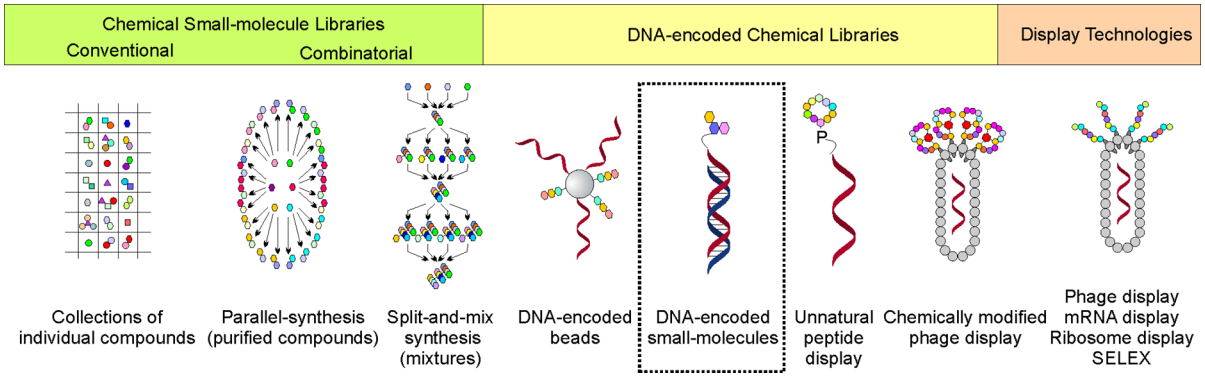


Figure 1. Comparison of different library types. DNA-encoded compound libraries are hybrid-type libraries bringing together beneficial aspects from conventional small-molecule chemical libraries and display technologies such as phage display. In particular, DNA-encoded chemical libraries combine the unparalleled level of high-throughput of phage display with the chemical space of conventional small-molecule chemical libraries. Several types of encoded chemical libraries have been described including DNA-encoded beads (one-bead−one-compound libraries), DNA-encoded small-molecule libraries (which are the focus of this Account) and libraries encoding peptides with chemical modifications and unnatural amino acids.

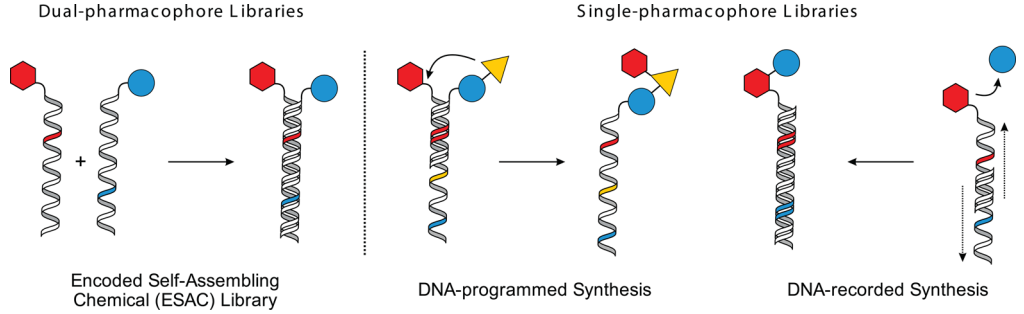


Figure 2. Basic strategies for the preparation of DNA-encoded small-molecule libraries. Encoded self-assembling chemical (ESAC) libraries are formed by the hybridization of two oligonucleotides displaying two pharmacophores at the adjoining termini. In single-pharmacophore libraries, the encoded compound is linked to only one oligonucleotide. Two main preparative approaches are available for the synthesis of single-pharmacophore libraries. In DNA-programmed synthesis, an initial oligonucleotide template contains the information for the synthesis of the final small-molecule product. The template sequence can be converted, for example, by sequential DNA-templated reactions or by hybridization-based sorting. Alternatively, DNA-recorded synthesis provides libraries by consecutive cycles of chemical reaction and encoding steps.

beads, thus allowing the construction of larger libraries and more efficient screening procedures [Figure 1].[17−19]

The most direct comparators for DECLs are probably conventional libraries of small organic molecules. These libraries can be obtained and screened as sets of individual molecules or as combinatorial mixtures of compounds prepared by split-and-pool methodologies [Figure 1]. Conventional small-molecule libraries provide great freedom for selecting the chemical structures of choice, and screening such libraries has provided numerous drug leads. Furthermore, the screening of individual molecules (i.e., those obtained from various sources or by parallel synthesis strategies) has the potential advantage of being compatible with phenotypic assays.[20] However, the synthesis and screening of individual molecules is an expensive and complex endeavor, requiring large amounts of target protein, a suitable bioassay, and expensive logistics. Even at large pharmaceutical companies screening campaigns are typically limited to a few hundred thousand compounds. By contrast, the use of combinatorial libraries not encoded by DNA (e.g., one-bead−one-compound libraries) has met with limited success, because it is extremely difficult to identify bioactive molecules present at low concentration in a large mixture of compounds. Positional scanning may be another possibility for identifying optimal binders but is mostly limited to peptides and requires a known lead structure.[21]

Several approaches have been proposed for the synthesis and screening of DECLs [Figure 2]. One fundamental aspect is the distinction between "single-pharmacophore" and "dual-pharmacophore" chemical libraries, depending on whether a single molecule is coupled to a DNA fragment, or whether two molecules are attached at the adjoining extremities of two complementary DNA strands [Figure 2]. This latter technology has also been named "encoded self-assembling chemical libraries" (ESAC library technology).[18]

For the synthesis of single-pharmacophore chemical libraries, two distinct general approaches are routinely applied. In *DNA-programmed synthesis* [Figure 2], molecules are assembled on an oligonucleotide strand, which may contain the DNA codes for the identification of building blocks. With the help of complementary reactant oligonucleotide derivatives, the oligonucleotide can drive the synthesis by DNA-templated chemical reactions.[22] Heteroduplex formation increases the local concentration of reagents, which may facilitate chemical reactions, followed by linker cleavage, purification, and restoration of the original single-stranded oligonucleotide template carrying the growing chemical structure of interest.[17] Additionally, the encoding oligonucleotides can be used to cause sequence-defined localization in specific reaction vessels for chemical modification.[19] The use of DNA-templated reactions has the advantage that the high effective molarity of

dx.doi.org/10.1021/ar400284t | *Acc. Chem. Res.* 2014, 47, 1247−1255

the reagents enables the use of reactions that normally would be incompatible with or low-yielding under standard bioconjugation conditions. A disadvantage of DNA-programmed synthesis is that it requires the preparation of reactive oligonucleotide conjugates, which can be laborious. Furthermore, because the DNA code is synthesized prior to compound assembly, the generation and purification of a large set of coding oligonucleotides is necessary. Two designs have been described to overcome this limitation. Researchers at Vipergen have made use of three-way junctions for concomitant compound synthesis and encoding obviating the need for an *a priori* template.[23] In a recent report, Li and co-workers have described a method where a universal template is used with inosine stretches, which are promiscuous with respect to base-pairing, and where the sequence tags for the building blocks are added during synthesis.[24] Alternatively, small organic molecules can be coupled to oligonucleotides using a split-and-pool synthesis approach, sometimes referred to as *DNA-recorded synthesis* [Figure 2].[25−28] Here, individual molecules are coupled to oligonucleotides carrying distinct short sequences serving as identification barcodes for the corresponding chemical moieties. After successful coupling and purification, encoded compounds are mixed and subsequently divided into aliquots, for a successive round of chemical synthesis with a second set of building blocks followed by elongation of the oligonucleotide codes with corresponding sequences [Figure 2].[25−27] The chemical coupling step is temporally separated from the DNA-encoding step, which can be performed using partially complementary oligonucleotide strands[25] or by ligation of double-stranded DNA fragments.[27]

LIBRARY SIZE AND NUMBER OF BUILDING BLOCKS

The probability of identifying hit compounds depends both on the number of molecules in the library and on their structures. When evaluating libraries, in addition to the total number of compounds also the number of modular building blocks used for synthesis is relevant. Reported DECLs vary substantially with respect to these two descriptors [Figure 3]. When libraries are built by the assembly of several sets of building blocks, library size grows exponentially with the number of reaction steps. This strategy has enabled the synthesis of compound collections of unprecedented size, including, for example, a library of 10^8 peptoid 8-mers[29] or $>10^9$ organic compounds consisting of four building blocks.[30] The multistep construction of very large combinatorial libraries may be appealing, yet it also has several limitations. Consecutive reactions tend to increase the average molecular weight of the compounds. As reported by Lipinski,[31] compounds with a molecular mass of >500 Da have an elevated risk of failing in clinical trials, and controlling the molecular mass is an important design consideration for DECLs. Library purity is also a concern when several reaction cycles are applied. DECLs with >4 building blocks have been prepared without purification steps and control over reaction yields, possibly affecting the quality of these libraries.[27,29,30] To the other extreme, the coupling of single molecules to individual oligonucleotides becomes prohibitively expensive for very large compound collections but may nevertheless be a practical avenue for libraries containing 10^2−10^3 members.[32,33]

Often, the hit discovery potential of DECLs is best achieved with libraries based on two to three sets of building blocks. Even relatively small libraries (10^3−10^6 compounds) may allow

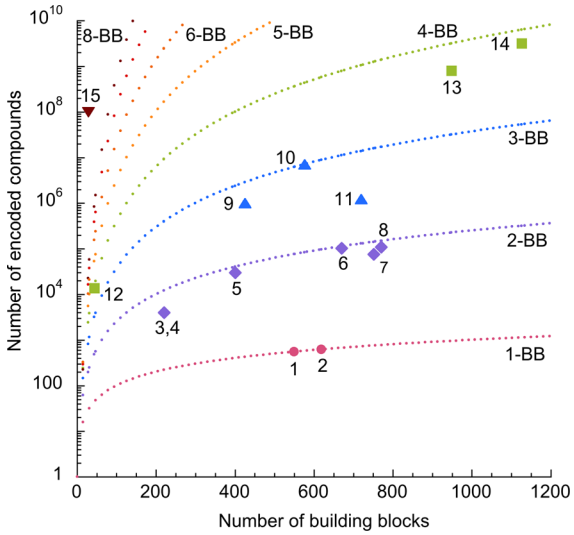


Figure 3. Summary of library sizes of representative DNA-encoded small molecule libraries. Libraries are grouped according to the number of building blocks (BB) per compound (1BB, red circles; 2BB, purple diamonds; 3BB, blue triangles; 4BB, green squares; 8BB, brown triangles). Dotted lines are curves of the total number of compounds as a function of the number of building blocks for hypothetical libraries consisting of equally sized sets of building blocks. Numbering of libraries: 1 = Scheuermann et al.;[32] 2 = Dumelin et al.;[33] 3 = Mannocci et al.;[25] 4 = Buller et al.;[26] 5 = Leimbacher et al.;[28] 6 = Franzini et al., unpublished; 7 = Samain et al. unpublished; 8 = Scheuermann et al., unpublished; 9 = Buller et al.;[35] 10 = Clark et al.;[27] 11 = Disch et al.;[36] 12 = Tse et al.;[37,38] 13 = Clark et al.;[27] 14 = Deng et al.;[30] 15 = Wrenn et al.[29]

the isolation of useful binding specificities [Table 1]. For example, a DECL with only 30 000 compounds enabled the discovery of a drug fragment for interleukin-2[28] and a lead expansion library with only 4000 compounds containing a benzamidine lead structure provided nanomolar inhibitors of serine proteases.[34] These results suggest that the total compound number in a library is not the only predictor of screening success.

Combinatorial libraries of DNA-encoded polypeptides can be very large (i.e., billions of compounds) but are typically made of the 20 proteinogenic amino acids and possibly a few unnatural amino acids.[15,39] High binding affinities can be achieved, but molecules tend to have a high molecular mass. As a consequence, oral administration and targeting of intracellular components remain challenging. DECLs constructed by DNA-templated synthesis methods so far rely on the use of a limited set of reactive oligonucleotide derivatives, which is typically on the order of 20−200.[38] As a consequence, the technology is particularly suited for the construction of small macrocyclic peptides,[17] but other structures could also be considered. Split-and-pool approaches, where sets of building blocks (often commercially available) can be added to a nascent chemical structure on solid phase or in solution, probably provide the largest versatility in library construction. Intermediates of library synthesis can be purified by HPLC only after the first reaction step, and subsequent modifications crucially rely on reactions with high yield and broad substrate specificity.

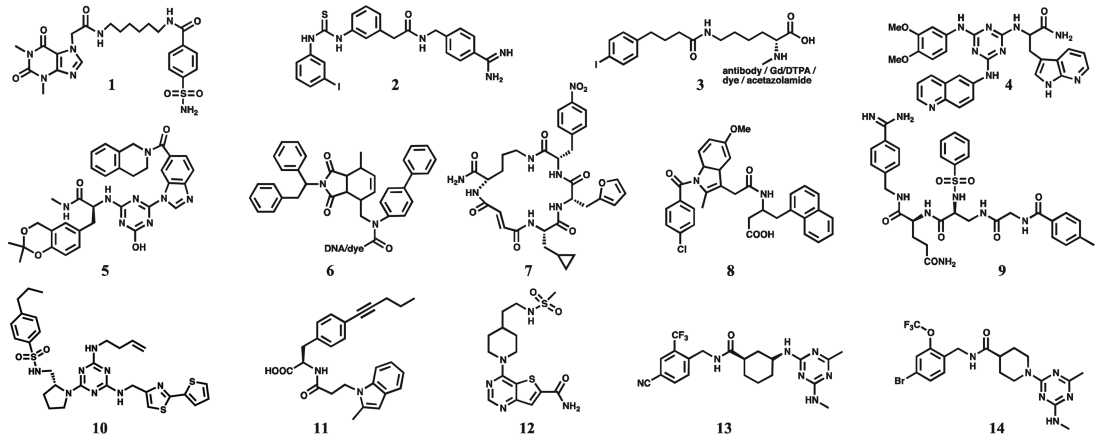
SUCCESS STORIES

The first DECLs were spiked with known protein ligands, which were later isolated in selection experiments (Table 1,

Table 1. DNA-Encoded Chemical Libraries in Practice

Timeline	Institution	Library Size	Hit Compound	Target Protein	Selection Results
Halpin & Harbury, 2004[19]	Stanford University	10^6		anti-enkephalin mAb	Proof-of-Principle
Gartner et al. 2004[17]	Harvard University	65		Carbonic anhydrase (CA)	Proof-of-Principle
Melkko et al. 2004[18]	ETH Zurich	138	**1**	CA II	12 nM (based on lead structure)
Melkko et al. 2007[56]	ETH Zurich	550	**2**	Trypsin	98 nM (based on lead structure)
Wrenn et al. 2007[29]	Stanford University	10^8		CrkSH3	16 μM (peptoid library)
Mannocci et al. 2008[25]	ETH Zurich	4000		Streptavidin	0.35 μM (de novo)[b]
Mannocci et al. 2008[25]	ETH Zurich	4000		IgGs	chromatography resin
Dumelin et al. 2008[33]	Philochem	550	**3**	HSA	3.2 μM (de novo)[a]
Scheuermann et al. 2008[32]	ETH Zurich	550		MMP-3	9.9 μM (de novo)[a]
Clark et al. 2009[27]	GlaxoSmithKline	ca. $7x10^6$		p38 MAP kinase	≤2 nM (based on lead structure)[a]
Clark et al. 2009[27]	GlaxoSmithKline	ca. $7x10^6$	**4**	Aurora A kinase	0.27 μM (de novo)[a]
Clark et al. 2009[27]	GlaxoSmithKline	ca. $8x10^8$	**5**	p38 MAP kinase	0.25 μM (de novo)[a]
Hansen et al. 2009[23]	Vipergen	100		Anti-enkephalin mAb	Proof-of-Principle
Buller et al. 2009[43]	ETH Zurich	4000	**6**	TNF	20 μM (de novo)[b]
Buller et al. 2009[43]	ETH Zurich	4000		Bcl-xL	10 μM (de novo)[b]
Kleiner et al. 2010[38]	Harvard University	13824	**7**	Src kinase	680 nM (de novo)[a]
Melkko et al. 2010[40]	Philochem	4000	**8**	Bcl-xL	0.93 μM (de novo)[b]
Mannocci et al. 2010[34]	Philochem	8000	**9**	Trypsin	3 nM (based on lead structure)[a]
Buller et al. 2011[35]	ETH Zurich/Philochem	10^6		CAIX	<0.5 μM(based on lead structure)[a]
Deng et al. 2012[30]	GlaxoSmithKline	$4x10^9$	**10**	ADAMTS-5	30 nM (de novo)[a]
Leimbacher et al. 2012[28]	ETH Zurich/Philochem	30000	**11**	IL-2	2.5 μM (de novo)[b]
Disch et al. 2013[36]	GlaxoSmithKline	$1.2x10^6$	**12**	SIRT1/2/3	4/1/7 nM (de novo)[a]
Podolin et al. 2013[41]	GlaxoSmithKline	undisclosed	**13**	sEH	27 pM (after MedChem)[a]
Thalji et al. 2013[42]	GlaxoSmithKline	undisclosed	**14**	sEH	0.5 nM (after MedChem)[a]

[a] **IC50 measurement**
[b] **Kd measurement by fluorescence polarization**



"proof-of-principle"), thereby proving both the completeness of library assembly and the principle of selection decoding.[17−19] Encouraged by such proof-of-principle results, large libraries were constructed, in order to use the technology for *de novo* selections.[27,29,30] However, even small libraries can yield novel active compounds. For example, interleukin-2 ligands (**11**),[28] portable albumin binders (**3**),[33] inhibitors of Bcl-xL (**8**),[40] and inhibitors of kinases (**7**)[38] have been identified.

GlaxoSmithKline has described several large DECLs based on split-and-pool synthesis and identified ligands of pharmaceutical interest. In 2009, they described the construction of a triazine-based 7 million-membered 3-building block library and of a 4-building block library containing 800 million compounds.[27] Potent aurora A kinase inhibitors (**4**) (IC_{50} = 270 nM) and p38MAP kinase inhibitors (**5**, IC_{50} = 250 nM hit compound from library screening; IC_{50} = 7 nM after optimization) were reported.[27] More recently, using libraries of a similar design, containing a central triazine core, the authors reported the isolation of a 30 nM ADAMTS-5 inhibitor (**10**).[30] Moreover, efficient triazine-based inhibitors have been

dx.doi.org/10.1021/ar400284t | *Acc. Chem. Res.* 2014, 47, 1247−1255

described for soluble epoxide hydrolase (sEH) (**13**, **14**),[41,42] but the design of the corresponding libraries used for selections has not been disclosed. The same company reported a high-affinity pan-inhibitor of the sirtuin-type histone deacetylase family inhibiting SIRT1 (**12**, IC_{50} = 4 nM), SIRT2 (IC_{50} = 1 nM), and SIRT3 (IC_{50} = 7 nM).[36]

Our group used a two-building block library derived from Diels–Alder chemistry[26] for the *de novo* discovery of a TNF-α binder (**6**, K_d = 10 μM), and a Bcl-xL binder (K_d = 10 μM).[43] A set of sulfonamide-containing carbonic anhydrase IX inhibitors (IC_{50} = 240 nM), which were able to recognize the cognate antigen in hypoxic tumor sections, were isolated out of a one million-membered three-building block library.[35] Other two-building block libraries yielded specific low micromolar binders to Bcl-xL (**8**, K_d = 930 nM)[40] and interleukin-2 (**11**, K_d = 2.5 μM).[28]

Philochem and collaborators described a small two-building block "lead-expansion" library, containing the known non-specific serine protease binder benzamidine, from which a high-affinity trypsin binder (**9**, IC_{50} = 3 nM) with high specificity could be isolated.[34]

Liu and co-workers were the first to report novel binders from a library synthesized by DNA-templated reactions. A library of 13 824 macrocycles was screened against a set of 36 proteins and yielded two specific high-affinity inhibitors of Src kinase (**7**, IC_{50} of 680 and 960 nM), as well as low micromolar inhibitors of Pim1, Akt3, MAPKAPK2, and p38-MAPKAP2.[38] From the initial Src kinase hit, a nanomolar inhibitor (IC_{50} = 99 nM) could be derived by systematically varying amino acid residues.[44]

In addition to libraries displaying conventional small molecules, DNA- or PNA-encoded peptide or peptoid libraries have been built and used for affinity-based screening.[45,46] Selections with a 100 million peptoid library were performed against the SH3 domain of Crk, yielding a 16 μM binder, which was superior to the natural Crk-SH3 ligands.[29]

COMPARISON OF CONVENTIONAL LIBRARIES AND DECLS

Conventional chemical libraries consist of individual compounds, which need to be synthesized, quality-controlled, and screened one by one. The costs for the synthesis, management, and screening of large chemical libraries (e.g., 1 million compounds) are enormous. The quality of compounds from commercial sources can also be substandard.[47]

In principle, pools of organic compounds could be screened simultaneously, for example, by affinity capture on a target protein of interest.[48] However, the procedure is prone to artifacts and typically relies on mass-spectrometric identification of compounds. In general, the screening of combinatorial libraries or of pools of compounds becomes more and more difficult with increasing library size, as the concentration of individual compounds progressively decreases. Similar considerations apply also for the "one bead–one compound" approach. Library encoding with a nucleic acid arguably represents the only general avenue for obtaining binding data in a very large compound collection, because the DNA "barcode" provides quantitative information about the ability of individual molecules to interact with the target protein(s) of interest. Indeed, PCR amplification can theoretically detect a single DNA-encoded molecule that survives the selection procedure.

Since DECLs are typically constructed by split-and-pool methods, library purity cannot be directly assessed using analytical methods, but only estimated based on representative reaction yields or assessed indirectly by the performance of selection experiments. However, when comparing conventional chemical libraries and DECLs, many features play to the advantage of encoded libraries, as discussed in the following paragraphs.

Costs and Logistics

The cost of reagents for the construction and screening of DECLs can be orders of magnitude lower, compared with conventional libraries. For example, the synthesis of a one-million compound DNA-encoded chemical library consisting of 100 × 100 × 100 building blocks requires merely the purchase of 300 oligonucleotides and of 300 chemical compounds. Once synthesized on nanomole scale, the library is sufficient for thousands of affinity-based selections and can be stored as a compound pool in a normal freezer. Selections need only minute amounts of target protein (i.e., microgram quantities), do not require expensive robotics, and can be performed in many different screening conditions. Also the decoding steps are inexpensive, in view of the enormous progress made in the field of high-throughput DNA sequencing.[25,49,50]

Time

The construction of a DNA-encoded chemical library can be much faster compared with the parallel synthesis of individual molecules. While conventional split-and-pool libraries can be rapidly synthesized, they are not compatible with efficient decoding procedures. The synthesis of a typical DECL can be performed in a few months. Once available, multiple targets can be screened in parallel in various experimental conditions, whereas a single conventional high-throughput screening campaign (one molecule at a time) may require days or weeks.

Purity and Ease of Synthesis

A DECL can be of variable purity, depending on design, execution and reaction schemes. Libraries constructed using multiple reaction cycles tend to have lower purity. As a notable exception, encoded self-assembling chemical (ESAC) libraries enjoy the benefit of being composed of individually purified sublibraries. Not all reactions are compatible with the presence of a DNA moiety, but a large variety of synthetic schemes have been shown to be feasible, including Diels–Alder cycloadditions,[26] nucleophilic aromatic substitution,[27] and the Wittig reaction,[17] to mention only a few. Most DECLs reported to date contain at least one peptide bond formation step during assembly. The DNA tag may also facilitate synthetic procedures, by ease of separation of products from starting reagents and by facile mass-spectrometric confirmation of reaction product identity. A particular case with respect to library purity is the synthesis of macrocycles using DNA-templated chemistry. A clever design of the reaction strategy enables the purification of the reaction product after each synthesis step.[37]

The Impact of the DNA Moiety of DECLs on Library Screening

The presence of DNA on individual compounds typically does not interfere with affinity selection procedures, since double-stranded DNA has a fairly rigid "rod-like" structure, which is conserved for all library members. Furthermore, DNA is typically spaced away from the compounds by a flexible linker of sufficient length, in order to minimize electrostatic effects. In

dx.doi.org/10.1021/ar400284t | *Acc. Chem. Res.* 2014, 47, 1247–1255

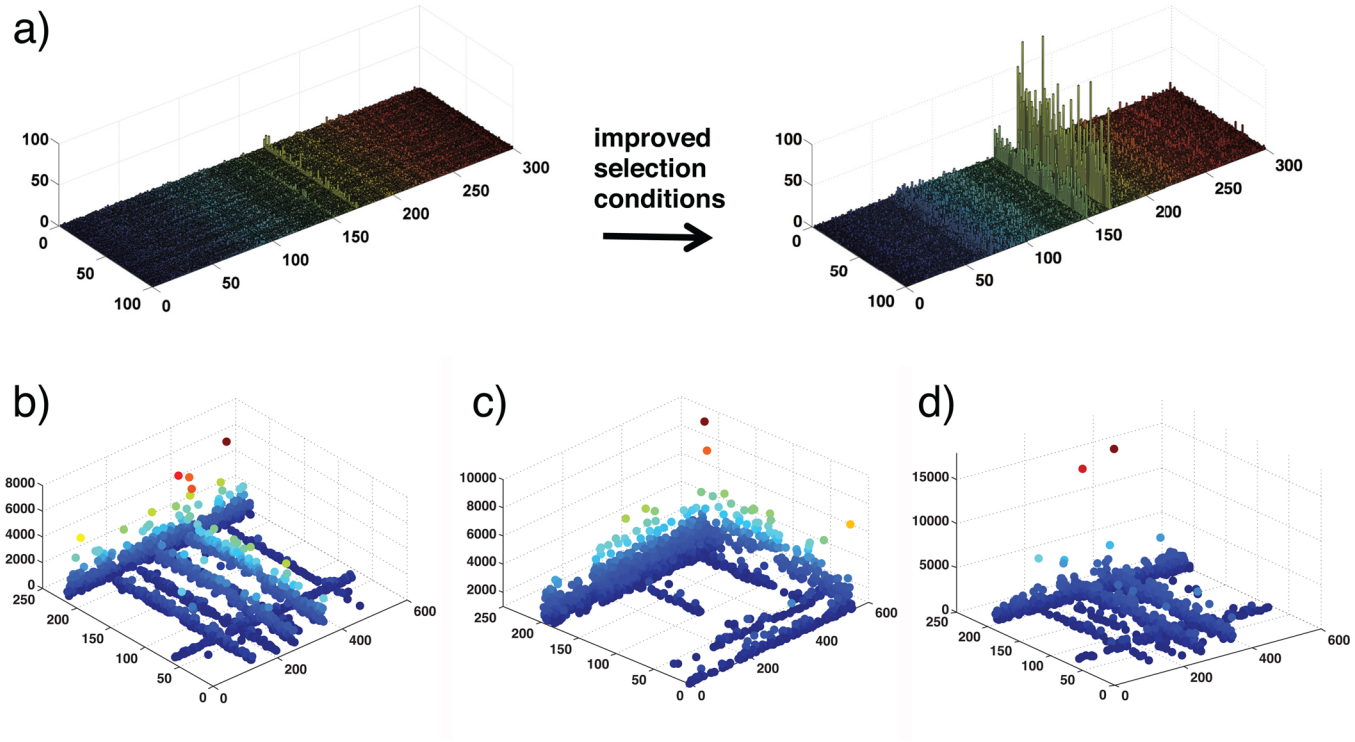


Figure 4. Screening results of DECL selections obtained by high-throughput DNA sequencing. (a) Histogram plots of two selections against a target protein (carbonic anhydrase IX) performed under different selection conditions, depicting library building blocks on the $x-y$ plane and relative frequencies after selection on the z-axis.[28] (b–d) Examples of fingerprints of selections against three undisclosed target proteins. Enrichment of compounds is given as colored dots, ranging from blue to red.

theory, potential problems could arise when screening for binders to DNA-binding proteins. Indeed, a large excess of herring or salmon sperm DNA is typically used in selection procedures for saturating potential DNA-binding sites.[27,51]

An important feature of DECLs relates to the fact that the linker between compound and DNA points away from the target protein of interest and provides a direct avenue for subsequent chemical modification steps (e.g., increase of compound solubility or attachment of a payload).

Solubility and Artifacts

DNA also confers high water solubility to otherwise insoluble molecules, thus facilitating selection procedures. Furthermore, individual library members can be used at low (e.g., femtomolar) concentrations, thus minimizing screening artifacts associated with precipitation or aggregation phenomena. Other artifacts may arise from several origins and may be common both to conventional libraries and DECLs, for example, artifacts resulting from target unfolding, lack of purity, or library impurities. However, DNA-encoded libraries can be screened in multiple conditions, thus increasing the chance to identify artifacts.

Ease of Screening

In analogy to protein-display technologies, DECLs are typically screened by affinity-capture selections [Figure 4]. A target protein of interest, usually immobilized on a solid support, is incubated with the DECL, and binding compounds are physically separated from nonbinders. Unlike conventional libraries, which need the *ad hoc* development of dedicated screening methods, DECL selections do not require a bioassay, but only the possibility of implementing an affinity capture procedure. Several conditions can be robotically implemented in library selection schemes, varying experimental parameters such as buffer composition or stringency of washing. While, in principle, compounds may be found to bind to any accessible site of the target protein, most identified hits tend to accommodate in clefts of the target protein, including the active sites of enzymes. Depending on the conditions used for selection the affinity of the obtained binders can vary broadly from single-digit nanomolar to high micromolar.

Structure–Activity Relationships (SAR)

A major bottleneck in small-molecule drug development is the evolution of early hit structures to lead compounds, which often involves the determination of structure–activity relationships (SAR) by the synthesis and evaluation of sets of hit-derived compounds. Simple structure–activity patterns may be directly accessible from the relative sequence enrichment of encoded compounds in DECL selection experiments, which would provide valuable information for subsequent medicinal chemistry optimization. In order to provide reliable SAR data, DECLs need to have an adequate level of purity and structural diversity that allow for a meaningful comparison, and further studies will be necessary to reveal the full potential of DECLs for this application.

Addressed Chemical Space

The chemical space of compounds that can be theoretically synthesized[52] vastly exceeds the number of compounds that can be screened, even using DNA-encoding procedures. The main strengths of DECL technology relate to the large number of compounds that can be synthesized and easily screened. However, the modular nature of typical DECL members may cause undesired liabilities for pharmaceutical development, such as suboptimal ligand efficiency or a difficult geometric fit in

dx.doi.org/10.1021/ar400284t | *Acc. Chem. Res.* 2014, 47, 1247–1255

tight protein pockets. These liabilities may be due to the intrinsic nature of chemical reactions used in DECL synthesis. On the other hand, the larger structures of DECL compounds consisting of multiple building blocks may be both an advantage when selecting against large epitopes and a disadvantage when selecting against enzymes and for downstream medicinal chemistry optimization. A long-term challenge and opportunity for DECL research consists in the development of novel synthetic schemes that are compatible with the creation of large compound libraries whose functional groups are arranged on a small and geometrically defined (rigid) scaffold.

Hit-to-Lead Development

A major challenge in drug discovery is the progression of hit compounds to lead structures for pharmaceutical development. Different screening technologies tend to provide different starting points for hit-to-lead development,[53] and hits from DECLs tend to be large with relatively low ligand efficiencies. Moreover, DNA conjugation may mask potential solubility problems. It is therefore important to take parameters such as drug-likeness and solubility into account already at the design stage.

DECL technology can however also be implemented as a tool for hit-to-lead optimization. Lead-expansion libraries may consist of a variable part combined with a known drug fragment for the examined target. For example, a lead-expansion library containing a benzamidine lead provided nanomolar inhibitors of serine proteases.[34] Furthermore, the obtained structure−affinity information can be helpful for hit-to-lead development.

■ CHALLENGES AND OPPORTUNITIES

DECL technology has advanced during recent years from proof-of-principle experiments to a potent method for pharmaceutical hit discovery. The possibility of preparing libraries with sizes previously unconceivable for small-molecule collections, together with efficient screening protocols, makes DECL a very promising technology for pharmaceutical research. However, further advances are required in order to establish DECLs as routine tools for drug discovery. A major challenge remains associated with the need to expand library size while preserving the drug-likeness of compounds and library purity. For this aim, it will be necessary to identify and optimize chemical reactions that are compatible with DNA and which provide access to structurally diverse building blocks. These reactions must then be implemented in library design schemes, which privilege structural compactness of the encoded compounds. Furthermore, the development of novel approaches for library purification and for quality control will be of great significance.

Several approaches to DECL technology have been conceived and successfully applied to the discovery of protein binders. Until now, however, binders to a relatively small set of target proteins have been reported [Table 1]. Moreover, little is known about the impact of library design and experimental factors on the successful (or unsuccessful) use of DECLs. More systematic studies are needed in order to gain statistical information on the performance of DECLs in pharmaceutical hit discovery. Structural details of library synthesis are sometimes not disclosed in publications, which hinders a comparative evaluation of different DECL approaches and different library designs. Some companies active in the field do not publish at all. In time, a comparative evaluation of the performance of different libraries on the same protein targets will shed light on general chemical rules, which are important for selection success. It is very well possible that different classes of protein targets may be best screened with different types of libraries and that no single "all-purpose" library exists.

In summary, DECLs have been established as an emerging tool for drug discovery, with a promising potential for hit discovery and lead expansion. Encoded libraries may provide structural information of broad utility to several areas of Medicinal Chemistry research, including fragment-based drug discovery.[54,55] DECLs are having an increasing impact on the drug discovery process not only in the pharmaceutical industry but also in academia, which can (for the first time) afford the construction and screening of libraries containing millions of compounds. Undoubtedly, encoded libraries will remain an active area of research, in which inventiveness in library construction and selection schemes may bear fruit not only for drug discovery applications but also for the study of biological processes.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: dario.neri@pharma.ethz.ch.
*E-mail: joerg.scheuermann@pharma.ethz.ch.

Notes

The authors declare the following competing financial interest(s): RF consults for Philochem, DN is a co-founder and shareholder of Philochem and JS is a board member of Philochem.

Biographies

Dr. Raphael M. Franzini has received his Master degree in Chemistry from EPF Lausanne (Switzerland). He then obtained his Ph.D. in Chemistry from Stanford University working on DNA-templated fluorescence detection of nucleic acids under the guidance of Prof. Eric Kool. He continued as an SNF postdoctoral researcher at Stanford University in the groups of Profs. Matthew Bogyo and Eric Kool. In 2012, he joined the group of Prof. Dario Neri at ETH Zurich as a VPFW-ETH postdoctoral fellow. In the group of Prof. Neri, he has advanced the development, synthesis, and application of DNA-encoded small-molecule libraries.

Prof. Dr. Dario Neri studied Chemistry at the Scuola Normale Superiore of Pisa (Italy) and obtained his Ph.D. at the ETH Zurich (Switzerland) under the guidance of professor Kurt Wüthrich. After postdoctoral research with Sir Gregory Winter at the Medical Research Council, Cambridge (U.K.), in 1996, he returned to the ETH Zürich as a professor. His research focuses on strategies for the targeted delivery of therapeutic effectors to sites of disease and the development of DNA-encoded chemical libraries. He is cofounder of Philogen SpA, an Italian−Swiss biotech company, which has brought several antibody drugs into clinical development programs.

Dr. Jörg Scheuermann studied Chemistry at the Ruprecht-Karls-University Heidelberg (Germany) and ETH Zurich (Switzerland). During his Ph.D. study in the group of professor Neri at the ETH Zurich, he worked on the identification of novel binding molecules to markers of angiogenesis. Together with Dario Neri, he co-developed DNA-encoded self-assembling chemical (ESAC) libraries. He is currently working on his habilitation in the field of DNA-encoded chemical libraries and targeted cytotoxics and their application in drug discovery and development.

■ ACKNOWLEDGMENTS

We thank ETH Zurich, Swiss National Science Foundation, and Philochem AG for financial support. R.M.F. acknowledges a postdoctoral fellowship of ETH and Marie Curie Actions (VPFW-ETH fellowship). The authors are grateful to Moreno Wichert, Willy Decurtins, Nikolaus Krall, and Dr. Florent Samain for comments on the manuscript.

■ REFERENCES

(1) McCafferty, J.; Griffiths, A. D.; Winter, G.; Chiswell, D. J. Phage antibodies: Filamentous phage displaying antibody variable domains. *Nature* **1990**, *348*, 552−554.
(2) Kang, A. S.; Barbas, C. F.; Janda, K. D.; Benkovic, S. J.; Lerner, R. A. Linkage of recognition and replication functions by assembling combinatorial antibody Fab libraries along phage surfaces. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*, 4363−4366.
(3) Boder, E. T.; Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **1997**, *15*, 553−557.
(4) Wilson, D. S.; Keefe, A. D.; Szostak, J. W. The use of mRNA display to select high-affinity protein-binding peptides. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 3750−3755.
(5) Hanes, J.; Pluckthun, A. In vitro selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 4937−4942.
(6) Ellington, A. D.; Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **1990**, *346*, 818−822.
(7) Tuerk, C.; Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **1990**, *249*, 505−510.
(8) Ponsel, D.; Neugebauer, J.; Ladetzki-Baehs, K.; Tissot, K. High affinity, developability and functional size: the holy grail of combinatorial antibody library generation. *Molecules* **2011**, *16*, 3675−3700.
(9) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. Expanding the genetic code of *Escherichia coli*. *Science* **2001**, *292*, 498−500.
(10) Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z.; Schultz, P. G. An expanded eukaryotic genetic code. *Science* **2003**, *301*, 964−967.
(11) Murakami, H.; Ohta, A.; Ashigai, H.; Suga, H. A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nat. Methods* **2006**, *3*, 357−359.
(12) Koivunen, E.; Wang, B.; Ruoslahti, E. Phage libraries displaying cyclic peptides with different ring sizes: Ligand specificities of the RGD-directed integrins. *Bio-Technology (New York)* **1995**, *13*, 265−270.
(13) Kawakami, T.; Murakami, H.; Suga, H. Messenger RNA-programmed incorporation of multiple N-methyl-amino acids into linear and cyclic peptides. *Chem. Biol.* **2008**, *15*, 32−42.
(14) Ng, S.; Jafari, M. R.; Derda, R. Bacteriophages and viruses as a support for organic synthesis and combinatorial chemistry. *ACS Chem. Biol.* **2012**, *7*, 123−138.
(15) Heinis, C.; Rutherford, T.; Freund, S.; Winter, G. Phage-encoded combinatorial chemical libraries based on bicyclic peptides. *Nat. Chem. Biol.* **2009**, *5*, 502−507.
(16) Brenner, S.; Lerner, R. A. Encoded combinatorial chemistry. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 5381−5383.
(17) Gartner, Z. J.; Tse, B. N.; Grubina, R.; Doyon, J. B.; Snyder, T. M.; Liu, D. R. DNA-templated organic synthesis and selection of a library of macrocycles. *Science* **2004**, *305*, 1601−1605.
(18) Melkko, S.; Scheuermann, J.; Dumelin, C. E.; Neri, D. Encoded self-assembling chemical libraries. *Nat. Biotechnol.* **2004**, *22*, 568−574.
(19) Halpin, D. R.; Harbury, P. B. DNA display II. Genetic manipulation of combinatorial chemistry libraries for small molecule evolution. *PLoS Biol.* **2004**, *2*, 1022−2030.
(20) Lee, J. A.; Uhlik, M. T.; Moxham, C. M.; Tomandl, D.; Sall, D. J. Modern phenotypic drug discovery is a viable, neoclassic pharma strategy. *J. Med. Chem.* **2012**, *55*, 4527−4538.
(21) Placzek, W. J.; Sturlese, M.; Wu, B.; Cellitti, J. F.; Wei, J.; Pellecchia, M. Identification of a novel Mcl-1 protein binding motif. *J. Biol. Chem.* **2011**, *286*, 39829−39835.
(22) Li, X.; Liu, D. R. DNA-templated organic synthesis: nature's strategy for controlling chemical reactivity applied to synthetic molecules. *Angew. Chem., Int. Ed.* **2004**, *43*, 4848−4870.
(23) Hansen, M. H.; Blakskjaer, P.; Petersen, L. K.; Hansen, T. H.; Hojfeldt, J. W.; Gothelf, K. V.; Hansen, N. J. A yoctoliter-scale DNA reactor for small-molecule evolution. *J. Am. Chem. Soc.* **2009**, *131*, 1322−1327.
(24) Li, Y.; Zhao, P.; Zhang, M.; Zhao, X.; Li, X. Multistep DNA-Templated Synthesis Using a Universal Template. *J. Am. Chem. Soc.* **2013**, *135*, 17727−17730.
(25) Mannocci, L.; Zhang, Y.; Scheuermann, J.; Leimbacher, M.; De Bellis, G.; Rizzi, E.; Dumelin, C.; Melkko, S.; Neri, D. High-throughput sequencing allows the identification of binding molecules isolated from DNA-encoded chemical libraries. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 17670−17675.
(26) Buller, F.; Mannocci, L.; Zhang, Y.; Dumelin, C. E.; Scheuermann, J.; Neri, D. Design and synthesis of a novel DNA-encoded chemical library using Diels-Alder cycloadditions. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5926−5931.
(27) 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.; Israel, D. I.; Jiang, J.; Kavarana, M. J.; Kelley, M. S.; Kollmann, C. S.; Li, F.; Lind, K.; Mataruse, S.; Medeiros, P. F.; Messer, J. A.; Myers, P.; O'Keefe, H.; Oliff, M. C.; Rise, C. E.; Satz, A. L.; Skinner, S. R.; Svendsen, J. L.; Tang, L.; van Vloten, K.; Wagner, R. W.; Yao, G.; Zhao, B.; Morgan, B. A. Design, synthesis and selection of DNA-encoded small-molecule libraries. *Nat. Chem. Biol.* **2009**, *5*, 647−654.
(28) Leimbacher, M.; Zhang, Y.; Mannocci, L.; Stravs, M.; Geppert, T.; Scheuermann, J.; Schneider, G.; Neri, D. Discovery of small-molecule interleukin-2 inhibitors from a DNA-encoded chemical library. *Chemistry* **2012**, *18*, 7729−7737.
(29) Wrenn, S. J.; Weisinger, R. M.; Halpin, D. R.; Harbury, P. B. Synthetic ligands discovered by in vitro selection. *J. Am. Chem. Soc.* **2007**, *129*, 13137−13143.
(30) Deng, H.; O'Keefe, H.; Davie, C. P.; Lind, K. E.; Acharya, R. A.; Franklin, G. J.; Larkin, J.; Matico, R.; Neeb, M.; Thompson, M. M.; Lohr, T.; Gross, J. W.; Centrella, P. A.; O'Donovan, G. K.; Bedard, K. L.; van Vloten, K.; Mataruse, S.; Skinner, S. R.; Belyanskaya, S. L.; Carpenter, T. Y.; Shearer, T. W.; Clark, M. A.; Cuozzo, J. W.; Arico-Muendel, C. C.; Morgan, B. A. Discovery of highly potent and selective small molecule ADAMTS-5 inhibitors that inhibit human cartilage degradation via encoded library technology (ELT). *J. Med. Chem.* **2012**, *55*, 7061−7079.
(31) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **2001**, *46*, 3−26.
(32) Scheuermann, J.; Dumelin, C. E.; Melkko, S.; Zhang, Y.; Mannocci, L.; Jaggi, M.; Sobek, J.; Neri, D. DNA-encoded chemical libraries for the discovery of MMP-3 inhibitors. *Bioconjugate Chem.* **2008**, *19*, 778−785.
(33) Dumelin, C. E.; Trussel, S.; Buller, F.; Trachsel, E.; Bootz, F.; Zhang, Y.; Mannocci, L.; Beck, S. C.; Drumea-Mirancea, M.; Seeliger, M. W.; Baltes, C.; Muggler, T.; Kranz, F.; Rudin, M.; Melkko, S.; Scheuermann, J.; Neri, D. A portable albumin binder from a DNA-encoded chemical library. *Angew. Chem., Int. Ed.* **2008**, *47*, 3196−3201.
(34) Mannocci, L.; Melkko, S.; Buller, F.; Molnar, I.; Bianke, J. P.; Dumelin, C. E.; Scheuermann, J.; Neri, D. Isolation of potent and specific trypsin inhibitors from a DNA-encoded chemical library. *Bioconjugate Chem.* **2010**, *21*, 1836−1841.
(35) Buller, F.; Steiner, M.; Frey, K.; Mircsof, D.; Scheuermann, J.; Kalisch, M.; Buhlmann, P.; Supuran, C. T.; Neri, D. Selection of carbonic anhydrase IX inhibitors from one million DNA-encoded compounds. *ACS Chem. Biol.* **2011**, *6*, 336−344.

(36) Disch, J. S.; Evindar, G.; Chiu, C. H.; Blum, C. A.; Dai, H.; Jin, L.; Schuman, E.; Lind, K. E.; Belyanskaya, S. L.; Deng, J.; Coppo, F.; Aquilani, L.; Graybill, T. L.; Cuozzo, J. W.; Lavu, S.; Mao, C.; Vlasuk, G. P.; Perni, R. B. Discovery of thieno[3,2-d]pyrimidine-6-carboxamides as potent inhibitors of SIRT1, SIRT2, and SIRT3. *J. Med. Chem.* **2013**, *56*, 3666−3679.
(37) Tse, B. N.; Snyder, T. M.; Shen, Y.; Liu, D. R. Translation of DNA into a library of 13,000 synthetic small-molecule macrocycles suitable for in vitro selection. *J. Am. Chem. Soc.* **2008**, *130*, 15611−15626.
(38) Kleiner, R. E.; Dumelin, C. E.; Tiu, G. C.; Sakurai, K.; Liu, D. R. In vitro selection of a DNA-templated small-molecule library reveals a class of macrocyclic kinase inhibitors. *J. Am. Chem. Soc.* **2010**, *132*, 11779−11791.
(39) Kawakami, T.; Ohta, A.; Ohuchi, M.; Ashigai, H.; Murakami, H.; Suga, H. Diverse backbone-cyclized peptides via codon reprogramming. *Nat. Chem. Biol.* **2009**, *5*, 888−890.
(40) Melkko, S.; Mannocci, L.; Dumelin, C. E.; Villa, A.; Sommavilla, R.; Zhang, Y.; Grütter, M.; Keller, N.; Jermutus, L.; Jackson, R. H.; Scheuermann, J.; Neri, D. Isolation of a Small-Molecule Inhibitor of the Antiapoptotic Protein Bcl-xL from a DNA-Encoded Chemical Library. *ChemMedChem* **2010**, *5*, 584−590.
(41) Podolin, P. L.; Bolognese, B. J.; Foley, J. F.; Long, E., 3rd; Peck, B.; Umbrecht, S.; Zhang, X.; Zhu, P.; Schwartz, B.; Xie, W.; Quinn, C.; Qi, H.; Sweitzer, S.; Chen, S.; Galop, M.; Ding, Y.; Belyanskaya, S. L.; Israel, D. I.; Morgan, B. A.; Behm, D. J.; Marino, J. P., Jr.; Kurali, E.; Barnette, M. S.; Mayer, R. J.; Booth-Genthe, C. L.; Callahan, J. F. In vitro and in vivo characterization of a novel soluble epoxide hydrolase inhibitor. *Prostaglandins Other Lipid Mediatiators* **2013**, *104−105*, 25−31.
(42) Thalji, R. K.; McAtee, J. J.; Belyanskaya, S.; Brandt, M.; Brown, G. D.; Costell, M. H.; Ding, Y.; Dodson, J. W.; Eisennagel, S. H.; Fries, R. E.; Gross, J. W.; Harpel, M. R.; Holt, D. A.; Israel, D. I.; Jolivette, L. J.; Krosky, D.; Li, H.; Lu, Q.; Mandichak, T.; Roethke, T.; Schnackenberg, C. G.; Schwartz, B.; Shewchuk, L. M.; Xie, W.; Behm, D. J.; Douglas, S. A.; Shaw, A. L.; Marino, J. P., Jr. Discovery of 1-(1,3,5-triazin-2-yl)piperidine-4-carboxamides as inhibitors of soluble epoxide hydrolase. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 3584−3588.
(43) Buller, F.; Zhang, Y.; Scheuermann, J.; Schafer, J.; Buhlmann, P.; Neri, D. Discovery of TNF inhibitors from a DNA-encoded chemical library based on diels-alder cycloaddition. *Chem. Biol.* **2009**, *16*, 1075−1086.
(44) Georghiou, G.; Kleiner, R. E.; Pulkoski-Gross, M.; Liu, D. R.; Seeliger, M. A. Highly specific, bisubstrate-competitive Src inhibitors from DNA-templated macrocycles. *Nat. Chem. Biol.* **2012**, *8*, 366−374.
(45) Debaene, F.; Da Silva, J. A.; Pianowski, Z.; Duran, F. J.; Winssinger, N. Expanding the scope of PNA-encoded libraries: Divergent synthesis of libraries targeting cysteine, serine and metallo-proteases as well as tyrosine phosphatases. *Tetrahedron* **2007**, *63*, 6577−6568.
(46) Svensen, N.; Diaz-Mochon, J. J.; Bradley, M. Decoding a PNA encoded peptide library by PCR: The discovery of new cell surface receptor ligands. *Chem. Biol.* **2011**, *18*, 1284−1289.
(47) Inglese, J.; Shamu, C. E.; Guy, R. K. Reporting data from high-throughput screening of small-molecule libraries. *Nat. Chem. Biol.* **2007**, *3*, 438−441.
(48) Muckenschnabel, I.; Falchetto, R.; Mayr, L. M.; Filipuzzi, I. SpeedScreen: Label-free liquid chromatography-mass spectrometry-based high-throughput screening for the discovery of orphan protein ligands. *Anal. Biochem.* **2004**, *324*, 241−249.
(49) Buller, F.; Steiner, M.; Scheuermann, J.; Mannocci, L.; Nissen, I.; Kohler, M.; Beisel, C.; Neri, D. High-throughput sequencing for the identification of binding molecules from DNA-encoded chemical libraries. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 4188−4192.
(50) Mardis, E. R. Next-generation sequencing platforms. *Annu. Rev. Anal. Chem.* **2013**, *6*, 287−303.
(51) Dumelin, C. E.; Scheuermann, J.; Melkko, S.; Neri, D. Selection of streptavidin binders from a DNA-encoded chemical library. *Bioconjugate Chem.* **2006**, *17*, 366−370.
(52) Reymond, J. L.; Awale, M. Exploring chemical space for drug discovery using the chemical universe database. *ACS Chem. Neurosci.* **2012**, *3*, 649−657.
(53) Keseru, G. M.; Makara, G. M. The influence of lead discovery strategies on the properties of drug candidates. *Nat. Rev. Drug Discovery* **2009**, *8*, 203−212.
(54) Murray, C. W.; Verdonk, M. L.; Rees, D. C. Experiences in fragment-based drug discovery. *Trends Pharmacol. Sci.* **2012**, *33*, 224−232.
(55) Pellecchia, M. Fragment-based drug discovery takes a virtual turn. *Nat. Chem. Biol.* **2009**, *5*, 274−275.
(56) Melkko, S.; Zhang, Y.; Dumelin, C. E.; Scheuermann, J.; Neri, D. Isolation of high-affinity trypsin inhibitors from a DNA-encoded chemical library. *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 4671−4674.