DNA-Encoded Libraries: Towards Harnessing their Full Power with Darwinian Evolution

Dockerill, Millicent; Winssinger, Nicolas

How to cite

DOCKERILL, Millicent, WINSSINGER, Nicolas. DNA-Encoded Libraries: Towards Harnessing their Full Power with Darwinian Evolution. In: Angewandte Chemie, 2022, vol. 62, n° 9. doi: 10.1002/anie.202215542

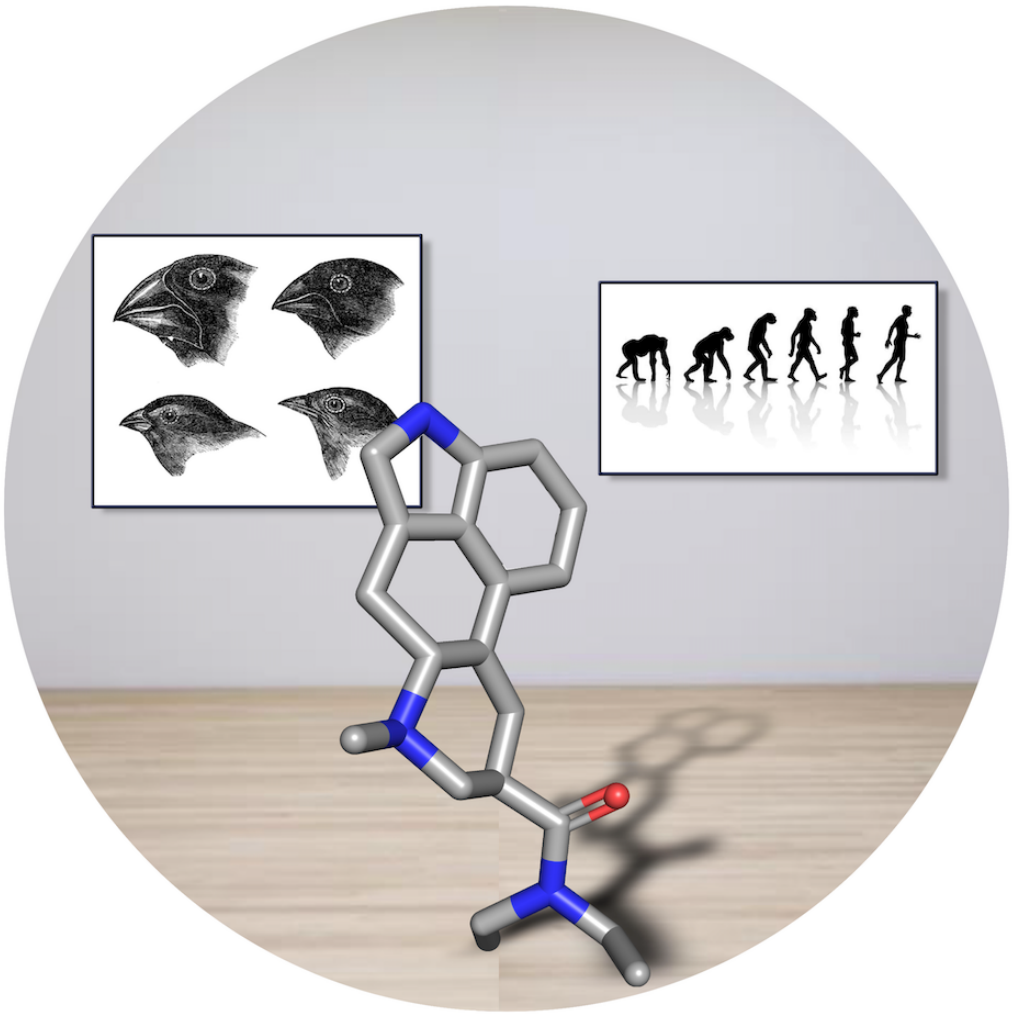
This publication URL: https://archive-ouverte.unige.ch/unige:168322
Publication DOI: 10.1002/anie.202215542

DNA-Encoded Libraries: Towards Harnessing their Full Power with Darwinian Evolution

Millicent Dockerill,[a] and Nicolas Winssinger*[a]

Department of Organic Chemistry, NCCR Chemical Biology, Faculty of Sciences, University of Geneva - 1211 Geneva (Switzerland)

Email: nicolas.winssinger@unige.ch



Millicent Dockerill received her master's degree in chemistry from the University of Edinburgh. She spent her final year research project in the medicinal chemistry department at AstraZeneca in Cambridge. In 2020, she joined the group of Professor Winssinger at the University of Geneva to pursue her PhD. Her interests lie in broadening the scope of DNA encoded libraries to bridge the gap between small drug-like molecules and biomacromolecules.

Nicolas Winssinger is professor of chemistry at the university of Geneva. His research interests lie at the crossroads of molecular sciences and life science. His lab uses nucleic acid as a supramolecular tag to encode molecules, program spatial organization of ligands and reactions. A long-term objective is to extend these principles towards complex systems that can emulate some of the fundamental features of living systems. His lab has been actively involved in DNA encoded libraries for the past 20 years. He received his PhD under the supervision of Prof. Nicolaou and carried out his postdoctoral training with Prof. Schultz.

Abstract: DNA-encoded library (DEL) technologies are transforming the drug discovery process, enabling the identification of ligands at unprecedented speed and scale. DEL makes use of libraries that are orders of magnitude larger than traditional high throughput screens. While a DNA tag alludes to a genotype-phenotype connection that is exploitable for molecular evolution, most of the work in the field is performed with libraries where the tag serves as an amplifiable bar code but does not allow 'translation' into the synthetic product it is linked to. In this review, we cover technologies that enable the 'translation' of the genetic tag into synthetic molecules, both biochemically and chemically, and explore how it can be used to harness Darwinian evolutionary pressure.

1. Introduction

The discovery of novel enzyme inhibitors, receptor agonists/antagonists, or simply binders is at the core of pharmaceutical innovations and central to biomedical research. Historically, natural products and chemical synthesis coupled to high throughput screening technologies (HTS) have been at the forefront of this process. The development of DNA-encoding technologies has brought a paradigm shift in this workflow. DNA-encoded library (DEL) technologies greatly facilitate the screening of synthetic molecules. The DNA tag essentially links the individual phenotype of synthetic molecules to an amplifiable barcode. It enables the identification of the fittest ligands in libraries of 10^6-10^9 within days to weeks, a transformative achievement compared to traditional HTS. The screening process is dramatically facilitated by the fact that affinity selections can be used, selecting for the tightest binder in the ensemble of the library rather than the discrete assays of HTS. This parallels biochemical library technologies such as phage display, ribosomal display, or systematic evolution of ligands by exponential enrichment (SELEX). However, DEL-screens are mostly practiced with a dsDNA-tag that record the synthetic path of individual members, hence encode their structures, but cannot be translated in their synthesis as do biochemically DNA-encoded processes (e.g., phage display). While screening libraries of 10^6-10^9 synthetic molecules is very impressive relative to what could be done a decade ago, it only scratches the surface of the theoretical diversity space, estimated at 10^{60} of drug like small molecules.[1-2] If one considers macromolecules, the diversity space is so large that a meaningful sampling could never be achieved with a single library. A library of all permutations of a 100 amino acid (very small protein), restraining it to the 20 canonical amino acids, would have 1.26 x 10^{130} permutations. There are not enough atoms in the universe to explore such a library. The breakthrough for macromolecules came from advances enabling evolution in a test tube using iterative cycles of selection, amplification, and diversification. This Darwinian evolution is facilitated by the fact that the DNA encodes the synthesis of the protein making it possible to amplify and translate the product of a first round of selection to reiterate the process. The power of Darwinian evolution (selection, amplification, diversification) comes from the fact that the fittest member in a screen needn't be present in the starting library but can emerge through the iterative process. It is thus not imperative to start with excessively large libraries provided the diversification mechanism can yield members that were not present in the starting library. While the world of biomolecules and drug-like small molecules are the remit of different disciplines that do not overlap experimentally, their goals converge in drug discovery and other biomedical applications, as well as catalysis. DNA-encoded technologies ought to be where these two fields also converge experimentally to bring the power of Darwinian evolution to the small(er) synthetic molecule realm and extend properties and function of macromolecules to a broader arsenal of chemical functionalities. This review covers different DNA-encoded library technologies, including both biochemical and chemical methods as well as the screening approach.

2.1. Biochemical Libraries

Phage display, pioneered by Smith in the 1980s makes use of bacteriophage with a well-defined coat protein that can be genetically modified to include the fusion peptide sequences to be displayed on the surface of the virion, thus providing a simple connection between the phenotype of the peptide library and its genotype. Linking these two properties enables *in vitro* selection since the encoding region can be amplified, enabling multiple selection cycles. It was rapidly shown that this method could be used to select for the fittest binder by affinity purification (biopanning).[13-14] This technology was embraced for the purpose of therapeutic antibodies, displaying the variable domain of a monoclonal antibody (mAb).[15-16] Attesting to the success of this technology, adalimumab (Humira®), the first therapeutic antibody discovered using this technology, was granted approval in 2002.[17] The impact of this technology was recognized by a Nobel prize awarded to Smith and Winter in 2018 (along with Arnold for directed evolution). However, in the realm of small molecule drugs, unmodified peptides typically do not have acceptable pharmacological properties. This limitation

was overcome by the development of methodologies harnessing selective functionalization of cysteines within a peptide library in order to screen macrocycles which have much better pharmacological properties. Numerous phage-compatible cyclisation strategies have now been developed to create large diverse mono- or bicyclic macrocyclic libraries (Figure 1.A.).[18-21] The Heinis lab have recently leveraged this technology to discover potent macrocycles that inhibit activated coagulation factor XI (FXIa), a step towards safer anti-coagulants.[22] Other approaches have been developed to further the scope of phage display libraires. Bogyo and co-workers developed a method to dually cyclize and incorporate a covalent warhead into their peptide library to develop selective irreversible inhibitors of the tobacco etch virus (TEV) protease and fluorophosphonate-binding hydrolases F (FphF).[23] Phage display is routinely used with library size of 10^{6-10} using multiple rounds of selection and amplification. The development of mechanisms for continuous evolution that include diversification is opening new horizons in the scope of these screens (*vide infra*, selection section). Other display platforms than phage have also been explored, such as *E. coli* display coupled to non-natural amino acid incorporation.[24]

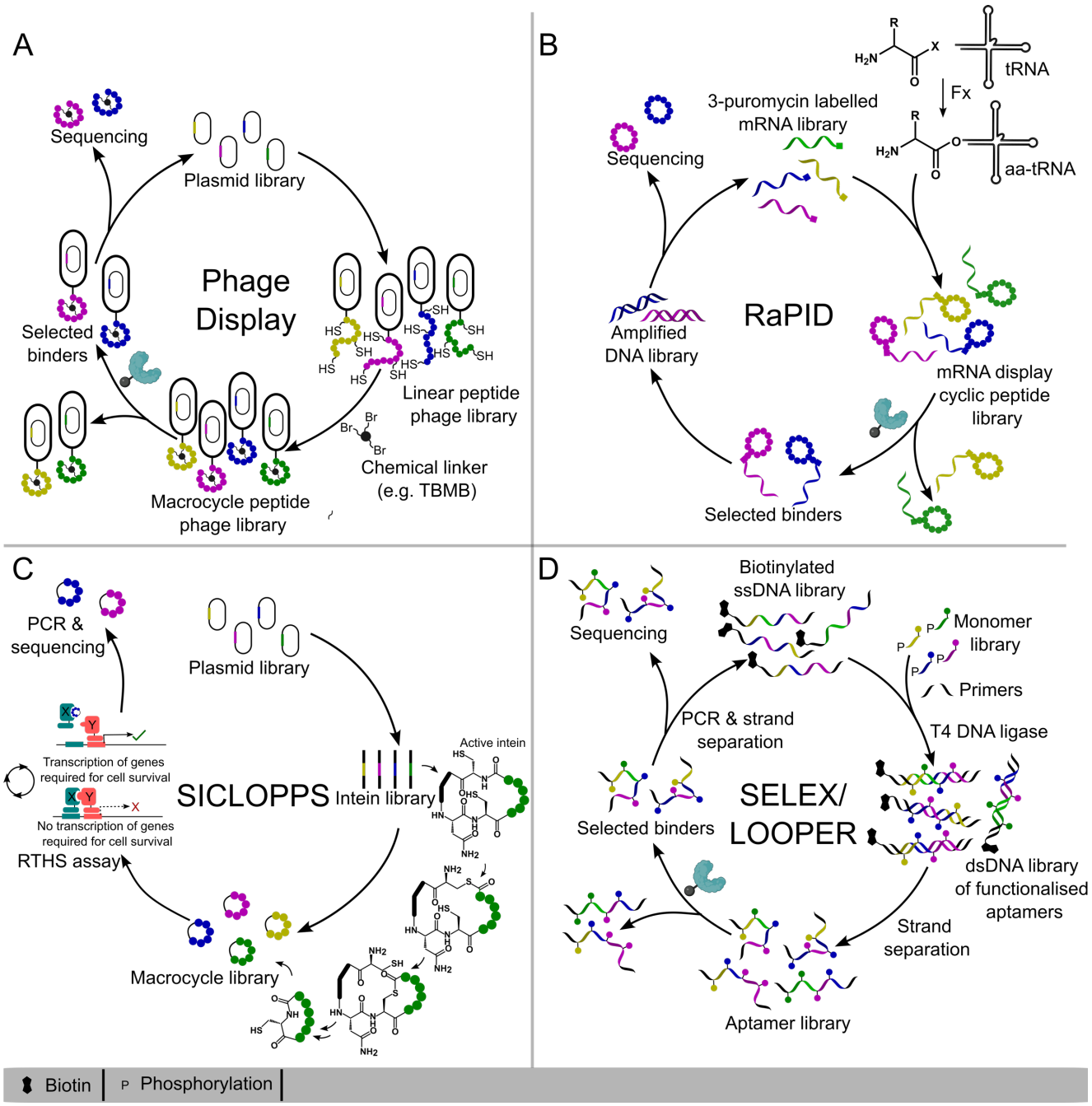


Figure 1. Biochemical Screening Technologies. (A) Phage display of cyclised macrocycles (B) mRNA display technology, RaPID (C) SICLOPPS screening of macrocycles (D) combination of LOOPER and SELEX for functionalised aptamer screening.

Ribosomal display is another technology that enables the direct translation of genetic information into peptides.[25-28] Since mRNA display capitalizes on nature's power to translate oligonucleotides into proteins, it was traditionally limited to canonical amino acids. In 2003, Suga developed the first flexizyme (Fx3),[29-30] a flexible tRNA acylation technology which enables the incorporation of non-natural amino acids on to tRNAs. Fx3 was followed by numerous generations of flexizymes which further broadened the scope of the non-proteinogenic amino acids that could be incorporated.[31] The reprogramming of the genetic code is exploited for Flexible *in vitro* Translation (FIT), since the non-standard acyl tRNAs, prepared with flexizyme, are used to produce peptides incorporating a wide range of amino acids (e.g. natural amino acids, *D*-amino acids, non-natural side chains, β-amino acids etc.). This technology has been used to screen large compound libraries in order to discover binders against specific targets, a process called Random non-standard Peptide Integrated Discovery (RaPID) (Figure 1.B).[32-33] Leveraging the ability to incorporate non-natural amino acids, it is often used to prepare and screen libraries of macrocycles with more desirable pharmacology over their linear counterparts. This technology has been commercialized (Peptidream) and embraced by a growing number of academic institutions. A recent example that illustrates the speed of execution is the identification of different cyclic peptides that bind to SARS-CoV-2 spike protein,[34] and the virus' main protease,[35] just months after the targets became known.

Cyclic peptides can also be generated by Split Intein-mediated Circular Ligation of Peptides and ProteinS (SICLOPPS). Benkovic and co-workers discovered that DnaE split intein from *Synechocystis sp.* PCC6803 could spontaneously form macrocycles *in vivo*.[36] The variable region of the peptide is encoded by a vector containing all the required components. The system can incorporate all 20 amino acids, with variable loop size, and simply requires cysteine or serine to perform transesterification and asparagine to release the cyclic peptide intracellularly (Figure 1.C.).[37] This technology, which starts with a simple library of plasmids, can be combined with a reverse two-hybrid system (RTHS) to screen large macrocyclic libraries *in cellulo*.[38-39] Whilst SICLOPPS is limited in the chemical diversity space (canonical amino acids, albeit *p*-benzoylphenylalanine was also used),[40] the library is screened directly in cell/bacteria/yeast,[41] providing an additional advantage, as protein binding *in vitro* does not necessarily translate to the desired phenotype in cells (*vide infra* for discussion on selection).

To screen large libraries of macrocycles, mRNA display, phage display and SICLOPPS libraries can undergo multiple rounds of selections in order to filter out false positives and enrich selected binders. SELEX also lends itself to this approach but delivers the discovery of oligonucleotides (aptamers) rather than peptides.[42-43] Oligonucleotide libraries have the added advantage that both genotype and phenotype are simultaneously represented and do not require a 'translation' step. Nucleic acids lend themselves to molecular evolution using a simple workflow: Start with a library of oligonucleotides, select for desirable properties, make more copies with some errors to introduce random change and repeat cycle. Joyce and coworkers elegantly leveraged this workflow to evolve different catalytic activities using nucleic acids.[44-46] However, whilst the peptide discovery techniques mentioned previously were once limited to the 20 natural amino acids, SELEX was traditionally limited to the 4 nucleobases. Methods have since been developed to encode and screen functionalized aptamers. A large range of artificial nucleobases and backbones have been developed enabling the diversification of the standard 'four letter alphabet'.[47-48] Hili *et al.* developed the Ligase-catalyzed OligO nucleotide PolymERisation (LOOPER) technology which combines SELEX with modified nucleotides to yield Highly Functionalized Nucleic Acid Polymers (HFNAPs) (Figure 1.D.).[49-50] Rather than swapping one natural nucleotide for a non-natural nucleotide, limiting the chemical modifications to 4, a 3 letter codon was used to encode the modifications enabling a wider range of diversity. Using T4 DNA ligase, the same group showed the polymerization of pentanucleotides derivatized with peptide octamers.[51] This technology was later used to screen a 10^{13}-member library which yielded a potent thrombin aptamer (K_D = 1.6 nM) incorporating 7 modifications.[52] A library of HFNAPs also yielded a PCSK-9 binder with a K_D of 3 nM.[53] A significant effort to expand aptamer's chemical space has focused on protein engineering in order to develop polymerases that recognize and amplify non-natural nucleic acid sequences. Chaput *et al*. synthesized chemically modified α-L-threofuranosyl uridine nucleoside triphosphates (tUTPs) and showed the efficient recognition of these modified bases by the engineered polymerase Kod-RSGA.[54] The same group also investigated alkyl phosphonate nucleic acids (phNAs) in which the canonical, negatively charged phosphodiester is replaced by an uncharged P-alkyl phosphonodiester backbone. These building blocks have also been used in DNA-templated synthesis and selection of phNA aptamers.[55] Non-enzymatic directed evolution of xeno nucleic acids (XNAs) has also been developed and will be discussed later in the review.

All these biochemical library syntheses leverage methods for 'translating' the DNA or RNA into the molecules of interest. It lends itself to reiterative selection/amplification cycles. In addition, error-prone PCR can be used to reintroduce diversity. This is frequently performed in SELEX experiments, thus truly taping in the full power of Darwinian evolution in its simplest embodiment. However, a limitation of the above-mentioned technologies is that they deliver specific biopolymers that don't share all the desirable pharmacological features of drug-like small molecules. The growing arsenal of bio-orthogonal chemistries has significantly pushed the boundaries of the chemical space that can be explored with these technologies,[21, 56-59] but other areas of the chemical diversity space such as diverse heterocycles and polycyclic frameworks known to be highly important in medicinal chemistry, remain inaccessible.

2.2. Chemical Libraries

Chemical libraries can be designed with a much wider range of functionalities based on drug-like properties. While not limited by the scope of enzymatic processes used in biochemical libraries, an important technical challenge is the restrictions imposed by DNA synthesis and none of the methods practiced today perform chemical DNA synthesis in conjunction with small molecule synthesis. Breakthrough in the synthesis of DEL came from five different technological approaches that overcome the limitations of chemical DNA synthesis: 1. DNA-templated synthesis (DTS);[6, 60] 2. PNA-encoded synthesis (hybridized to complementary DNA for decoding);[8, 61] 3. Directed sorting, routing DNA containing specific codons through a given path of split and mix synthesis;[9, 62] 4. Self-assembling libraries where the diversity arises from the combinatorial pairing of DNA-tagged fragments (a supramolecular approach);[7, 63] 5. Primer extension[64-65] or enzymatic ligation of dsDNA tags[10, 66] to encode split and mix synthesis. Technologically, the latter method (enzymatic tagging) is the most accessible and is the most practiced with several commercial services (HitGen, X-Chem) and commercial libraries available.[67] The scope of each of these methods has been extensively reviewed.[68-75] Here we will focus on technologies that enable the 'translation' of the DNA code into the synthetic molecule in order to amplify selected library members from a first round of selection, essentially recapitulating two important steps of Darwinian evolution: selection and amplification.

DTS,[60] first disclosed in 2001,[76] involves sequence specific synthesis of products unrelated to DNA by harnessing hybridization to bring the reactive partners together and modify the molecule attached to the DNA. This technology has since evolved to be compatible with a range of chemistries broadening its applications.[60] Implementing DTS for library preparation allows for iterative selection cycles and therefore increased enrichment of the selected binders. This was first showcased with a pilot library of macrocycles (Figure 2.A.), demonstrating an enrichment of a carbonic anhydrase binder after two selection cycles, recapitulating two essential steps towards evolution of synthetic products: selection and amplification.[6] The library size was then increased to 13,824 members leading to the discovery of a selective sub-μM inhibitor of Src kinase, which was further developed to yield macrocycles with single digit nanomolar IC_{50}s.[77-79] Improvement of this screening technology enabled the preparation of a 256,000 member library which was screened against insulin-degrading enzyme (IDE), yielding a novel macrocycle displaying an IC_{50} of 40 nM.[80] Library synthesis with DTS can be carried out in a single reaction vessel since each reactant is directed to the appropriate place through hybridization but requires laborious preparation of the reagent-DNA conjugates. Another important consideration in DTS is the choice of DNA sequences used for each reagent. Since it must balance duplex stability and hybridization specificity, the entire repertoire of sequence space cannot be used. DTS of synthetic compounds require single strand linear DNA which have potential drawbacks at the screening stage, including non-specific hybridization between library members and folding of the ssDNA tag. The 'Yoctoreactor' (yR) is conceptually related since the synthesis is achieved in a templated fashion but harmonizes the distance between reacting partner using a three-way junction.[81] The term comes from the estimated volume of the center of a DNA three-way junction (10^{-24} L). This technology involves three DNA strands (combinatorial assembly) that fold to bring the reactive functional groups in proximity, whilst keeping the encoding region distant from the reaction site (Figure 2.B.). Prior to screening, the synthetic compound is exposed by disassembly of the Yoctoreactor via transformation to dsDNA. The latter can be amplified and reintegrate the chemical moiety using 'rolling translation', enabling multiple cycles of selection. This rolling translation method involves successive enzymatic digestion, ligation with DNA-building blocks, reaction, purification, and refolding steps. This workflow was demonstrated to yield 150,000-fold enrichment of [Leu]enkephalin, an endogenous opioid peptide neurotransmitter, after two selection cycles against anti-ENK monoclonal antibody. This technology also has the added advantage that it does not require one DNA strand per compound, unlike the previous macrocyclic libraries mentioned, which can be laborious to prepare and design, however, the translation requires more steps. Similar assemblies have been published by Xiaoyu Li and co-workers.[82] Prior to this work, their lab also designed a DTS-based DEL that relies on a universal template thus also alleviating complex template preparation. It exploits the promiscuity of deoxy inosine, which can hybridize with all 4 nucleobases. Their method utilizes multiple hybridization, reaction, photocleavage cycles to produce a diverse library (Figure 2.C.).[83] They showed 94.7-fold enrichment of a known binder of carbonic anhydrase II protein after one selection cycle. However, iterative cycles were not exploited.

DTS is the chemical equivalent of many biological processes which enable the translation of DNA (genotype) into synthetic molecules with specific properties (phenotype), discussed previously. Non-enzymatic approaches involving templated-directed synthesis of XNAs have been developed to yield larger molecular weight species. The lab of David Liu has developed a translation, selection, amplification system for peptide nucleic acids (PNAs).[84] This approach was developed further to encode synthetic oligomers with PNAs, which would undergo sequence dependent ligation, followed by cleavage of the PNA to release a synthetic polymer encoded by the DNA template (which can be amplified to reiterate the cycle) (Figure 2.D.).[85] More recently, non-enzymatic templated synthesis of acyclic L-threoninol nucleic acid (L-αTNA), requiring N-cyan imidazole, has also been developed.[86] Whilst no selection cycles were described, this development could be a promising platform for the discovery of diverse XNA polymers via *in vitro* selection.

Concurrently to the development of DTS, the Harbury lab developed directed sorting of DNA for the preparation of DEL.[9] In this approach, a library of ssDNA sequences is split into different pools by hybridization to immobilized sequences corresponding to the anticodon of a given synthetic transformation (Figure 2E). After each step, the library is pooled, and a further synthetic cycle can be performed. Thus, the phenotype of the molecule is encoded in its synthetic path through this DNA sorting. The method is thus compatible with cycles of selection and amplification since the DNA tags obtained after a first round of selection can be subjected to the same synthetic scheme.[62, 87-88] One issue limiting molecular evolution using this technology is the fact that the codons used represent a small fraction of all possible genetic

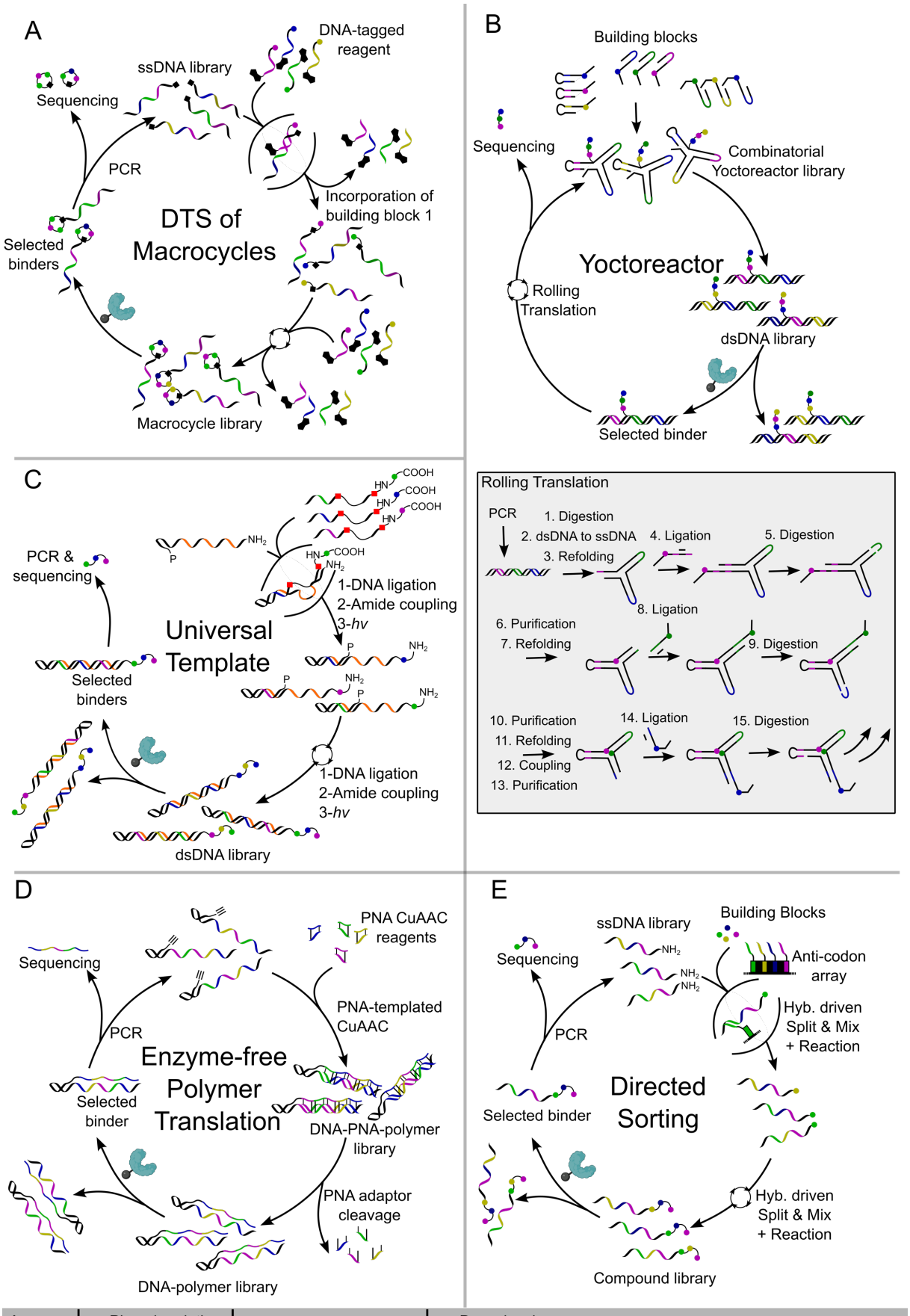


Figure 2. Chemical screening technologies (A) DTS of macrocycles (B) Yoctoreactor with regeneration of library species by rolling translation (C) Universal template synthesis (D) Enzyme-free polymer translation (E) Directed sorting.

permutations and, mutations in the genetic information would lead to non-coding sequences. This was later resolved by the use of an outsized genetic code (i.e. the same fragments are encoded by multiple codons).[89] This outsized genetic code was utilized to perform 4 iterative cycles with a $2x10^{10}$ member DNA library, encoding 88,434 peptides. Harbury and coworkers showed a 175,000-fold median enrichment for the DNA sequences that encoded the highest- ranking peptide when screened for fitness against protein kinase A (PKA) substrate. They demonstrated clear progression of PKA substrate motifs after three rounds. DELs prepared with directed sorting using an outsized genetic code have all the features necessary for molecular evolution. However, this methodology has thus far only been demonstrated with peptidic libraries. Extension to drug-like synthetic products would be a major milestone in molecular evolution.

2.3. Self-assembled libraries

Self-assembled DEL libraries are formed by the assembly of fragments through hybridization rather than covalent linkers. The hybridization can be thermodynamically stable, static fragment pairs, or dynamic with exchange of the fragment combinations. An attractive feature of this supramolecular approach to DEL is that numerically large libraries can be obtained by pairing smaller and better characterized libraries. The first example of fragment pairing was reported by Neri and co-workers and focused on the affinity maturation of a pharmacophore by pairing it with a library of diverse fragments

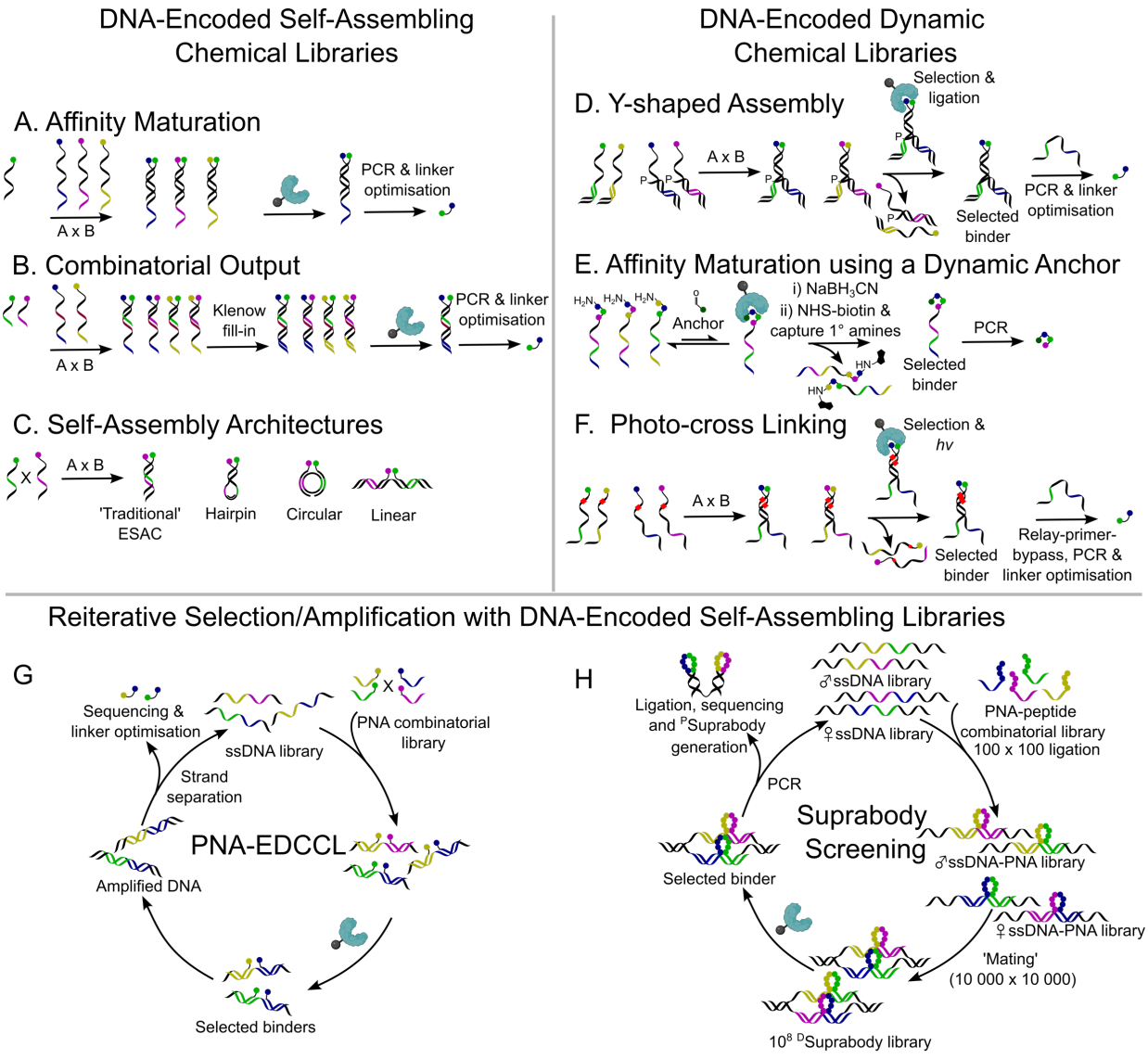


Figure 3. Self-assembled libraries. Different formats of self-assembly with static (A-C) or dynamic properties (D-F) and their application to reiterative selection/amplification (G-H).

(Figure 3.A.).[7] In this first example, the library size was not amplified by the pairing but the technology rapidly progressed to pair libraries of fragments affording a combinatorial output.[90] Methods were developed to encode this dual-output, notably via Klenow fill-in reaction (Figure 3.B.).[91-93] These works were performed with libraries which do not dynamically shuffle. Similar techniques have been applied to DNA-encoded dynamic libraries,[94] in order to distinguish the relationship between fragments. For this type of libraries, the protein of interest (POI) shifts the dynamic equilibrium generating the best binder transiently. A 'freezing' step is required to read out the best combinations.[95-96] As was first demonstrated with dynamic combinatorial libraries[97-98] (not DNA-encoded), a very attractive feature is the amplification of the fittest member through this equilibrium shift. The fittest member in a theoretical dynamic library of 100 x 100 is present at 0.01% but, the presence of the target would shift the equilibrium to 1% (assuming none of the other combinations are binders in this hypothetical library). Zhang *et al.* developed a Y-shape DNA assembly which enables the production of a ssDNA encoding both fragments only when the POI is present (Figure 3.D.).[99] This method combines combinatorial output with an additional enrichment step. Imine reduction has also been used to stop the reversible reaction between an amine DEL library and an aldehyde-anchor in order to screen for fragment affinity maturation, without post-screen linker optimization (Figure 3.E.).[100] Other methods include photo-cross linking the ssDNA of each pharmacophore using *p*-stilbazoles (B-base) (Figure 3.F.).[101] Photo-chemistry has also been used to interconvert between dsDNA and ssDNA encoded libraires using a reversible covalent headpiece (RCHP) with 3-cyanovinylcarbazole (^{CNV}K).[102-103] This method combines the ease of dsDNA library synthesis with the versatility of ssDNA, as well as the possibility of re-incorporating a different 2nd strand. Diversifying the scaffolds of self-assembled DELs enables different applications but it can also yield novel hits due to the different conformation of individual pharmacophores. The comparison of self-assembly architectures such as hetero-dimer, hairpin, circular and linear displays has been well studied (Figure 3.C),[92] with only the latter being used for iterative cycles of selection/ amplification to date. Attesting to the power of self-assembled libraries, isoform-selective fragments for therapeutic targets of interest were identified using a DNA-encoded dynamic library.[104]

The assembly of the library by hybridization to a template (Figure 3.C, linear architecture) offers the unique opportunity to amplify the assembly instructions following a selection in order to reassemble the library with fragments emerging from a first selection round (Figure 3.G.).[105-106] This approach was demonstrated with selections against carbonic anhydrase and lectins. In this case, the synthetic molecules were encoded with PNA, which is compatible with standard solid phase peptide synthesis (SPPS) and facilitates traditional split and mix approach. The architecture makes

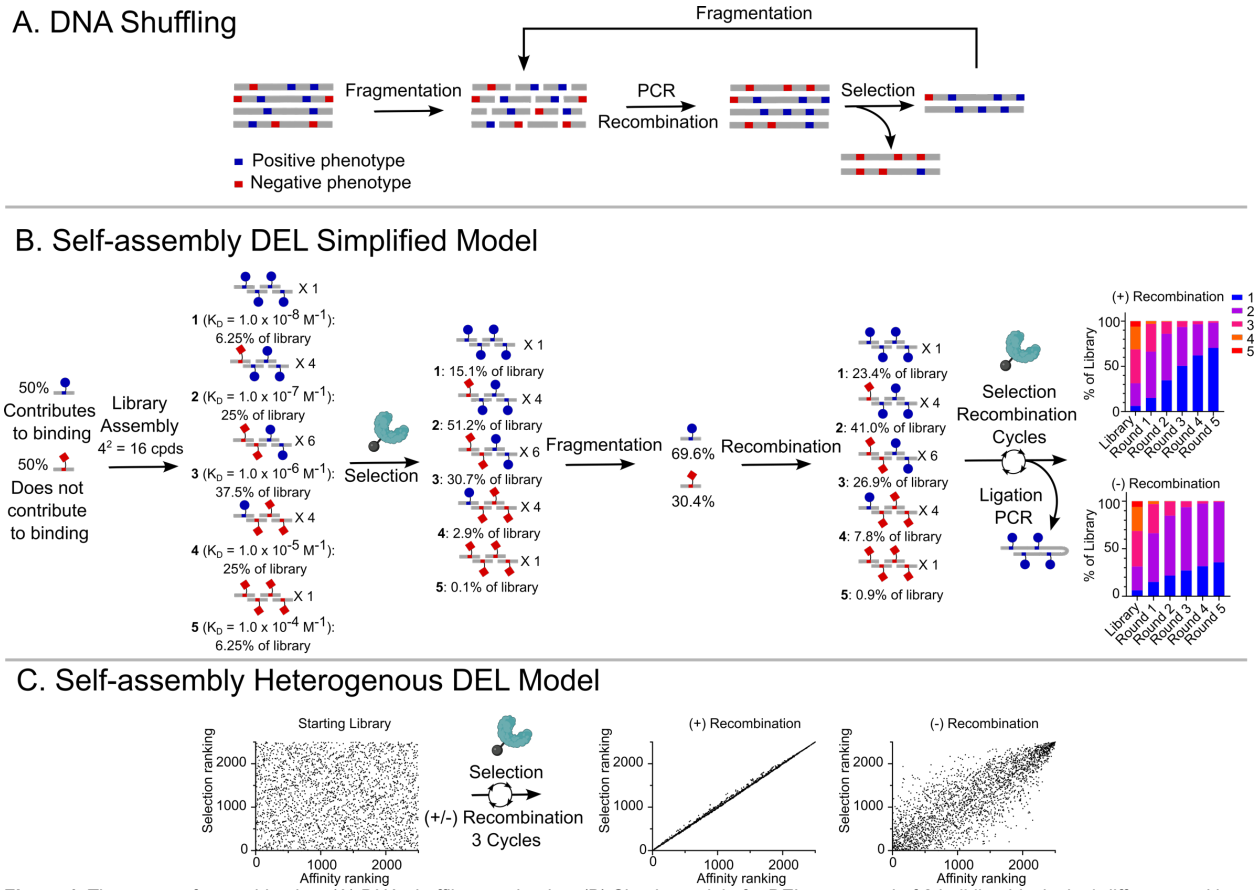


Figure 4. The power of recombination. (A) DNA shuffling mechanism (B) Simple model of a DEL composed of 2 building blocks in 4 different positions, yielding 16 different compounds which can be screened and recombined, gradually increasing the enrichment of the best binder (C) Modelling performed by Vummidi and Farrera-Soler *et al.* of a heterogenous library to display the selection ranking versus affinity ranking with and without recombination.

use of DNA templates that pair two PNA-encoded fragments, which can be amplified following a selection with the information of the fittest pair combination. Reassembly of the library is facilitated by the transient immobilization of the amplified DNA on a streptavidin resin and treatment with an excess of the PNA-conjugates to capture the desired synthetic product. This strategy offers fast reiterative selection / amplification cycles since it does not require chemical synthesis.

The chemical and self-assembly technologies mentioned previously enable large library preparation, efficient screening and decoding of molecules as well as iterative screening cycles to enrich the best binders, however none of them enable the generation of molecular diversity which is a key step of Darwinian evolution. Biological technologies enable this step by using error-prone PCR causing mutations which are then translated into the phenotype. However, this would not be functional with the previously discussed chemical and self-assembly technologies as it would result in non-coding sequences. Another important component of evolution is lineage crossing (sexual reproduction) or DNA cross-over, facilitating the emergence of combined mutations relative to random mutagenesis. At the molecular level, this can be recapitulated by DNA shuffling (Figure 4.A.).[107-108] This technique, also known as 'sexual PCR',[109] inspired the development of self-assembled library which involves the 'mating' of multiple DNA libraries (Figure 3.H).[110] This encoded supramolecular assembly starts as two ssDNA libraries (♂ and ♀) which have complementary primer regions. They each encode for a 10 000-compound library which can be prepared by DNA-templated ligation of 100 x 100 PNA-encoded fragments, making such a library readily accessible. The two templated libraries (ssDNA) are paired by hybridization to form a 10^8-member library. This architecture was used to generate peptidic loops akin to the hypervariable region of antibodies that encode the antigen binding site. There are a number of examples demonstrating that the conformation of peptides oligomers can be controlled by hybridization of flanking oligonucleotide rather a covalent bond of macrocycles.[111-114] These assemblies, termed DNA Suprabody (^{D}Suprabody) can be screened against a target of interest to select the best binders. The individual strands are then PCR amplified, re-hybridized with the PNA-peptide building blocks and reassembled into ^{D}Suprabodies. It is this recombination step which parallels DNA shuffling as enriched binders randomly reassemble introducing novel combinations.

The ^{D}Suprabody library can undergo multiple evolutionary cycles, followed by a simple ligation between the two encoding strands prior to PCR and sequencing in order to decode the best combination. The hits can be transformed into a PNA Suprabody (^{P}Suprabody), which mimics the architecture of the ^{D}Suprabody, but enables easy preparation (in line SPPS), purification and hit validation. This method was validated by screens against both streptavidin and PD-L1. To really understand the true power of recombination, mathematical modelling, based on the simulations described by Satz for traditional DEL,[115] was performed factoring in 'shuffling' after each selection cycle.[110] To exemplify the benefits, we reproduced the modeling with a self-assembled DEL composed of 2 building blocks (blue and red) in four different positions (4^2=16 compounds) (Figure 4.B.). We assume the blue fragment contributes to binding to a protein of interest, whilst the red does not. For this simplified simulation, it is assumed that having all four binding moieties leads to a K_D of 10 nM, and each time a binding functionality is lost, the affinity is decreased by one log unit. Assuming all compounds are equally present in the starting library (6.25% each), the best binder (**1**) would represent 6.25%, good binders (**2**) 25%, moderate binders (**3**) 37.5%, bad binders (**4**) 25% and non-binder (**5**) 6.25%. After the first round of selection, the best binder and good binders are enriched compared to the weaker compounds. The compounds then undergo a shuffling step where the building blocks are redistributed between all possible combinations, as would be in DNA shuffling, simulating the random hybridization of both templates. This increases the concentration of the best binder at the expense of lesser binder since there are numerically more combinations of moderate binders than the best binder. As shown in Figure 4.B, the convergence to the fittest binder progresses more rapidly with recombination than without. This is further accentuated in larger libraries. Mathematical simulations were performed using large libraries (10^8) where 1% of the library is assigned an affinity ranging from low nM to high nM, assuming a standard deviation of concentration in the initial library. Plot of the selection ranking vs affinity ranking of the top 2,500 binders clearly show the benefit of selections including recombination compared to without recombination (Figure 4.C.). As selection rounds are performed, the best binders should correspond to the highest-ranking compound. This linear correlation was significantly improved when recombination was used (R^2 = 0.999 vs 0.696).

3. Selection Approaches

Traditionally, affinity selection is utilized to screen for the best binders within a DEL library. Whilst this method is efficient, quick and suitable to a range of proteins obviating the need of functional assay development, the pertinence of the hit may be compromised by the artificial setting of the affinity selection.[116] First, the hit may not interact with a biologically relevant active site on the target, limiting its biological activity (unless used for PROTAC development). Secondly, immobilization of the POI on a solid support does not represent the protein's native environment which can result in poor hit validation when tested in a more complex assay. Developing novel and more elaborate screening technologies would enable the screening of more complex POIs, increase the hit quality, and could potentially be used to implement Darwinian evolution directly in the screening step.

Liu and coworkers utilized interaction-dependent PCR (IDPCR) to selectively amplify the sequences of molecules bound to a POI.[117] This was first done by conjugating the target to ssDNA via NHS chemistry which would form a duplex

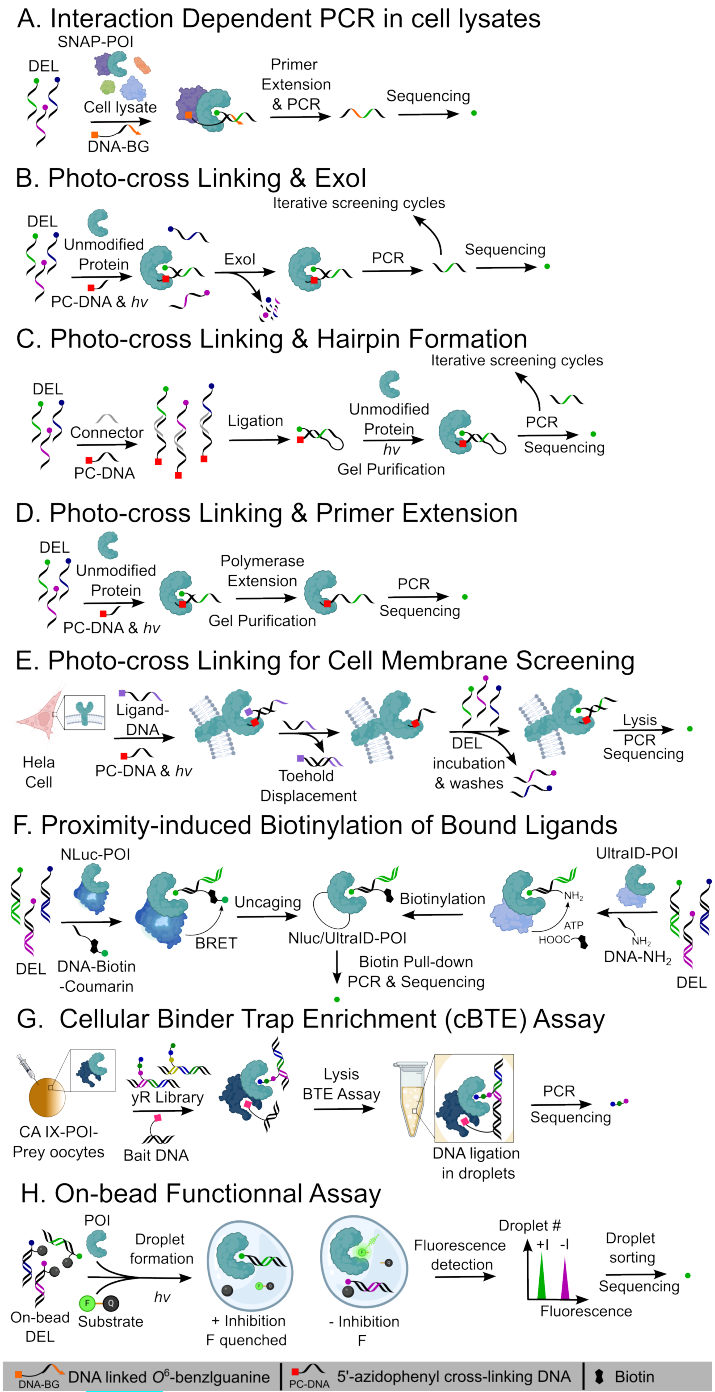


Figure 5. Selection approaches for chemical DELs (A-H)

with the library members' DNA only if encoding a good enough binder. The 'hairpin' is then extended, and PCR amplified to yield hit sequences. This technology was further developed to screen DELs against unpurified proteins in cell lysates.[118] Two novel techniques were developed to anchor the oligonucleotide tag; (i) covalently by expressing the target protein with SNAP, a self-labeling protein that reacts O^6-benzylguanine-DNA (DNA-BG) (Figure 5.A.) or (ii) non-covalently via a DNA-tagged antibody which either recognizes the target protein (this may hinder the binding site) or a His_6 epitope tag. The method was validated with a range of model systems (e.g., streptavidin, carbonic anhydrase II) and showed good enrichment of specific binders. The power of this technology is that the protein is found in a more native environment reflecting the complex mixture of partners present (and protein assemblies). To verify this function, rapamycin, which moderately binds to FRB (K_D 26 µM) but becomes potent in the presence of FKBP (K_D 12 nM), was conjugated to a ssDNA and screened in cell lysates containing overexpressed SNAP-FRB or overexpressed SNAP-FRB and FKBP. A 100-fold enrichment of the DNA encoding rapamycin was achieved when both proteins were overexpressed, versus 10-fold enrichment for SNAP-FRB alone. This result clearly demonstrates the power of screening in complex mixtures but requires a pre-screening modification of the target which may hinder ligand binding. Photo-chemistry has been widely utilized to screen against unmodified, non-immobilized protein targets,[119-122] as well as membrane-associated proteins.[123-124] Li and co-workers have extensively worked with photo crosslinking group (aryl azide, benzophenone, diazirene) to identify the best binders in DEL selections.[125] Their first design utilized a short 8-nt ssDNA containing the photo-cross linking group (PC-DNA) which hybridizes with a primer binding site (PBS) of the DEL, close to the binding moiety (Figure 5.B.). Upon protein binding and short light irradiation, the binders and their PC-DNA strand become covalently linked to the target. Upon addition of Exonuclease I (ExoI) the binders are protected from degradation and can be decoded or amplified to undergo another iterative cycle.[120] This technology was further developed into a 'ligate-cross-link-purify'[121] workflow, which covalently links the PC-DNA to the DEL via hairpin ligation, resulting in a fully covalent complex upon protein binding and light irradiation (Figure 5.C.). This complex can therefore be purified by gel, removing the risk of over-digestion by ExoI. Both methods utilized a 4,800-member library synthesized by DTS of macrocycles previously discussed,[77] applicable to multiple rounds of selection (although not performed). In another variant of the technology, the same group used PC-DNA to covalently react with the target and then undergo polymerase extension to copy the binder sequence information, which can be readout after gel purification (Figure 5.D.).[122] Importantly, these technologies were implemented to screen against membrane-associated proteins on live cells.[123] Screening of cell membrane proteins is highly desirable for drug discovery but trickier as the proteins are not stable outside of the membrane. This has been overcome in some instances by addition of detergents,[126] or using nano disc technology.[127] Screening targets directly on the cell membrane is difficult due to low concentration of protein and the requirement for

selective hit detection. Exploiting their DNA proximity affinity labelling (DPAL) technology,[119, 128-130] Li and co-workers were able to selectively tag the POI on the cell membrane with a short DNA strand via photo crosslinking (this required a known binder or antibody) (Figure 5.E.). This short strand is a 'homing beacon' for the DEL to overcome the issues mentioned previously. The DEL binds to the POI-DNA complex but only when the ligand has affinity for the target is the complex stable enough to resist washing steps. The binders undergo elution by heat denaturation, centrifugation, and PCR amplification of the supernatant to decode sequences. The potential of this technology was demonstrated by screening a 30.42 million-member DEL against carbonic anhydrase 12 (CA-12), the folate receptor (FR) and the epidermal growth factor receptor (EGFR).

In parallel, Krusemark and co-workers also developed a method to screen DELs against cell surface or cytosolic targets.[131] The method involved the expression of a fusion-POI (e.g. SNAP-DOR (δ opioid receptor)) to pull-down the POI after cell lysis. It also required the incorporation of a cyclic cell penetrating peptide (cCCP) tag onto the DNA to enable cellular uptake as well as a DNA linked sulfonyl fluoride cross linker to covalently attach the binders to the protein.[132] More recently, Krusemark and coworkers used enzyme-proximity labelling to enrich binders from DELs. Two methods were developed to selectively uncage or incorporate a biotin onto the protein bound ligands: (1) Photo-uncaging of a coumarin-biotin moiety via Nluc (fused to POI) induced BRET; (2) proximity induced biotinylation of NH_2-DNA by an engineered biotin ligase (UltraID) fused to the POI (Figure 5.F).[133] In addition they validated the screening on live cells by performing a test selection against UltraID-DOR.

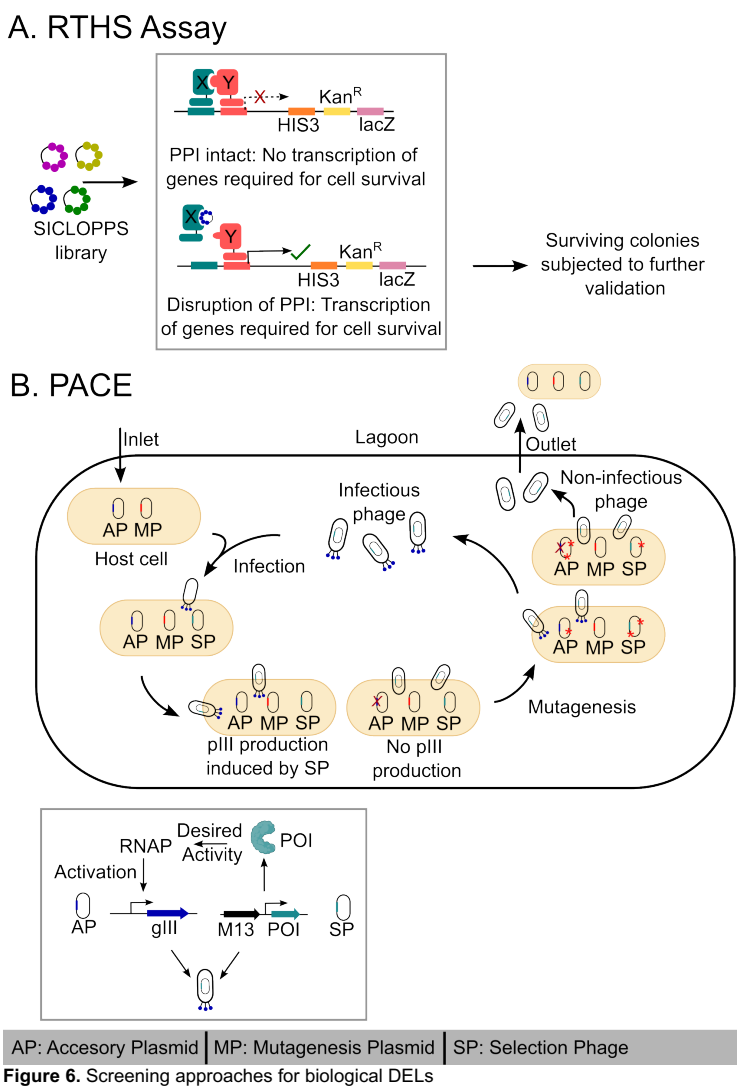


Figure 6. Screening approaches for biological DELs

Yet another approach to screen DELs in living cells was developed by Vipergen using libraries prepared using the Yoctoreactor approach.[134-136] This was made possible using a binder trap enrichment (BTE) assay (Figure 5.G). In this case, the protein of interest is fused to the catalytic domain of carbonic anhydrase IX (prey) such that it will bind to 'bait' dsDNA. The method relies on isolation of individual complex in droplets of water-oil emulsion which compartmentalizes and traps ligand-DNA with a protein-bait-DNA to catalyze binding-dependent DNA ligation between the two DNA species. Ligated DNA can then be selectively amplified by PCR for sequencing. This technology was used to screen a DEL, which was injected (50 nL) alongside bait DNA into *Xenopus* oocytes (1 µL volume). The cells were lysed and submitted to the BTE assay.

Microfluidic techniques have also been leveraged to perform functional assays using DEL. Paegel and co-workers adapted the one-bead-one-compound (OBOC)[5] method to screen functional assays on flow cytometry.[137-139] One channel dispenses protein, whilst another dispenses assay substrate which are combined into a droplet that encapsulates the bead (Figure 5.H.). The compound is released into the droplet by photocleavage, and the fluorescence of each droplet is measured followed by sorting and sequencing of selected beads. Whilst each compound cannot be amplified, beads could theoretically be put through multiple screening cycles. This technology also has the additional advantage of using minute amounts of material.

In parallel to chemical DEL screens, biological DELs have also evolved beyond the simple affinity assay. Notably, the previously discussed SICLOPPS technology is screened via the reverse two hybrid system (RTHS) (Figure 6.A.).[36, 39, 41, 140-141] This approach can study PPI disruption by engineering *E. Coli* cells to only survive when inhibition occurs. The surviving colonies can be cultured and be subjected to other selection rounds to eliminate false positives.

A screening technology which results in directed evolution of biomolecules with little human intervention is phage-assisted continuous evolution (PACE).[142-143] Unlike traditional phage display selection cycles, this technology removes the need for DNA library design and preparation, cell transformation, gene extraction, or cloning of DNA, rendering it almost 'autonomous' and enabling many cycles (e.g. 38 generations per 24 hours). The technology relies on a continuous flow 'lagoon' that contains M13 bacteriophages ('selection phage' (SP)), which rely on protein III (pIII) to grow. These phages have been modified so that pIII production is linked to activity rather than the phage itself. The gene encoding pIII (gIII) is therefore displayed on an 'accessory plasmid' (AP) and triggered when the engineered protein (POI) is active. If the SP encodes for an active POI, then pIII is produced and the phage displaying this protein can replicate and survive in the lagoon (Figure 6.B.). A mutagenesis plasmid (MP) is also used to trigger random mutagenesis to all the host genomes but only affects the SP since other mutated species replicate slower and are washed away in the continuous flow. The mutations which enable the protein to produce pIII are therefore retained and submitted to further rounds of mutagenesis resulting in continuous evolution of the protein. This technology is versatile but has a few requirements: (i) pIII production must be able to be linked to the POI, (ii) the gene encoding the POI must be below 5 kb, and (iii) proteins which require PTMs or disulfide folding are not suitable (protein production occurs in the cytosol of *E. Coli*).[144] Alternative versions have since been developed, including phage-assisted non continuous evolution (PANCE) which replaces the continuous flow by step-wise dilutions, enabling the evolution of proteins with slow or low activity.[144-145] A negative selection process was also developed in order to increase the selection pressure towards a specific role rather that broadening the proteins scope (hence avoiding undesired properties).[146-148] This technology has been used for the evolution of a range of proteins, including CRISPR-Cas9,[149] and botulinum neurotoxin proteases.[150]

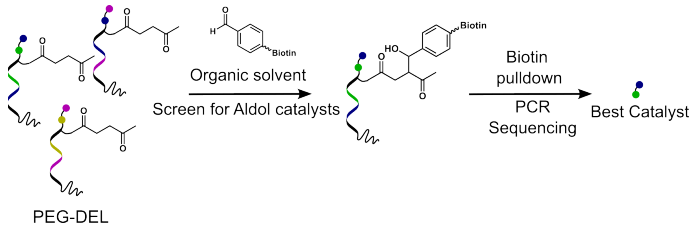


Figure 7. PEG derivatized DEL to enable organic solvent-based catalyst screen

Whilst DELs are mainly used for the discovery of pharmacologically relevant compounds, they have recently been expanded to catalyst development. One-bead-one-compound (OBOC) catalyst discovery was previously limited by its throughput.[151-153] The use of DNA (genotype) to encode the catalyst (phenotype) enabled the screening of a 16.7 million catalyst library.[154] The DNA was chemically modified to contain a PEG 40,000 tail to ensure solubility in a range of organic solvents. Diproline was used as a positive control to screen catalysts for amine-catalysed aldol reaction yielding the conjugation of biotin. The product was enriched using streptavidin/biotin pulldown and subjected to PCR (Figure 7.). The known catalyst was enriched 1,200-fold suggesting this workflow is applicable for the discovery of novel catalyst or the optimization of known ones.

4. Summary and Outlook

The encoding of discrete synthetic organic molecules with DNA tags as an amplifiable barcode has enabled screens of unprecedented size at unprecedent speed with fewer resources. Remarkably, many of the problems that plagued the early days of combinatorial chemistry (solubility of library members, impurities leading to false positives in biochemical assays, heterogenous reaction yield) have proven to be less severe problems in DEL by virtue of the fact that the DNA tag provide homogeneous solubilization and affinity screens are less prone to artifacts than activity-based assays or cell-based assays. Truncated reaction products tend to have lower affinity than final products and even if present as a fraction, the power of DNA amplification can compensate for the low abundance of the final product, in an ideal case. In practice, the noise that arise from libraries of poor quality can make the analysis complicated. Screening large libraries yields overwhelming data sets that require bioinformatic treatment and proper controls to be interpretable. There is not always a straight correlation between binding fitness and sequence count of the tag due to heterogeneous concentration in the starting library, heterogeneous yield in the synthetic sequence (pollution from common truncated intermediates) and amplification aptitude of the tag. This can be filtered out to some extent by analyzing the sequence count of tags vs a control selection or the starting library (enrichment vs sequence count) but that limits the number of compounds since current next-generation sequencing is limited to 10^9 sequence counts.[155] Simulations have shown that screening libraries >10^8 members would yield more false negatives, outweighing the benefit of deep molecular space exploration.[156] Experimental evidence points to the need for 10^4 copies of each library member for good performance in the screen.[157] Thus selection performed in 100 µL at 10 µM should not exceed 10^9 members. Furthermore, binding fitness may not correlate to the sought inhibition or functional fitness. Significant advances have been made on the selection/screening side, extending the scope from a protein of interest immobilized on a bead to selections performed in crude cell extracts with the pertinent complexity of environment or on whole cell, for membrane proteins which were formerly inaccessible to the technology, or functional assays in droplets. Another area of interest for further progress is technologies enabling the 'translation' of the tag into the synthetic molecule it encodes. While the examples presented in this review show that this is possible, the full repertoire of DEL compatible chemistry has not yet been harnessed. These advances are essential to

explore a diversity space that is numerically larger than the library screened. While the chemical diversity space that biochemical methods explore is restricted to biopolymers, iterative cycles of screening and amplification are routinely practiced in order to converge on the fittest results.

Tremendous advances have also been made on the scope of the chemical diversity and screening of such biochemically encoded libraries. On the chemical diversity front, macrocycles, or polycyclic peptides with more desirable pharmacology[158-159] are accessible as well as oligonucleotides that encompass functionality extending far beyond the canonical nucleobases. On the screening side, the selection can be performed based on a functional assay rather than a simple binding fitness. Last but not least, technologies for the continuous evolution (PACE) have been reported, combining the amplification of fittest members in a functional assay with diversification, tapping into the full power of Darwinian evolution at warp speed (more than 1 cycle per hour). However, these technologies are still confined to a chemical diversity space that is quite different from drug-like small molecules. Progress in DEL synthesis has shown that the scope of chemistry that can be used to prepare libraries is far broader than originally anticipated but most of these advances were made using DEL-encoding that does not allow molecular evolution. From a drug discovery perspective, natural products are the crib of therapeutics and benefit from the wisdom of Darwinian evolution to achieve their function. DNA-encoding techniques offer the opportunity to replicate this mechanism and extend it to new chemistries at an unprecedented scale. It is clear that only a fraction of the chemical space has been explored in traditional screens, DNA-encoded technologies are poised to allow new horizons to be discovered. It is now possible to perform continuous evolution using phage display (PACE), coupling these advances to semisynthetic derivatization of phage libraries to extend PACE to peptidic macrocycles is within reach. Similarly, functional screens have been reported with synthetic DELs using OBOC. Coupling this screening technology to DEL formats enabling the translation of DNA into synthetic product is also within reach in order to tie the evolutionary pressure of synthetic chemistry transformation to a cellular phenotype. While the recipe for Darwinian evolution is simple (select for desirable properties, produce more of the selected molecules while introducing some changes, reiterate); its practical implementation to evolve small molecules requires technologies to translate DNA into synthetic molecules. This translation step imposes some constrains that other DEL formats (select/decode) don't have. It may be that the select/decode format is good enough for some, even most drug discovery efforts, but current DEL libraries are reaching the numerical limits of the technology. Molecular evolution circumvents this numerical limitation, provided the translation chemistry can give access to the fittest entity.

Keywords: Combinatorial Chemistry • Darwinian evolution • DNA encoded libraries • Self-assembly • Supramolecular Chemistry

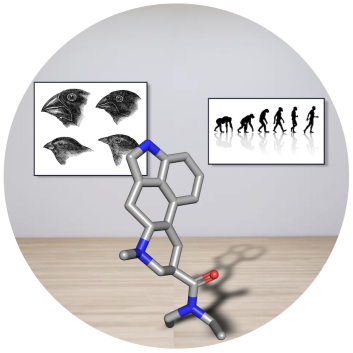
Acknowledgments: We thank our collaborators in the area of DNA encoding and the funding agencies that contributed to our work (ERC, SNSF, NCCR chemical biology). This perspective is not intended to be comprehensive but instead aims to highlight subset of studies that contribute to molecular evolution of synthetic products, and we apologize for arbitrary omissions. Figures were created using BioRender.com.

[1] R. S. Bohacek, C. McMartin, W. C. Guida, *Med. Res. Rev.* **1996**, *16*, 3-50.
[2] J. L. Reymond, *Acc. Chem. Res.* **2015**, *48*, 722-730.
[3] S. Brenner, R. A. Lerner, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 5381-5383.
[4] Á. Furka, F. Sebestyén, M. Asgedom, G. Dibó, *Int. J. Pept. Prot. Res.* **1991**, *37*, 487-493.
[5] K. S. Lam, S. E. Salmon, E. M. Hersh, V. J. Hruby, W. M. Kazmierski, R. J. Knapp, *Nature* **1991**, *354*, 82-84.
[6] Z. J. Gartner, B. N. Tse, R. Grubina, J. B. Doyon, T. M. Snyder, D. R. Liu, *Science* **2004**, *305*, 1601-1605.
[7] S. Melkko, J. Scheuermann, C. E. Dumelin, D. Neri, *Nat. Biotechnol.* **2004**, *22*, 568-574.
[8] F. Debaene, L. Mejias, J. L. Harris, N. Winssinger, *Tetrahedron* **2004**, *60*, 8677-8690.
[9] D. R. Halpin, P. B. Harbury, *PLOS Biol.* **2004**, *2*, e173.
[10] M. A. Clark, R. A. Acharya, C. C. Arico-Muendel, S. L. Belyanskaya, D. R. Benjamin, N. R. Carlson, P. A. Centrella, C. H. Chiu, S. P. Creaser, J. W. Cuozzo, C. P. Davie, Y. Ding, G. J. Franklin, K. D. Franzen, M. L. Gefter, S. P. Hale, N. J. V. Hansen, D. I. Israel, J. Jiang, M. J. Kavarana, M. S. Kelley, C. S. Kollmann, F. Li, K. Lind, S. Mataruse, P. F. Medeiros, J. A. Messer, P. Myers, H. O'Keefe, M. C. Oliff, C. E. Rise, A. L. Satz, S. R. Skinner, J. L. Svendsen, L. Tang, K. van Vloten, R. W. Wagner, G. Yao, B. Zhao, B. A. Morgan, *Nat. Chem. Biol.* **2009**, *5*, 647-654.
[11] J. Harris, D. E. Mason, J. Li, K. W. Burdick, B. J. Backes, T. Chen, A. Shipway, G. Van Heeke, L. Gough, A. Ghaemmaghami, F. Shakib, F. Debaene, N. Winssinger, *Chem. Biol.* **2004**, *11*, 1361-1372.
[12] G. P. Smith, *Science* **1985**, *228*, 1315-1317.
[13] S. F. Parmley, G. P. Smith, *Gene* **1988**, *73*, 305-318.
[14] J. K. Scott, G. P. Smith, *Science* **1990**, *249*, 386-390.
[15] J. McCafferty, A. D. Griffiths, G. Winter, D. J. Chiswell, *Nature* **1990**, *348*, 552-554.
[16] T. Clackson, H. R. Hoogenboom, A. D. Griffiths, G. Winter, *Nature* **1991**, *352*, 624-628.
[17] A. Frenzel, T. Schirrmann, M. Hust, *mAbs* **2016**, *8*, 1177-1194.
[18] K. T. O'Neil, R. H. Hoess, S. A. Jackson, N. S. Ramachandran, S. A. Mousa, W. F. DeGrado, *Prot. Struct. Funct. Bioinform.* **1992**, *14*, 509-515.
[19] C. Heinis, T. Rutherford, S. Freund, G. Winter, *Nat. Chem. Biol.* **2009**, *5*, 502-507.
[20] C. Heinis, G. Winter, *Curr. Opin. Chem. Biol.* **2015**, *26*, 89-98.
[21] S. S. Kale, C. Villequey, X. D. Kong, A. Zorzi, K. Deyle, C. Heinis, *Nat. Chem.* **2018**, *10*, 715-723.
[22] V. Carle, Y. Wu, R. Mukherjee, X.-D. Kong, C. Rogg, Q. Laurent, E. Cecere, C. Villequey, M. S. Konakalla, T. Maric, C. Lamers, C. Díaz-Perlas, K. Butler, J. Goto, B. Stegmayr, C. Heinis, *J. Med. Chem.* **2021**, *64*, 6802-6813.
[23] S. Chen, S. Lovell, S. Lee, M. Fellner, P. D. Mace, M. Bogyo, *Nat. Biotechnol.* **2021**, *39*, 490-498.

[24] T. Navaratna, L. Atangcho, M. Mahajan, V. Subramanian, M. Case, A. Min, D. Tresnak, G. M. Thurber, *J. Am. Chem. Soc.* **2020**, *142*, 1882-1894.
[25] J. Hanes, A. Plückthun, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 4937-4942.
[26] R. W. Roberts, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12297-12302.
[27] K. Josephson, A. Ricardo, J. W. Szostak, *Drug Discov. Today* **2014**, *19*, 388-399.
[28] Y. C. Huang, M. M. Wiedmann, H. Suga, *Chem. Rev.* **2019**, *119*, 10360-10391.
[29] H. Murakami, H. Saito, H. Suga, *Chem. Biol.* **2003**, *10*, 655-662.
[30] J. Morimoto, Y. Hayashi, K. Iwasaki, H. Suga, *Acc. Chem. Res.* **2011**, *44*, 1359-1368.
[31] Y. Goto, T. Katoh, H. Suga, *Nat. Protoc.* **2011**, *6*, 779-790.
[32] Y. Goto, H. Suga, *Acc. Chem. Res.* **2021**, *54*, 3604-3617.
[33] T. Passioura, H. Suga, *Chem. Commun.* **2017**, *53*, 1931-1940.
[34] A. Norman, C. Franck, M. Christie, P. M. E. Hawkins, K. Patel, A. S. Ashhurst, A. Aggarwal, J. K. K. Low, R. Siddiquee, C. L. Ashley, M. Steain, J. A. Triccas, S. Turville, J. P. Mackay, T. Passioura, R. J. Payne, *ACS Cent. Sci.* **2021**, *7*, 1001-1008.
[35] J. Johansen-Leete, S. Ullrich, S. E. Fry, R. Frkic, M. J. Bedding, A. Aggarwal, A. S. Ashhurst, K. B. Ekanayake, M. C. Mahawaththa, V. M. Sasi, S. Luedtke, D. J. Ford, A. J. O'Donoghue, T. Passioura, M. Larance, G. Otting, S. Turville, C. J. Jackson, C. Nitsche, R. J. Payne, *Chem. Sci.* **2022**, *13*, 3826-3836.
[36] C. P. Scott, E. Abel-Santos, M. Wall, D. C. Wahnon, S. J. Benkovic, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13638-13643.
[37] A. Tavassoli, S. J. Benkovic, *Nat. Protoc.* **2007**, *2*, 1126-1133.
[38] A. Tavassoli, S. J. Benkovic, *Angew. Chem. Int. Ed.* **2005**, *44*, 2760-2763.
[39] A. R. Horswill, S. N. Savinov, S. J. Benkovic, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15591-15596.
[40] T. S. Young, D. D. Young, I. Ahmad, J. M. Louis, S. J. Benkovic, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 11052-11056.
[41] K. R. Lennard, A. Tavassoli, *Chem. Eur. J.* **2014**, *20*, 10608-10614.
[42] C. Tuerk, L. Gold, *Science* **1990**, *249*, 505-510.
[43] A. D. Ellington, J. W. Szostak, *Nature* **1990**, *346*, 818-822.
[44] A. A. Beaudry, G. F. Joyce, *Science* **1992**, *257*, 635-641.
[45] G. F. Joyce, *Curr. Opin. Struct. Biol.* **1994**, *4*, 331-336.
[46] M. C. Wright, G. F. Joyce, *Science* **1997**, *276*, 614-617.
[47] K. Duffy, S. Arangundy-Franklin, P. Holliger, *BMC Biol.* **2020**, *18*, 112.
[48] M. López-Tena, S.-K. Chen, N. Winssinger, *Bioconj. Chem.* **2022**, doi: 10.1021/acs.bioconjchem.1022c00292.
[49] R. Hili, J. Niu, D. R. Liu, *J. Am. Chem. Soc.* **2013**, *135*, 98-101.
[50] Y. Lei, D. Kong, R. Hili, *ACS Comb. Sci.* **2015**, *17*, 716-721.
[51] C. Guo, C. P. Watkins, R. Hili, *J. Am. Chem. Soc.* **2015**, *137*, 11191-11196.
[52] D. H. Kong, W. Yeung, R. Hili, *J. Am. Chem. Soc.* **2017**, *139*, 13977-13980.
[53] Z. Chen, P. A. Lichtor, A. P. Berliner, J. C. Chen, D. R. Liu, *Nat. Chem.* **2018**, *10*, 420-427.
[54] Q. F. Li, V. A. Maola, N. Chim, J. Hussain, A. Lozoya-Colinas, J. C. Chaput, *J. Am. Chem. Soc.* **2021**, *143*, 17761-17768.
[55] S. Arangundy-Franklin, A. I. Taylor, B. T. Porebski, V. Genna, S. Peak-Chew, A. Vaisman, R. Woodgate, M. Orozco, P. Holliger, *Nat. Chem.* **2019**, *11*, 533-542.
[56] T. R. Oppewal, I. D. Jansen, J. Hekelaar, C. Mayer, *J. Am. Chem. Soc.* **2022**, *144*, 3644-3652.
[57] J. Y. K. Wong, R. Mukherjee, J. Y. Miao, O. Bilyk, V. Triana, M. Miskolzie, A. Henninot, J. J. Dwyer, S. Kharchenko, A. Iampolska, D. M. Volochnyuk, Y. S. Lin, L. M. Postovit, R. Derda, *Chem. Sci.* **2021**, *12*, 9694-9703.
[58] A. I. Ekanayake, L. Sobze, P. Kelich, J. Youk, N. J. Bennett, R. Mukherjee, A. Bhardwaj, F. Wuest, L. Vukovic, R. Derda, *J. Am. Chem. Soc.* **2021**, *143*, 5497-5507.
[59] R. L. Redman, I. J. Krauss, *J. Am. Chem. Soc.* **2021**, *143*, 8565-8571.
[60] X. Y. Li, D. R. Liu, *Angew. Chem. Int. Ed.* **2004**, *43*, 4848-4870.
[61] J. L. Harris, N. Winssinger, *Chem. Eur. J.* **2005**, *11*, 6792-6801.
[62] D. R. Halpin, P. B. Harbury, *PLOS Biol.* **2004**, *2*, e174.
[63] J. Scheuermann, D. Neri, *Curr. Opin. Chem. Biol.* **2015**, *26*, 99-103.
[64] F. Buller, L. Mannocci, Y. X. Zhang, C. E. Dumelin, J. Scheuermann, D. Neri, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5926-5931.
[65] L. Mannocci, Y. X. Zhang, J. Scheuermann, M. Leimbacher, G. De Bellis, E. Rizzi, C. Dumelin, S. Melkko, D. Neri, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 17670-17675.
[66] A. D. Keefe, M. A. Clark, C. D. Hupp, A. Litovchick, Y. Zhang, *Curr. Opin. Chem. Biol.* **2015**, *26*, 80-88.
[67] R. A. Lerner, S. Brenner, *Angew. Chem. Int. Ed.* **2017**, *56*, 1164-1165.
[68] R. E. Kleiner, C. E. Dumelin, D. R. Liu, *Chem. Soc. Rev.* **2011**, *40*, 5707-5717.
[69] R. M. Franzini, D. Neri, J. Scheuermann, *Acc. Chem. Res.* **2014**, *47*, 1247-1255.
[70] C. Zambaldo, S. Barluenga, N. Winssinger, *Curr. Opin. Chem. Biol.* **2015**, *26*, 8-15.
[71] R. A. Goodnow, C. E. Dumelin, A. D. Keefe, *Nat. Rev. Drug Discov.* **2017**, *16*, 131-147.
[72] D. Neri, R. A. Lerner, *Annu. Rev. Biochem.* **2018**, *87*, 479-502.
[73] G. X. Zhao, Y. R. Huang, Y. Zhou, Y. Z. Li, X. Y. Li, *Expert Opin. Drug Discov.* **2019**, *14*, 735-753.
[74] P. R. Fitzgerald, B. M. Paegel, *Chem. Rev.* **2021**, *121*, 7155-7177.
[75] A. L. Satz, A. Brunschweiger, M. E. Flanagan, A. Gloger, N. J. V. Hansen, L. Kuai, V. B. K. Kunig, X. Lu, D. Madsen, L. A. Marcaurelle, C. Mulrooney, G. O'Donovan, S. Sakata, J. Scheuermann, *Nat. Rev. Meth. Pri.* **2022**, *2*, 3.
[76] Z. J. Gartner, D. R. Liu, *J. Am. Chem. Soc.* **2001**, *123*, 6961-6963.
[77] B. N. Tse, T. M. Snyder, Y. Shen, D. R. Liu, *J. Am. Chem. Soc.* **2008**, *130*, 15611-15626.
[78] R. E. Kleiner, C. E. Dumelin, G. C. Tiu, K. Sakurai, D. R. Liu, *J. Am. Chem. Soc.* **2010**, *132*, 11779-11791.
[79] G. Georghiou, R. E. Kleiner, M. Pulkoski-Gross, D. R. Liu, M. A. Seeliger, *Nat. Chem. Biol.* **2012**, *8*, 366-374.
[80] D. L. Usanov, A. I. Chan, J. P. Maianti, D. R. Liu, *Nat. Chem.* **2018**, *10*, 704-714.
[81] M. H. Hansen, P. Blakskjaer, L. K. Petersen, T. H. Hansen, J. W. Hojfeldt, K. V. Gothelf, N. J. V. Hansen, *J. Am. Chem. Soc.* **2009**, *131*, 1322-1327.
[82] C. Cao, P. Zhao, Z. Li, Z. Chen, Y. Huang, Y. Bai, X. Li, *Chem. Commun.* **2014**, *50*, 10997-10999.
[83] Y. Z. Li, P. Zhao, M. D. Zhang, X. Y. Zhao, X. Y. Li, *J. Am. Chem. Soc.* **2013**, *135*, 17727-17730.

[84] Y. Brudno, M. E. Birnbaum, R. E. Kleiner, D. R. Liu, *Nat. Chem. Biol.* **2010**, *6*, 148-155.
[85] J. Niu, R. Hili, D. R. Liu, *Nat. Chem.* **2013**, *5*, 282-292.
[86] K. Murayama, H. Okita, T. Kuriki, H. Asanuma, *Nat. Commun.* **2021**, *12*, 804.
[87] S. J. Wrenn, R. M. Weisinger, D. R. Halpin, P. B. Harbury, *J. Am. Chem. Soc.* **2007**, *129*, 13137-13143.
[88] R. M. Weisinger, S. J. Wrenn, P. B. Harbury, *PLoS One* **2012**, *7*, e28056.
[89] C. J. Krusemark, N. P. Tilmans, P. O. Brown, P. B. Harbury, *PLoS One* **2016**, *11*, e0154765.
[90] S. Melkko, Y. Zhang, C. E. Dumelin, J. Scheuermann, D. Neri, *Angew. Chem. Int. Ed.* **2007**, *46*, 4671-4674.
[91] M. Wichert, N. Krall, W. Decurtins, R. M. Franzini, F. Pretto, P. Schneider, D. Neri, J. Scheuermann, *Nat. Chem.* **2015**, *7*, 241-249.
[92] S. Oehler, L. Plais, G. Bassi, D. Neri, J. Scheuermann, *Chem. Commun.* **2021**, *57*, 12289-12292.
[93] L. Plais, A. Lessing, M. Keller, A. Martinelli, S. Oehler, G. Bassi, D. Neri, J. Scheuermann, *Chem. Sci.* **2022**, *13*, 967-974.
[94] B. B. Shi, Y. Zhou, X. Y. Li, *RSC Chem. Biol.* **2022**, *3*, 407-419.
[95] F. V. Reddavide, W. L. Lin, S. Lehnert, Y. X. Zhang, *Angew. Chem. Int. Ed.* **2015**, *54*, 7924-7928.
[96] G. Li, W. L. Zheng, Z. T. Chen, Y. Zhou, Y. Liu, J. R. Yang, Y. Y. Huang, X. Y. Li, *Chem. Sci.* **2015**, *6*, 7097-7104.
[97] I. Huc, J. M. Lehn, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 2106-2110.
[98] O. Ramstrom, J. M. Lehn, *Nat. Rev. Drug Discov.* **2002**, *1*, 26-36.
[99] F. V. Reddavide, M. Y. Cui, W. L. Lin, N. Q. Fu, S. Heiden, H. Andrade, M. Thompson, Y. X. Zhang, *Chem. Commun.* **2019**, *55*, 3753-3756.
[100] Y. Q. Deng, J. Z. Peng, F. Xiong, Y. A. Song, Y. Zhou, J. F. Zhang, F. S. Lam, C. Xie, W. Y. Shen, Y. R. Huang, L. Meng, X. Y. Li, *Angew. Chem. Int. Ed.* **2020**, *59*, 14965-14972.
[101] Y. Zhou, C. Li, J. Z. Peng, L. X. Xie, L. Men, Q. R. Li, J. F. Zhang, X. D. Li, X. Li, X. H. Huang, X. Y. Li, *J. Am. Chem. Soc.* **2018**, *140*, 15859-15867.
[102] G. X. Zhao, S. T. Zhong, G. Zhang, Y. F. Li, Y. Z. Li, *Angew. Chem. Int. Ed.* **2022**, *61*, e202115157.
[103] G. X. Zhao, X. H. Fan, Y. F. Li, G. Zhang, Y. Z. Li, *ChemBioChem* **2022**, e202200025.
[104] Y. Zhou, W. Y. Shen, J. Z. Peng, Y. Q. Deng, X. Y. Li, *Bioorg. Med. Chem.* **2021**, *45*, 116328.
[105] J. P. Daguer, M. Ciobanu, S. Alvarez, S. Barluenga, N. Winssinger, *Chem. Sci.* **2011**, *2*, 625-632.
[106] M. Ciobanu, K. T. Huang, J. P. Daguer, S. Barluenga, O. Chaloin, E. Schaeffer, C. G. Mueller, D. A. Mitchell, N. Winssinger, *Chem. Commun.* **2011**, *47*, 9321-9323.
[107] W. P. C. Stemmer, *Nature* **1994**, *370*, 389-391.
[108] A. Crameri, S.-A. Raillard, E. Bermudez, W. P. C. Stemmer, *Nature* **1998**, *391*, 288-291.
[109] G. P. Smith, *Nature* **1994**, *370*, 324-325.
[110] B. R. Vummidi, L. Farrera-Soler, J. P. Daguer, M. Dockerill, S. Barluenga, N. Winssinger, *Nat. Chem.* **2022**, *14*, 141-152.
[111] T. Machida, S. Dutt, N. Winssinger, *Angew. Chem. Int. Ed.* **2016**, *55*, 8595-8598.
[112] P. M. Sabale, M. Imiolek, P. Raia, S. Barluenga, N. Winssinger, *J. Am. Chem. Soc.* **2021**, *143*, 18932-18940.
[113] L. Roglin, M. R. Ahmadian, O. Seitz, *Angew. Chem. Int. Ed.* **2007**, *46*, 2704-2707.
[114] S. Thurley, L. Roglin, O. Seitz, *J. Am. Chem. Soc.* **2007**, *129*, 12693-12695.
[115] A. L. Satz, *ACS Chemical Biology* **2015**, *10*, 2237-2245.
[116] Y. R. Huang, Y. Z. Li, X. Y. Li, *Nat. Chem.* **2022**, *14*, 129-140.
[117] L. M. McGregor, D. J. Gorin, C. E. Dumelin, D. R. Liu, *J. Am. Chem. Soc.* **2010**, *132*, 15522-15524.
[118] L. M. McGregor, T. Jain, D. R. Liu, *J. Am. Chem. Soc.* **2014**, *136*, 3264-3270.
[119] G. Li, Y. Liu, Y. Liu, L. Chen, S. Wu, Y. Liu, X. Li, *Angew. Chem. Int. Ed.* **2013**, *52*, 9544-9549.
[120] P. Zhao, Z. Chen, Y. Li, D. Sun, Y. Gao, Y. Huang, X. Li, *Angew. Chem. Int. Ed.* **2014**, *53*, 10056-10059.
[121] B. B. Shi, Y. Q. Deng, P. Zhao, X. Y. Li, *Bioconj. Chem.* **2017**, *28*, 2293-2301.
[122] B. B. Shi, Y. Q. Deng, X. Y. Li, *ACS Comb. Sci.* **2019**, *21*, 345-349.
[123] Y. R. Huang, L. Meng, Q. G. Nie, Y. Zhou, L. D. Chen, S. L. Yang, Y. M. E. Fung, X. M. Li, C. Huang, Y. Cao, Y. Z. Li, X. Y. Li, *Nat. Chem.* **2021**, *13*, 77-88.
[124] Y. R. Huang, Y. Q. Deng, J. F. Zhang, L. Meng, X. Y. Li, *Chem. Commun.* **2021**, *57*, 3769-3772.
[125] Y. A. Song, X. Y. Li, *Acc. Chem. Res.* **2021**, *54*, 3491-3503.
[126] S. Ahn, A. W. Kahsai, B. Pani, Q.-T. Wang, S. Zhao, A. L. Wall, R. T. Strachan, D. P. Staus, L. M. Wingler, L. D. Sun, J. Sinnaeve, M. Choi, T. Cho, T. T. Xu, G. M. Hansen, M. B. Burnett, J. E. Lamerdin, D. L. Bassoni, B. J. Gavino, G. Husemoen, E. K. Olsen, T. Franch, S. Costanzi, X. Chen, R. J. Lefkowitz, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 1708-1713.
[127] S. Ahn, B. Pani, A. W. Kahsai, E. K. Olsen, G. Husemoen, M. Vestergaard, L. Jin, S. Zhao, L. M. Wingler, P. K. Rambarat, R. K. Simhal, T. T. Xu, L. D. Sun, P. J. Shim, D. P. Staus, L.-Y. Huang, T. Franch, X. Chen, R. J. Lefkowitz, *Mol. Pharmacol.* **2018**, *94*, 850-861.
[128] Y. Liu, W. Zheng, W. Zhang, N. Chen, Y. Liu, L. Chen, X. Zhou, X. Chen, H. Zheng, X. Li, *Chem. Sci.* **2015**, *6*, 745-751.
[129] X. Bai, C. Lu, J. Jin, S. Tian, Z. Guo, P. Chen, G. Zhai, S. Zheng, X. He, E. Fan, Y. Zhang, K. Zhang, *Angew. Chem. Int. Ed.* **2016**, *55*, 7993-7997.
[130] D.-Y. Wang, Y. Cao, L.-Y. Zheng, L.-D. Chen, X.-F. Chen, Z.-Y. Hong, Z.-Y. Zhu, X. Li, Y.-F. Chai, *Chem. Eur. J.* **2017**, *23*, 10906-10914.
[131] B. Cai, D. Kim, S. Akhand, Y. Sun, R. J. Cassell, A. Alpsoy, E. C. Dykhuizen, R. M. Van Rijn, M. K. Wendt, C. J. Krusemark, *J. Am. Chem. Soc.* **2019**, *141*, 17057-17061.
[132] B. Cai, C. J. Krusemark, *Angew. Chem. Int. Ed.* **2022**, *61*, e202113515.
[133] B. Cai, A. Mhetre, C. Krusemark, *ChemRxiv* **2022**, https://doi.org/10.26434/chemrxiv-22022-l26097f.
[134] P. Blakskjaer, T. Heitner, N. J. V. Hansen, *Curr. Opin. Chem. Biol.* **2015**, *26*, 62-71.
[135] L. K. Petersen, P. Blakskjær, A. Chaikuad, A. B. Christensen, J. Dietvorst, J. Holmkvist, S. Knapp, M. Kořínek, L. K. Larsen, A. E. Pedersen, S. Röhm, F. A. Sløk, N. J. V. Hansen, *MedChemComm* **2016**, *7*, 1332-1339.
[136] L. K. Petersen, A. B. Christensen, J. Andersen, C. G. Folkesson, O. Kristensen, C. Andersen, A. Alzu, F. A. Slok, P. Blakskjaer, D. Madsen, C. Azevedo, I. Micco, N. J. V. Hansen, *J. Am. Chem. Soc.* **2021**, *143*, 2751-2756.
[137] A. B. MacConnell, A. K. Price, B. M. Paegel, *ACS Comb. Sci.* **2017**, *19*, 181-192.
[138] W. G. Cochrane, A. L. Hackler, V. J. Cavett, A. K. Price, B. M. Paegel, *Anal. Chem.* **2017**, *89*, 13227-13234.
[139] W. G. Cochrane, M. L. Malone, V. Q. Dane, V. Cavett, A. L. Satz, B. M. Paegel, *ACS Comb. Sci.* **2019**, *21*, 425-435.

[140] M. Vidal, R. K. Brachmann, A. Fattaey, E. Harlow, J. D. Boeke, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 10315-10320.
[141] M. A. White, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 10001-10003.
[142] K. M. Esvelt, J. C. Carlson, D. R. Liu, *Nature* **2011**, *472*, 499-503.
[143] S. C. Popa, I. Inamoto, B. W. Thuronyi, J. A. Shin, *ACS Omega* **2020**, *5*, 26957-26966.
[144] S. M. Miller, T. Wang, D. R. Liu, *Nat. Protoc.* **2020**, *15*, 4101-4127.
[145] T. B. Roth, B. M. Woolston, G. Stephanopoulos, D. R. Liu, *ACS Synth. Biol.* **2019**, *8*, 796-806.
[146] J. B. Doyon, V. Pattanayak, C. B. Meyer, D. R. Liu, *J. Am. Chem. Soc.* **2006**, *128*, 2477-2484.
[147] C. A. Tracewell, F. H. Arnold, *Curr. Opin. Chem. Biol.* **2009**, *13*, 3-9.
[148] J. C. Carlson, A. H. Badran, D. A. Guggiana-Nilo, D. R. Liu, *Nat. Chem. Biol.* **2014**, *10*, 216-222.
[149] J. H. Hu, S. M. Miller, M. H. Geurts, W. Tang, L. Chen, N. Sun, C. M. Zeina, X. Gao, H. A. Rees, Z. Lin, D. R. Liu, *Nature* **2018**, *556*, 57-63.
[150] T. R. Blum, H. Liu, M. S. Packer, X. Xiong, P.-G. Lee, S. Zhang, M. Richter, G. Minasov, K. J. F. Satchell, M. Dong, D. R. Liu, *Science* **2021**, *371*, 803-810.
[151] P. A. Lichtor, S. J. Miller, *ACS Comb. Sci.* **2011**, *13*, 321-326.
[152] P. A. Lichtor, S. J. Miller, *Nat. Chem.* **2012**, *4*, 990-995.
[153] K. Akagawa, J. Sen, K. Kudo, *Angew. Chem. Int. Ed.* **2013**, *52*, 11585-11588.
[154] K. D. Hook, J. T. Chambers, R. Hili, *Chem. Sci.* **2017**, *8*, 7072-7076.
[155] J. Shendure, S. Balasubramanian, G. M. Church, W. Gilbert, J. Rogers, J. A. Schloss, R. H. Waterston, *Nature* **2017**, *550*, 345-353.
[156] A. L. Satz, R. Hochstrasser, A. C. Petersen, *ACS Comb. Sci.* **2017**, *19*, 234-238.
[157] A. Sannino, E. Gabriele, M. Bigatti, S. Mulatto, J. Piazzi, J. Scheuermann, D. Neri, E. J. Donckele, F. Samain, *ChemBioChem* **2019**, *20*, 955-962.
[158] A. Zorzi, K. Deyle, C. Heinis, *Curr. Opin. Chem. Biol.* **2017**, *38*, 24-29.
[159] R. Obexer, L. J. Walport, H. Suga, *Curr. Opin. Chem. Biol.* **2017**, *38*, 52-61.



Taking cues from Darwinian evolution, is it possible to apply molecular evolution to drug-like molecules? The review covers technologies for the synthesis of DNA-encoded libraries enabling translation of the DNA code into synthetic products, explores the different selection format and discuss how these can be brought together to evolve synthetic molecule with Darwinian selection pressure.