Chemical Evolution as a Tool for Molecular Discovery

S. Jarrett Wrenn and Pehr B. Harbury

Department of Biochemistry, Stanford University, Stanford, California 94305;
email: sjwrenn@stanford.edu, harbury@cmgm.stanford.edu

Annu. Rev. Biochem. 2007. 76:331–49

The *Annual Review of Biochemistry* is online at
biochem.annualreviews.org

This article's doi:
10.1146/annurev.biochem.76.062205.122741

0066-4154/07/0707-0331$20.00

Key Words

combinatorial chemistry, DNA display, DNA-templated synthesis, in vitro evolution, in vitro selection, small molecules

Abstract

In modern academic and industrial laboratories, evolutionary strategies are used routinely to identify biopolymers with novel activities. Large libraries of nucleic acids ($\sim 10^{15}$) or peptides and proteins ($\sim 10^{13}$) can be subjected to multiple rounds of selective pressure, amplification, and diversification, yielding individual sequences with desirable properties. Although the evolutionary approach is a powerful search tool, the chemical nature of biopolymers is not suited for all purposes. Application of evolutionary strategies to libraries of arbitrary chemical composition would overcome this problem, and radically change traditional small-molecule discovery. The chemical make-up of in vitro evolution libraries has necessarily been limited, because library synthesis relies on enzymes. A great deal of current research focuses on expanding the chemical repertoire of in vitro evolution by (*a*) broadening enzyme substrate specificities to include unnatural building blocks, or (*b*) developing methods to translate DNA sequences into multistep organic syntheses. We discuss the strengths and weaknesses of the approaches, review the successes, and consider the future of chemical evolution as a tool.

Contents

INTRODUCTION

Nearly 40 years ago, Spiegelman and coworkers introduced the concept of an extracellular Darwinian experiment (1). In that work, nucleic acid sequences capable of self-duplication in the presence of nucleotides and Qβ replicase were serially diluted over decreasing time intervals. Fast-growing variants were preferentially replicated, producing mutants with altered base composition and genome deletions. The long-term significance of the work was profound. It demonstrated that the same basic evolutionary principles responsible for organismal evolution could be applied to biomolecule populations in vitro, yielding variants with exaggerated attributes dictated by an artificial selective pressure. Through iterative rounds of selection, amplification, and diversification, rare molecules with new functions would emerge from an enormous collection of molecular possibilities.

The notion of in vitro evolution was explored extensively in the 1980s and 1990s, leading to the development of several powerful techniques for the selection of functional nucleic acids (2–4) and proteins (5–9). These approaches rely on the translation of nucleic acid genes (DNA) into corresponding gene products (RNA, protein) while maintaining a physical link between gene and gene product (**Figure 1**). The coupling between gene and gene product enables the use of bulk functional selections, post-selection amplification, and library propagation through multiple generations. For libraries with theoretical complexities larger than the physical library size, diversification between rounds allows for large chemical spaces to be sampled indirectly. This diversification distinguishes in vitro evolution from in vitro selection, and can produce functional molecules unlikely to have existed in a random starting population. Selection and evolution methods are now routinely used by both academic research laboratories and biotechnology companies to isolate potent ligands, inhibitors, and catalysts [reviewed in (10)].

Although in vitro evolution libraries have been based predominately on biopolymers, the strategy is conceptually generalizable to any type of chemical library, as long as a means exists to direct the synthesis of a heterologous gene product with a nucleic acid gene. Application of evolutionary techniques to semi- or wholly synthetic molecular libraries can provide access to compound collections orders of magnitude more complex than those accessible with state-of-the-art high-throughput screening technology ($>10^{12}$ vs 10^6). The past decade has seen a strong focus directed to expanding the chemical repertoire of in vitro evolution techniques in hopes of using the approach to identify functional unnatural molecules. Below we review the major

progress reported thus far, and discuss the future of chemical evolution.

MODIFIED BIOPOLYMERS

One approach toward chemical evolution is to exploit the promiscuity of biosynthetic enzymes (polymerases, the ribosome, etc.) to incorporate unnatural building blocks (unnatural nucleotides or amino acids) into biopolymers. Using this strategy, modified biopolymer libraries can be generated enzymatically and then subjected to in vitro selections.

Modified Nucleic Acids

The productive enzymatic incorporation of unnatural nucleotides into nucleic acids requires that DNA and RNA polymerases efficiently utilize modified nucleotides while faithfully replicating a nucleic acid sequence. The modified building blocks must be well-behaved as nucleotide triphosphates during incorporation, and as components of the template strand during amplification. Although there are multiple reports of polymerases that incorporate monomers with entirely unnatural nucleobases (11), most efforts toward in vitro evolution have involved nucleotide modifications that do not disrupt Watson-Crick base pairing [e.g., C5-substitutions of pyrimidines (12, 13)].

The earliest example of this approach involved the replacement of thymidine with 5-(1-pentynyl)-2′-deoxyuridine (**Figure 2*a***) in a 60-nucleotide single-stranded DNA library. The library was selected for binding to thrombin (14). The resulting pentynyl-dU aptamers showed dissociation constants for thrombin similar to those derived from natural aptamer libraries, but the binding motifs differed from the natural aptamers. In addition to validating the use of unnatural monomers for in vitro selection, the result demonstrated that the unnatural moieties could alter the outcome of the selection. A wide variety of C5-substituted deoxyuridines have been successfully incorporated by DNA polymerases and could be included in nucleic acid libraries (15). For example, ATP-binding DNA molecules have been selected from an oligonucleotide library incorporating 5-(3″-aminopropynyl)-2′-deoxyuridine (**Figure 2*a***) (16).

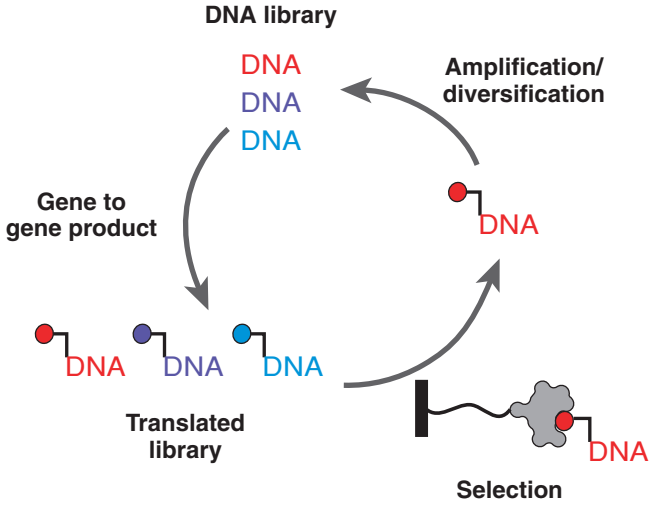


Figure 1

In vitro evolution. A library of nucleic acid genes (*colored DNA*) is translated into corresponding gene products (*colored balls*), while maintaining a physical link between gene and gene product. The translated library is then subjected to selective pressure (binding to immobilized *gray widget*). The post-selection population is amplified and diversified, generating the second-generation gene pool for translation. The process is repeated until the population converges to genes encoding gene products fit to survive the selection.

The impetus for some of the work on incorporating unnatural nucleotides into nucleic acid libraries has been to improve the stability of the resulting aptamers. Several early examples reported the use of nuclease-resistant 2′-amino pyrimidine containing RNA libraries for in vitro selection. These efforts yielded tight binding ligands to vascular endothelial growth factor (VEGF), basic fibroblast growth factor (bFGF), and human neutrophil elastase with increased stability in human serum (17–19). Similarly, VEGF was targeted with nuclease-resistant aptamers incorporating 2′-fluoropyrimidines, giving ligands of extremely tight binding affinity ($K_d \approx 10^{-10}$) (20). A slightly modified variant of one of these ligands was recently approved by the U.S. F.D.A. for the treatment

of neovascular age-related macular degeneration (pegaptanib or Macugen), making it the first aptamer-derived human therapeutic (21).

Beyond selections for ligands, multiple examples of nucleic acid catalysts containing unnatural moieties have been reported. In the first example, a Cu^{2+}-dependent

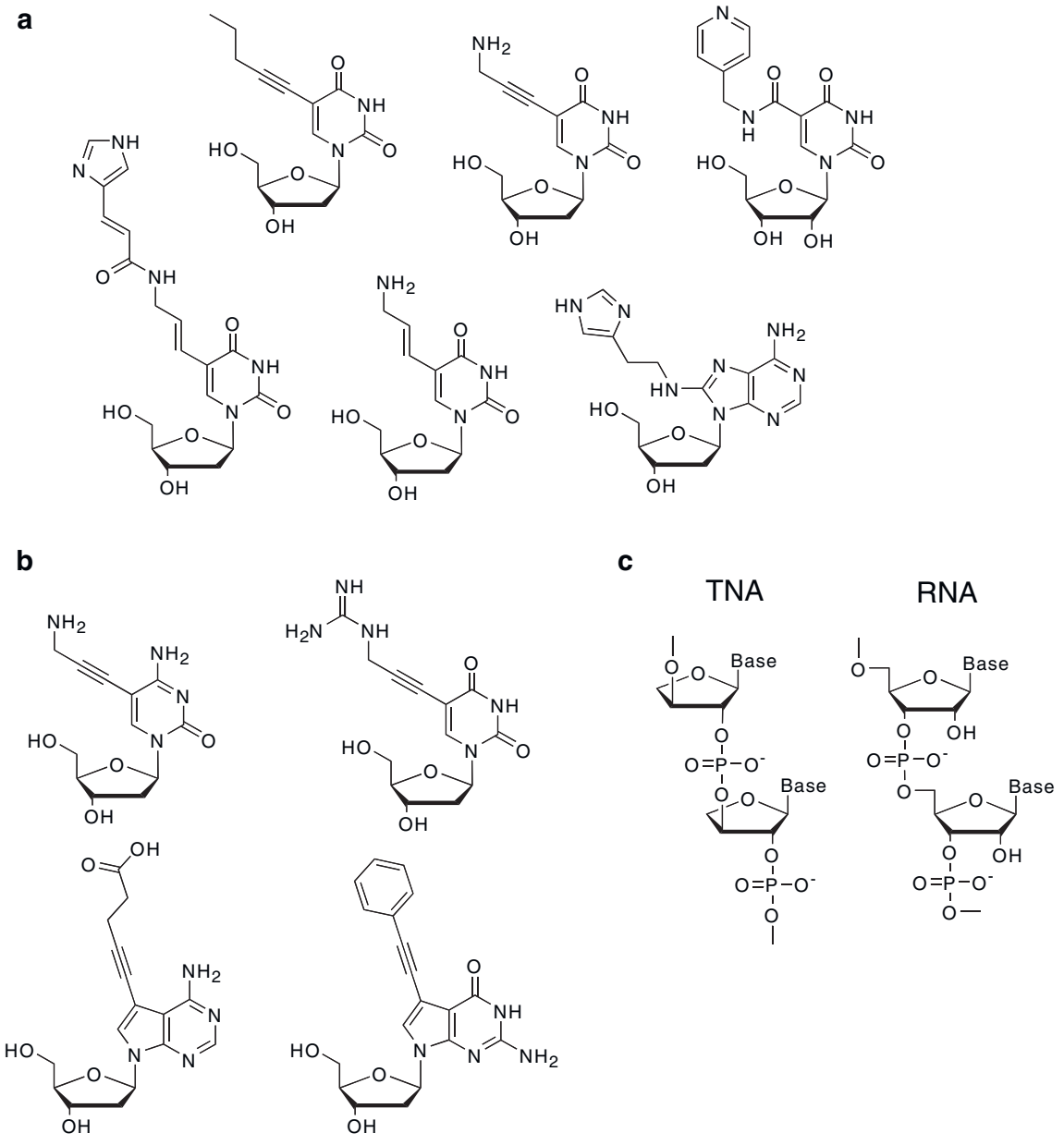


Figure 2

Modified nucleic acids. (*a*) Examples of unnatural nucleic acid monomers incorporated into in vitro selection libraries. (*b*) A fully replicable DNA system using C5-modified pyrimidines and C7-modified 7-deaza purines. (*c*) (3′,2′)-α-L-threose nucleic acid (TNA) shown next to ribose nucleic acid (RNA).

Diels-Alderase was evolved from a library of $\sim 10^{14}$ unique RNA sequences containing C5-pyridylmethylcarboxamid-uracil (**Figure 2*a***) in place of uracil (22). The result was striking because it represented the first observed nucleic acid catalyzed carbon-carbon bond formation reaction. However, the unnatural nucleotide was not essential for creating a Diels-Alderase, as a similar catalyst was later selected from a library of unmodified RNAs aptamers (23). Barbas and coworkers selected a metal-independent RNA cleaving enzyme from a DNA library in which a C5-imidazole-functionalized deoxyuridine (**Figure 2*a***) replaced thymidine (24). Later, the selection of a DNA-based RNAse containing two modified nucleotides was reported (25). In that work, C5-amino-functionalized deoxyuridine and C8-imidazole-functionalized deoxyadenosine replaced thymidine and deoxyadenosine, respectively (**Figure 2*a***).

Selections with more extensively derivatized nucleic acid libraries will likely be forthcoming, as a number of fully unnatural, polymerizable nucleic acid systems have been introduced. For example, a faithfully replicating DNA system using C5-modified pyrimidines and C7-modified 7-deaza purines was reported (26). The appended functionalities included a carboxylic acid, a primary amine, a guanidino group, and a hydrophobic phenyl group (**Figure 2*b***). Recently, Ichida and coworkers described an in vitro selection system for (3′,2′)-α-L-threose nucleic acids (TNAs) (27); however, a selection experiment has not yet been reported (**Figure 2*c***).

Modified Peptides

The biosynthetic machinery responsible for translation is highly complex, making the incorporation of new building blocks into peptides and proteins a challenging prospect. However, given the success of nonribsomal peptides (NRPs) as therapeutics (28), several groups have pursued strategies to translate RNA libraries into unnatural peptide libraries.

Several approaches exist for ribosomal translation with unnatural monomers. The most widely studied system involves nonsense suppression of an in vitro translation extract by a suppressor-tRNA bearing an unnatural amino acid (29). The strategy was used to carry out selections on mRNA display libraries containing the unnatural amino acid biocytin (30) (**Figure 3*a***). Because the approach is inherently limited to the incorporation of two or three different unnatural monomers (there are only three stop codons), it has little utility for in vitro selection of predominately unnatural peptide libraries. This issue has been addressed by introducing four-base codons for mRNA display (31). In that work, five unnatural amino acids (**Figure 3*a***), including biocytin, were included in 10^{12}-sized mRNA display libraries and subjected to selection for streptavidin binding, leading to enrichment of biocytin-containing sequences. The four-base codon experiments represent the most complex peptide selection libraries reported that utilize unnatural amino acids.

An alternative strategy for unnatural peptide synthesis involves the reassignment of sense codons to unnatural amino acids. Such reassignment requires that the natural aminoacyl-tRNAs not be present to compete with the unnatural ones. Using tRNA-depleted rabbit reticulocyte lysates and chemically charged tRNAs, Frankel and coworkers demonstrated the translation of short peptides containing *N*-methylated phenylalanine (**Figure 3*b***) in an mRNA display format (32). Although the strategy suffers from the preferential incorporation of trace contaminating natural aminoacyl-tRNA, it might represent a viable option for unnatural peptide selections if more extensive tRNA depletion methods are developed.

Alternatively, multiple sense codons can be reassigned in purified translation systems, where the concentration of all components is controlled. Forster and coworkers used a purified translation system with

TNA: (3′,2′)-α-L-threose nucleic acid

NRP: nonribsomal peptide

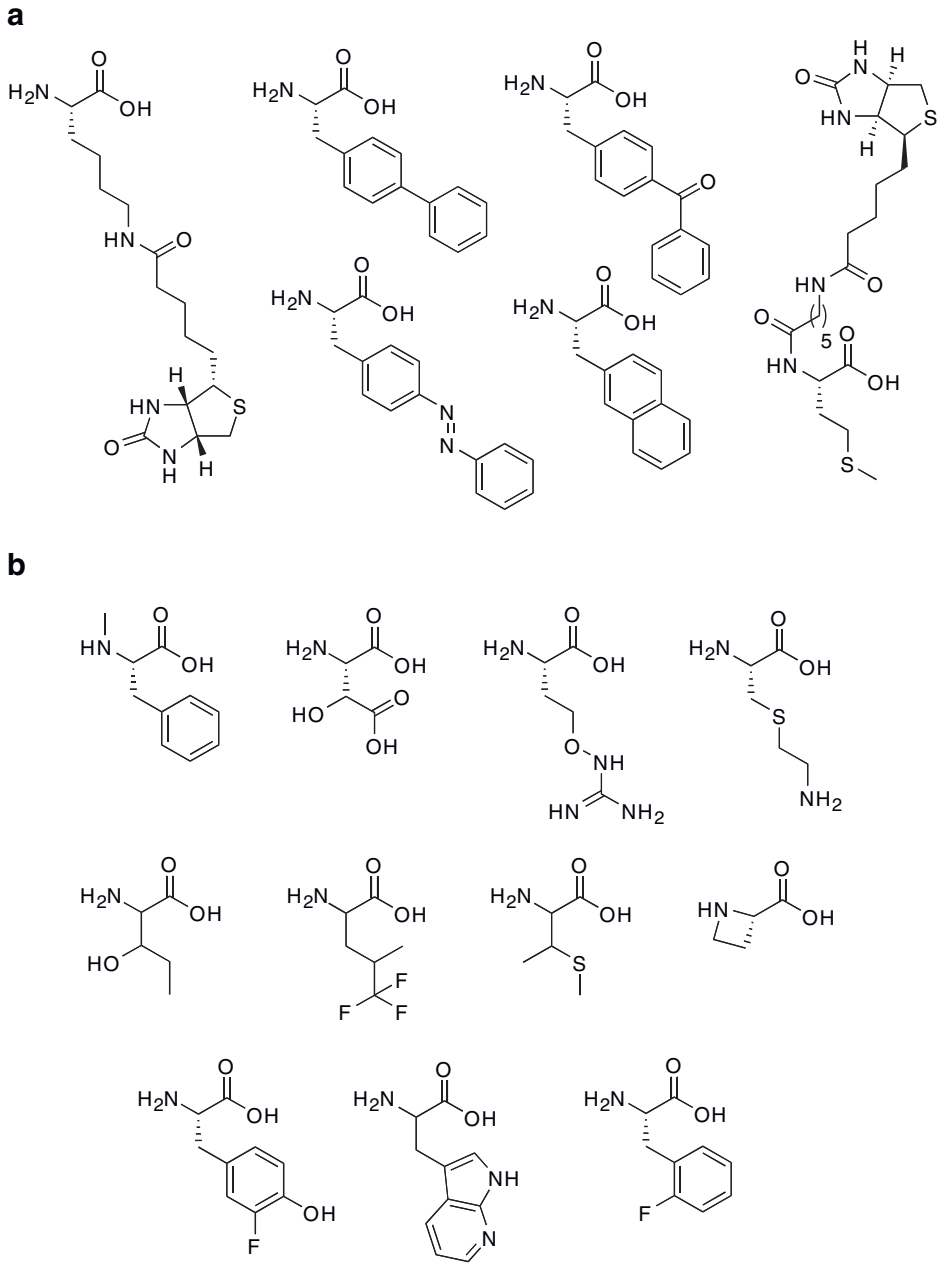


Figure 3

Modified peptides. (*a*) Examples of unnatural amino acids incorporated into in vitro selection libraries. (*b*) Examples of unnatural amino acids shown to be compatible with mRNA display.

chemoenzymatically synthesized aminoacyl-tRNAs to translate peptides containing up to three adjacent unnatural amino acids (33) and later offered "pure translational display," a relative of ribosome display, as a potential system for in vitro selection of unnatural peptides (34). The authors used this system to enrich *N*-biotinylated methionine-containing peptides (**Figure 3*a***). Merryman & Green carried out the bulk *N*-methylation of aminoacyl-tRNAs and used them in a purified translation system to

generate short *N*-methylated peptides (35). The result was striking because *N*-methylated peptides are more protease resistant than natural peptides, and frequently appear in NRPs.

In a recent exciting report, Szostak and coworkers reported the use of enzymatically charged unnatural aminoacyl-tRNAs and a purified translation system to translate peptides containing up to ten different unnatural amino acids (**Figure 3*b***) fused to their encoding mRNAs (36). Since then, the same authors described an assay for examining the promiscuity of aminoacyl-tRNA synthetases, and validated the enzymatic charging of more than 90 unnatural monomers onto tRNAs (37). Although a selection using the strategy has not yet been reported, it clearly holds great promise as a means for selecting and evolving wholly unnatural peptide libraries.

Perspective

Substantial progress has been made toward expanding the chemical make-up of enzymatically produced biopolymers. Once optimized, these methods have the potential to generate vast libraries of unnaturally functionalized nucleic acids and peptides. Given the proven utility of NRPs and aptamers such as pegaptanib, these libraries are likely to provide useful ligands, inhibitors, and potential drug leads.

From a chemical standpoint, the enzyme-based approaches have one intrinsic limitation: the substrate promiscuity of biosynthetic enzymes is bounded. Although a variety of new functional groups and side chains can be introduced, the chemical architecture of the polymer backbone will likely not differ drastically from the native precursor. Polyamide, polyester, and sugar phosphate-based polymers are not well suited for all applications, such as catalysis in organic solvents or use as orally bioavailable drugs. It will be interesting to see how far the biosynthetic enzymes can ultimately be pushed to produce heterologous materials.

IN VITRO SELECTION AND EVOLUTION OF SYNTHETIC COMPOUNDS

DTS: DNA-templated synthesis

ssDNA: single-stranded DNA

Within the past decade, evolution approaches have been proposed that are based on synthetic organic chemistry. Rather than relying on enzymes to translate nucleic acid genes into gene products, these methods exploit the sequence-specific hybridization of nucleic acids to direct a corresponding sequence of chemical reactions. Although DNA has long been recognized as a useful material for encoding information (38) and for templating chemistry (39), these two characteristics have only recently been exploited together. Two distinct yet complementary approaches, DNA display and DNA-templated synthesis (DTS), represent the major strategies reported thus far. The most significant results of each approach are discussed below.

DNA Display

In 2000, Halpin & Harbury offered a strategy for translating DNA libraries into combinatorial chemistry libraries of arbitrary chemical make-up (40). The concept, referred to as DNA display, exploits the hybridization properties of DNA to route single-stranded DNA (ssDNA) molecules bearing a synthetic handle through a split-and-pool combinatorial synthesis (**Figure 4**). Hybridization of DNA templates to oligonucleotide-coated solid supports acts to physically separate a DNA library into subpools based on sequence. The subpools are then subjected to distinct synthetic transformations with different chemical building blocks, covalently modifying the chemical handles on the DNA templates in a sequence-encoded manner. The splitting and chemistry are repeated until the entire DNA-template sequence is read and translated. The DNA sequence serves as a molecular recipe for the attached synthetic product, and acts to move the molecule along an encoded trajectory through a combinatorial synthetic tree. The final small molecule-DNA conjugates are

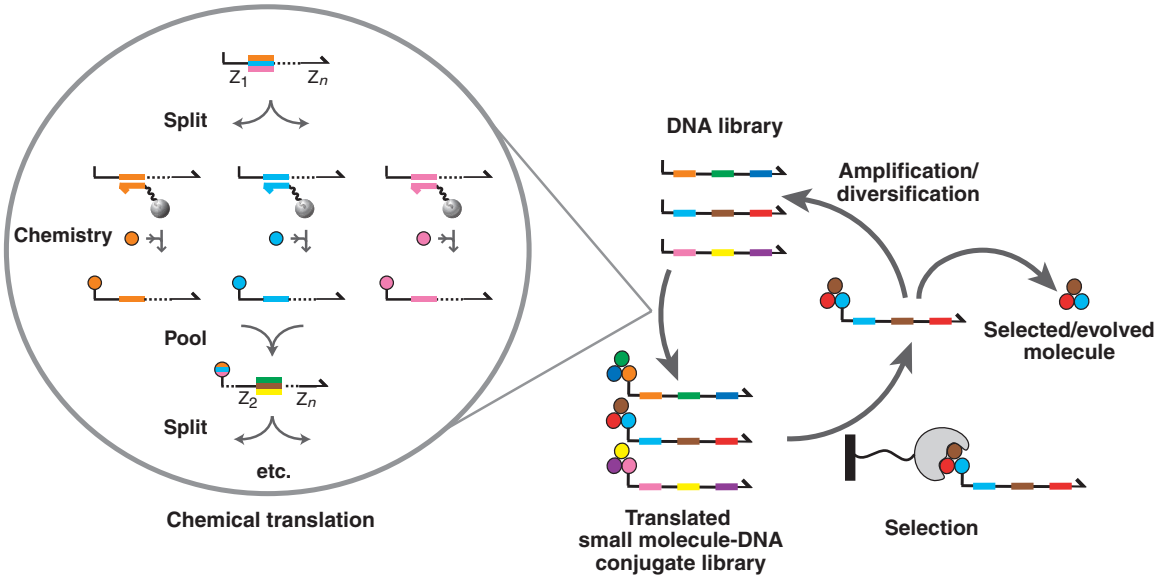


Figure 4

DNA display. A library of ssDNA molecules (*top center*) is chemically translated into small molecule-DNA conjugates. (*top left*) The DNA library is split into subpools by hybridization of 20-base codons to complementary oligonucleotide anticodons that are immobilized on separate columns (*orange*, *cyan*, *pink bars*). A distinct chemical transformation is carried out on each subpool, resulting in the covalent attachment of a chemical building block to the DNA (*orange*, *cyan*, *pink balls*). The library is pooled and then split based on the next coding region (*green*, *brown*, *yellow bars*), and distinct chemical transformations are carried out for each subpool. The process is iterated until the entire DNA sequence is read. Each codon can exist at only one coding region. The translated library is subjected to selection for a function of interest (binding to immobilized *gray widget*). The DNA linked to the small molecules that bind to the target is amplified, potentially diversified, and used as input for the subsequent round of chemical translation. The entire process is repeated until the library converges. Enriched molecules can be identified by DNA sequence and assayed for function as pure compounds.

subjected to selection, amplification, and diversification, as with standard in vitro selection systems. The matured genetic material is used to generate input for a subsequent round of chemical translation.

The DNA constructs used thus far in DNA display experiments incorporate multiple 20-base coding and noncoding regions, as well as a 5′ amine that serves as a chemical handle for synthesis (**Figure 4**) (41). At each coding region, one of several mutually exclusive sequences (codons) can exist. These sequences direct the hybridization-based routing during chemical translation. The noncoding regions that flank each coding region facilitate the initial library assembly and diversification between selection rounds. The number of coding regions, and the number of different codons present at each coding region, can be adjusted arbitrarily to accommodate a wide variety of library schemes. Generation of the ssDNA is carried out using a modified nucleic acid sequence-based amplification (NASBA) protocol (42) with 5′-modified primers.

The mechanics of sequence-encoded routing for chemical translation involves a series of column-to-column transfers, without solution manipulations of the library (41). Splitting into subpools is accomplished by flowing the library over a series of anticodon microcolumns bearing oligonucleotides complementary to the codons present at a single

coding region. The DNA hybridized to each anticodon microcolumn is then transferred to an anion exchange microcolumn, where the chemical transformation is carried out. These optimized routing steps give >90% yield per step, without measurable mishybridization.

Because DNA substrates are bound to columns during chemical steps, transformations are carried out in solid-phase synthesis format (43). The strategy provides great chemical flexibility. Organic and anhydrous solvents, in which DNA is insoluble, are suitable for chemical transformations. High-yielding reactions have been performed in neat methanol, ethanol, isopropanol, acetone, tetrahydrofuran, dioxane, acetonitrile, *N*,*N*-dimethylformamide, dimethylsulfoxide, ethyl acetate, and methylene chloride. In addition, a vast excess of reagents is used to drive reactions to completion. Excess reagents and byproducts are washed away upon completion of a coupling step.

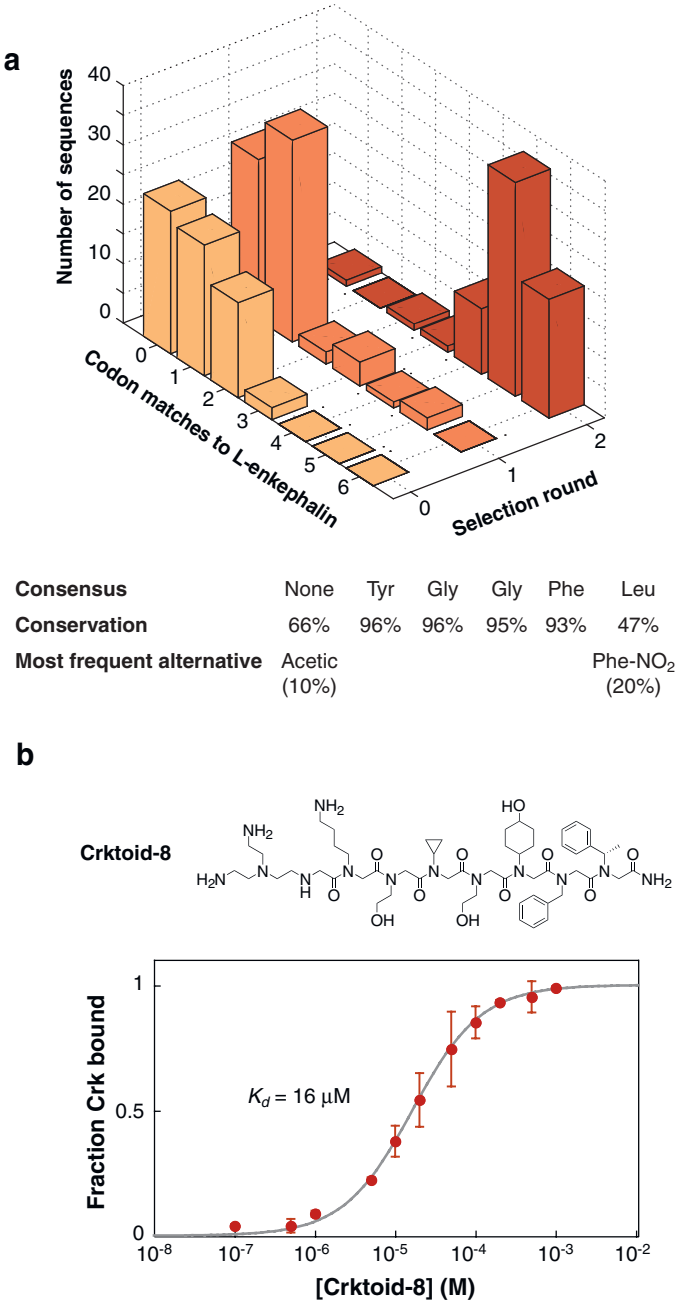
As an initial chemical system, DNA-compatible methods for peptide synthesis were developed and validated to give near-quantitative coupling for the 20 amino acids (side-chain protection strategies for carboxylic acids, amines, imidazoles, guanidinium groups, and thiols were described) (43). DNA-linked peptides up to 12 residues long were synthesized with high yield using these methods. The protocols are efficient and general enough to produce unnatural peptide libraries for selection. More recently, methods for the DNA-compatible synthesis of *N*-substituted glycines, or peptoids, have been developed and verified for approximately 30 different building blocks (S.J. Wrenn, R.M. Weisinger, D.R. Halpin & P.B. Harbury, submitted). Peptoids represent a protease-resistant and potentially druggable class of peptidomimetics. The DNA-directed synthesis of peptoids up to eight residues long has been accomplished. As with peptides, a wide variety of functionalized side chains is DNA-compatible. The types of chemical reactions known to proceed efficiently in the DNA display format include acylations of primary and secondary amines, nucleophilic displacements with amines and thiols, reductive aminations, electrophilic aromatic substitutions, periodate-based oxidations, and phosphine-based reductions.

In a critical experiment to validate the mechanics of DNA display, Halpin & Harbury designed a 10^6-member acylated pentapeptide library, and subjected it to a selection for binding to the 3-E7 antibody (44). Although the library was predominately unnatural, one of every million starting DNA sequences encoded the [Leu]-enkephalin peptide, which binds the 3-E7 antibody with seven nanomolar affinity. The experiment was a stringent test because six properly executed chemical translation steps were required for ligand synthesis, and the library was sufficiently large to require propagation through multiple generations before enrichment could be observed. When the experiment was conducted, sequencing of three library generations revealed a shift from a degenerate DNA distribution to one dominated by a consensus sequence encoding [Leu]-enkephalin (**Figure 5*a***). The result demonstrated that large synthetic libraries could be accessed with evolution-based methods using only common equipment and reagents. To put this into perspective, the library used in the Halpin & Harbury experiment was as complex as a typical compound collection at a pharmaceutical company, but was manipulated by a single individual without any automation.

The work of Halpin & Harbury did not address whether DNA-directed combinatorial-chemistry strategies could produce unknown binding motifs from a random library. Within the combinatorial chemistry community, the utility of screening random libraries, instead of libraries focused around a known ligand, is debated. It was not clear whether synthetic libraries would act like biopolymer libraries, in the sense that sufficiently large libraries would include ligands to essentially any target. To answer the question, Wrenn et al. synthesized a 10^8-member 8-mer peptoid library and subjected it to a

selection for binding to the N-terminal Src homology 3 domain of the proto-oncogene Crk (N-CrkSH3) (S.J. Wrenn, R.M. Weisinger, D.R. Halpin & P.B. Harbury, submitted). Over six library generations, the molecular population converged to a small number of sequence families. When the corresponding peptoids were re-synthesized without the appended DNA, and assayed for N-CrkSH3 binding, more than half proved to be N-CrkSH3 ligands with micromolar K_d values (**Figure 5*b***).

This result suggests that complex libraries will generally produce functional ligands to arbitrary targets. At a complexity of 10^8, this synthetic compound library is the largest that has been exhaustively explored to date, and it demonstrates that DNA-directed approaches can greatly increase access to the chemical universe. The ligand K_d values derived from the peptoid library are similar to those of ligands derived from equally complex phage display libraries targeted against SH3 domains. By analogy to peptide selections, larger compound libraries ($\sim 10^{13}$) should generate small-molecule ligands with nanomolar or better K_ds (6).



DNA-Templated Synthesis

The ability of DNA or RNA hybridization to promote chemical reactions by bringing nucleic acid-linked reagents into close proximity has been studied extensively (45). On the basis of this phenomenon, multiple groups have proposed that templated chemical polymerization of nucleotide-like units could

Figure 5

DNA display libraries. (*a*) Genetic material from each generation of the synthetic peptide library was subcloned, and roughly 70 full sequences were examined. For each sequence, the number of codon matches to the [Leu]-enkepahlin was recorded, and a histogram was generated for each round of selection. Over three generations, the library converged to the [Leu]-enkephalin sequence with good conservation. (*b*) Chemical structure and binding assay (intrinsic tryptophan fluorescence) of a molecule (Crktoid-8) that emerged from a 10^8-member library targeted against the N-terminal SH3 domain of the proto-oncoprotein Crk (NCrkSH3). Dissociation constant for the binding interaction is shown (K_d).

constitute a strategy for synthetic polymer evolution (46–49). This idea was generalized by Gartner & Liu, who demonstrated that reactions driven by DNA-induced proximity could mediate the translation of a DNA sequence into a covalently attached synthetic molecule (50). In the Gartner & Liu strategy, termed DNA-templated synthesis, single-stranded template DNA molecules bearing a reactive functional group are allowed to anneal with complementary oligonucleotides linked to a reagent. Hybridization increases the effective molarity of the two reactive species with respect to one another, accelerating a reaction between them that is effectively directed by the DNA sequence (**Figure 6**). Reacted templates are purified away from unreacted templates using a biotin handle present on the reagent oligonucleotide. The covalent link between the reagent oligonucleotide and the reaction product is finally broken. This process corresponds to the translation of one region of DNA sequence into one synthetic chemical unit, and it is iterated until an entire template sequence is "read." The translated compound-DNA conjugate library is then subjected to functional selection, amplification, and diversification.

Substantial progress has been made in developing DTS-compatible chemical transformations. The reactions reported thus far include thiol alkylation, amine acylation, reductive amination, nitro-aldol and nitro-Michael reactions, Wittig olefination, dipolar cycloaddition, Heck coupling, alkene-alkyne

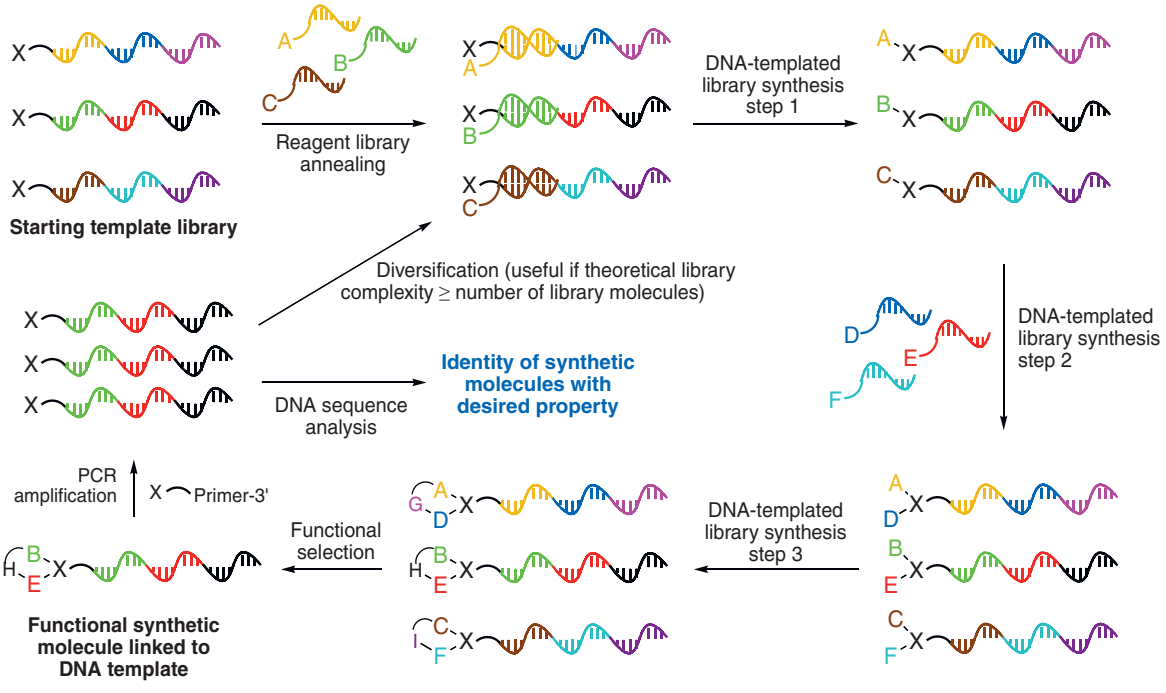


Figure 6

DNA-templated synthesis. Template DNA molecules bearing a synthetic handle are allowed to anneal with incoming reagent-linked oligonucleotides in one pot. Reacted species are purified, and reagent oligonucleotides, removed. The process is repeated until the entire template sequence is read. The translated library can be subjected to selection, followed by amplification and potentially diversification, in a manner similar to DNA display. Modified and reprinted with permission from (60). Copyright 2004 AAAS.

coupling, and phosphine-based reductions (50–54). This array of transformations is remarkable given the constraint that the reaction conditions must be compatible with DNA hybridization. The reported yields vary depending on the reaction type and the identity of the specific reactants. For example, Wittig olefinations proceed nearly quantitatively (>90%) in under two hours, whereas dipolar additions require long reaction times (~24 h) for reasonable product formation (15%–50%) (51).

A series of studies has been carried out to demonstrate the potential of DTS for library synthesis and selection. Because the DNA-linked reagents need only be present at nanomolar concentration, incompatible reactions can be performed in a single solution without cross-reactivity (55). Thus, library chemistry steps are amenable to a "one-pot" format. Multistep synthesis utilizing a variety of linkers and several reaction types has been demonstrated. The synthetic products include tripeptides, branched thioethers (56), a cyclic *N*-acyloxazolidine (**Figure 7*a***) (52), and a number of products derived from branching reaction pathways (57). The overall yields for these multistep syntheses have fallen between 1% and 10%. A clever strategy for ordered translation of a template in a single solution can reportedly increase yields for multistep syntheses (58).

Although no more than four sequential encoded chemical transformations have been reported, the DTS translation process can theoretically be applied to longer syntheses. However, the effective concentration of reagents hybridized at distal positions of a template may be insufficient to adequately accelerate intramolecular product formation. Indeed, the rates of some reactions, for example, reductive amination, are strongly dependent on the distance between the template and oligonucleotide reactants (59). Several alternate template architectures have been proposed to increase the number of reactions that can be encoded. An interesting loop-out "omega" architecture greatly improved the yields of distance-dependent reactions while utilizing template regions up to 20 bases away from the template reactive group.

As a proof-of-principle selection, a library of 65 fumaramide-containing cyclic peptides was synthesized using DTS (60). One of the starting DNA sequences encoded

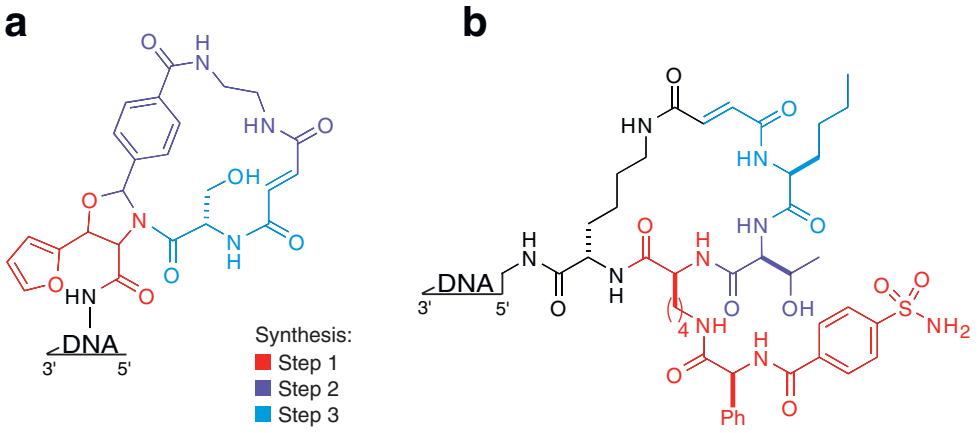


Figure 7

DTS products. (*a*) Cyclic *N*-acyloxazolidine synthesized in three templated steps. (*b*) The member of the cyclic-peptide DTS library designed to contain a known carbonic anhydrase binding residue (phenyl sulfonamide). Library members were synthesized in three steps, followed by Wittig-mediated ring closure (carbon-carbon double bond).

a synthetic product that included a known phenyl-sulfonamide ligand to carbonic anhydrase (**Figure 7*b***). This ligand-encoding template contained a unique restriction site, so its population in the library could be assessed by PCR and restriction digest. The synthesized library was subject to two serial affinity purifications for carbonic anhydrase binding, giving a roughly 200-fold enrichment of the ligand-encoding sequence. The experiment demonstrated that compound collections could be generated by DTS and enriched for ligands to a protein target. It remains to be seen how DTS will perform with large libraries ($\geq 10^6$).

Part of the beauty of DTS lies in its versatility as an information-encoding tool. In an impressive report, Kanan and coworkers applied DTS to the problem of reaction discovery (54). Because reacted templates can be easily purified away from template starting material, and because the template sequence encodes the reactant identities, DTS can be applied to assay a panel of potential reactants and conditions for covalent bond formation. In the work of Kanan et al., a palladium-catalyzed alkene-alkyne coupling reaction was discovered and then validated in a non-DNA-templated format. Such reactivity data can be read out using microarrays, with the potential to report the results of thousands of reactant combinations simultaneously.

Perspective

The fundamental distinction between DNA display and DTS is that the former relies on partition-based chemical translation, whereas the latter relies on proximity-based chemical translation. Each approach offers distinct advantages and suffers from inherent limitations. The partitioning strategy employs common off-the-shelf reagents for chemical transformations. This makes the technology modular and simple with respect to chemistry and allows the genetic code to be arbitrarily redefined at any point. Furthermore, the partitioning approach is compatible with synthetic environments where DNA is not soluble (organic/anhydrous solvents) and where DNA hybridization is not favorable (low salt, high temperature, high pH), providing a great deal of chemical flexibility. However, the partitioning-based chemical translation process can be quite laborious, as distinct chemical reactions must be carried out after each partition step. Proximity-induced reactivity enables one-pot chemical translation steps and the ability to purify reacted templates away from those that did not react. This simplifies library synthesis. In principle, proximity-based chemical translation also enables unique reactivity control, such as heterocoupling reactions with reagents that would typically favor homocoupling (61). On the other hand, the proximity method requires that reagents be linked to oligonucleotides via a removable linker, which is synthetically demanding and may preclude certain reaction types. Finally, the proximity-mediated reactions do not proceed efficiently in anhydrous solvents (62) and are not compatible with conditions that denature the DNA duplex. Because the partitioning and proximity approaches offer unique advantages, hybrid strategies might ultimately prove valuable.

In the coming years, we anticipate that DNA-directed combinatorial chemistry strategies will provide access to synthetic libraries as complex as large aptamer libraries ($\sim 10^{15}$), and we expect that the in vitro evolution of such libraries will become a widely used tool for general molecular discovery.

FUTURE DIRECTIONS

A number of advances will be required for chemical evolution to reach its full potential as a robust molecular discovery tool. However, the field is gaining momentum, and countless applications of chemical evolution are surely on the horizon. Below, we briefly discuss some of interesting directions and challenges that lie ahead.

Versatile Biosynthetic Enzymes: A Challenge for Protein Engineers

As our ability to engineer protein function improves, it will become possible to tailor enzymes for novel biosynthetic chemistries. Already, directed evolution efforts have produced mutant DNA polymerases with expanded substrate specificities relative to their parents (63, 64). In the area of ribosomal translation, engineered tRNAs and charging enzymes have been used to reprogram protein synthesis in bacterial cell culture, resulting in the production of proteins containing unnatural amino acids [reviewed in (65)]. Surely, the replication and translation machinery will be modified further to accommodate increasingly exotic biopolymer variants, and these materials will be evolved.

Another role for protein engineering will be to produce enzymes for the chemoenzymatic synthesis of small-molecule libraries. Enzymatic transformations are particularly attractive in the context of DNA-encoded combinatorial chemistry because they are generally carried out under gentle conditions. Moreover, the creation of stereocenters and carbon-carbon bonds is comparatively simple with enzymes. The use of biocatalysts in process chemistry continues to grow (66), and directed evolution is being used to tune the stereospecifity, regioselectivity, and catalytic efficiency of biocatalysts (67, 68). However, notwithstanding a few special exceptions, enzymes have not been used for the synthesis of small-molecule libraries. The problem is modularity (69). Few enzymes that transform a specific functional group are capable of acting on a broad spectrum of reactants containing that group. The challenge for chemoenzymatic combinatorial synthesis will be to engineer enzymes that are both modular and very specific. Good precedents for such enzymes (for example, penicillin G acylase) exist in the enzymatic protecting group literature (70).

One Quadrillion Small Molecules: A Challenge for Synthetic Organic Chemists

An important application for in vitro evolution of small molecules will be the discovery of compounds that perturb specific biological functions in animals. Presumably, such molecules will have to exhibit many of the properties that drugs do. Thus, a logical future direction is to construct DNA-directed combinatorial chemistry libraries that conform to accepted empirical specifications for druggability, such as the Lipinski rules of five (71). This future direction raises an interesting problem: Whereas typical combinatorial-chemistry libraries comprise ~10,000 compounds (72), in vitro evolution calls for libraries with complexities exceeding 10^{12} compounds. Few combinatorial synthetic schemes exist, even on paper, that can generate 10^{12} Lipinski-compatible molecules from commercially available building blocks. Such schemes are chemically demanding because the organic transformations used to build the library must be very tolerant of functional group variation within the reactants. The realization of complex small-molecule libraries will require chemical innovation and new DNA-compatible transformations. Chemoenzymatic synthesis could turn out to be an important part of the solution.

New Roles for "Molecule-Autonomous" Selections: A Challenge for Biochemists

All in vitro evolution techniques require selective propagation of gene products that possess a desired property. In general, the property must be manifested at the single-molecule level, and the selection process must decide whether to propagate or to discard each molecule as an individual (this makes it possible to test libraries as heterogeneous mixtures rather than as purified compounds). We refer to selections that discriminate based

on single-molecule properties as molecule autonomous. By contrast, the cell-based assays used in high-throughput screening are decidedly not molecule autonomous. Such assays score how multiple copies of a molecule collectively perturb a cell.

The prototypical molecule-autonomous selection is the affinity purification of gene products that bind to a matrix-immobilized target. A variant of this binding selection, based on isolating molecules with slow k_{off} rates, can efficiently enrich for gene products with maximum affinity (73). Alternatively, binding counterselections can be used to deplete gene products that bind promiscuously to off-targets (74). These counterselections optimize specificity and are particularly important when the relevant environment of the target contains closely related homologues. Binding selections can also be applied to heterogeneous targets that exist in their normal biological setting. For example, phage display in live mice has produced peptides that home variously to the vascular endothelium, to the heart, to the lung, to adipose tissue, and to the islets of Langerhans (75). More recently, Cerchia et al. reported the selection of RNA aptamers that bind to rat adrenal cells expressing the human RET receptor tyrosine kinase [the authors counterselected against aptamers that bound to cells lacking the receptor (76)]. Such whole-cell selections are applicable to targets whose biologically active forms are refractory to purification. Finally, single-turnover catalysts can be isolated based on binding to reactants and products that are tethered to the catalyst.

An important future direction will be to expand the range of properties that can be engineered using molecule-autonomous selections. In the case of small molecules intended for use in animals, it will be important to select for desirable pharmacological properties, such as good water solubility, membrane permeability, metabolic stability, and the absence of toxicity. Along these lines, some of the techniques developed for high-throughput screening can likely be adapted as molecule-autonomous selections. For example, the fractionation of small molecules over immobilized artificial membrane columns correlates with their membrane permeability in animal models (77, 78). This column fractionation can presumably be applied to libraries. Similarly, a small molecule's metabolic stability correlates with its resistance to modification when incubated with rat liver microsome preparations (77, 79). Digestion of translated libraries with microsomes between rounds of affinity selection would thus select for metabolically stable gene products. Ideally, application of novel pharmacological selections to complex libraries in the early stages of the molecular discovery process will decrease the likelihood of late-stage failure in animal studies.

Another future direction for molecule-autonomous selections involves the isolation of multiple-turnover catalysts from catalyst libraries. These selections present the unique problem of attributing reaction products to the catalyst that generated them. High-throughput sorting methods based on double-emulsions might provide the answer (80, 81), but significant technological advances will be necessary.

Automation and Throughput: Integration with Other Technologies

The integration of modern in vitro evolution approaches with other emerging technologies is an important future direction. One key advance will be the introduction of robotics. Many tedious steps in building-block preparation and in the translation of genes into gene products are amenable to automation. As the chemical alphabets used for in vitro evolution grow, the advantages of such automation will increase. Automation can also be used to multiplex the selection process so that binding motifs for many different targets can be evolved in parallel. The automated discovery of affinity reagents across a broad target spectrum will have important diagnostic

applications, particularly in conjunction with microarrays. Simultaneous interaction of several weak binding motifs with a target can produce multivalent binding that is both high affinity and high specificity (82). Evolution approaches are ideal for generating sets of noncompetitive binding motifs that recognize a common target.

High-throughput shotgun sequencing is another emerging technology that will affect in vitro evolution. The ability to sequence one million clones from an evolution experiment in a single afternoon will facilitate deeper analysis of the results. In the context of biopolymers, multiple sequence variants from a single family of solutions will allow the inference of three-dimensional structure (83). The relative abundance of different binding solutions should correlate with affinity. In the context of small-molecule evolution, the data will likely be useful for constructing quantitative structure-activity relationships. These relationships are used to characterize the nature of binding pockets on a target, and can be used as inputs for computational discovery methods.

Finally, it will be important to integrate small-molecule in vitro evolution approaches with more conventional high-throughput screening technology. The cell-based and animal-based assays used to judge the in vivo utility of small molecules will never conform to a molecule-autonomous selection format, and will have to be carried out in a screening format. Conversely, screening formats will never be able to access the vast chemical diversity available to in vitro selections. A sensible hybrid strategy will be first to affinity-mature complex libraries by in vitro evolution and read out the result by shotgun sequencing. This step will reduce the complexity of the molecular discovery problem from the order of 10^{15} possibilities to the order of one million, and will greatly enrich for molecules capable of binding to the target. After computational clustering of DNA sequences, a subset of the corresponding molecules can then be resynthesized robotically for high-throughput screening. The entire process could proceed without human intervention.

Prospectus

The notion of chemical evolution is still in its infancy, but the field has advanced rapidly. Enormous chemical libraries will be realized over the next decade. This unprecedented access to the uncharted reaches of chemical space will certainly bring exciting discoveries. As the molecular discovery process becomes increasingly accessible to independent research laboratories, areas as diverse as material science and medicine will reap the benefits.

SUMMARY POINTS

1. In vitro selection and evolution have become routine methods for isolating functional nucleic acids and proteins; efforts are under way to apply the same approaches to unnatural molecules.
2. A variety of modified nucleotides have been enzymatically incorporated into nucleic acid libraries for in vitro selections, yielding ligands and catalysts containing unnatural functional groups; a number of fully unnatural, polymerizable nucleic acid systems have also been introduced.
3. Using approaches such as nonsense suppression, four-base codons, tRNA depletion with sense codon reassignment, and purified translation systems, unnatural amino acids have been incorporated into mRNA display and ribosome display libraries.

4. Two complementary approaches for the sequence-directed translation of DNA into synthetic molecules have been developed as a means for in vitro selection and evolution of libraries of arbitrary chemical makeup; they are DNA display and DNA-templated synthesis (DTS).
5. DNA display employs a partition-based strategy for chemical translation; compound libraries with complexities as large as 100 million have been synthesized, subjected to selection for protein binding, and propagated over multiple generations, yielding novel synthetic ligands.
6. DTS employs a proximity-based strategy for chemical translation; a large variety of reaction types and complex synthetic structures are accessible, and the technology is suitable for library synthesis and in vitro selection.

FUTURE ISSUES

1. Protein engineering strategies should be applied to expand the chemical repertoire of biopolymer libraries and to produce enzymes for the chemoenzymatic synthesis of DNA-encoded compound libraries.
2. Combinatorial synthetic schemes should be developed that can generate complex libraries (10^{12}) of druggable compounds from commercially available building blocks.
3. Advanced molecule-autonomous selection strategies should be pursued to isolate molecules with useful attributes such as high-affinity binding, specific molecular recognition in complex environments, membrane permeability, metabolic stability, and multiple-turnover catalysis.
4. Emerging technologies such as automation, high-throughput sequencing, and high-throughput screening should be integrated into in vitro evolution technologies to efficiently and rapidly pinpoint and validate functional compound from large libraries.

ACKNOWLEDGMENTS

We thank David R. Liu for kindly providing **Figure 6** and Rebecca M. Weisinger for thoughtful comments on the manuscript. S.J.W. was partially supported by the Stanford University Medical Scientist Training Program. P.B.H. is supported by an N.I.H. Pioneer Award (NIH 1 DP1 0D000429).

LITERATURE CITED

1. Mills DR, Peterson RL, Spiegelman S. 1967. *Proc. Natl. Acad. Sci. USA* 58:217–24
2. Tuerk C, Gold L. 1990. *Science* 249:505–10
3. Ellington AD, Szostak JW. 1990. *Nature* 346:818–22
4. Joyce GF. 1989. *Gene* 82:83–87
5. Scott JK, Smith GP. 1990. *Science* 249:386–90
6. Wilson DS, Keefe AD, Szostak JW. 2001. *Proc. Natl. Acad. Sci. USA* 98:3750–55
7. Hanes J, Pluckthun A. 1997. *Proc. Natl. Acad. Sci. USA* 94:4937–42
8. Mattheakis LC, Bhatt RR, Dower WJ. 1994. *Proc. Natl. Acad. Sci. USA* 91:9022–26

9. Nemoto N, Miyamoto-Sato E, Husimi Y, Yanagawa H. 1997. *FEBS Lett.* 414:405–8
10. Roberts RW, Ja WW. 1999. *Curr. Opin. Struct. Biol.* 9:521–29
11. Kool ET. 2002. *Acc. Chem. Res.* 35:936–43
12. Bittker JA, Phillips KJ, Liu DR. 2002. *Curr. Opin. Chem. Biol.* 6:367–74
13. Tarasow TM, Eaton BE. 1998. *Biopolymers* 48:29–37
14. Latham JA, Johnson R, Toole JJ. 1994. *Nucleic Acids Res.* 22:2817–22
15. Sakthivel K, Barbas CF. 1998. *Angew. Chem. Int. Ed. Engl.* 37:2872–75
16. Battersby TR, Ang DN, Burgstaller P, Jurczyk SC, Bowser MT, et al. 1999. *J. Am. Chem. Soc.* 121:9781–89
17. Green LS, Jellinek D, Bell C, Beebe LA, Feistner BD, et al. 1995. *Chem. Biol.* 2:683–95
18. Jellinek D, Green LS, Bell C, Lynott CK, Gill N, et al. 1995. *Biochemistry* 34:11363–72
19. Lin Y, Qiu Q, Gill SC, Jayasena SD. 1994. *Nucleic Acids Res.* 22:5229–34
20. Ruckman J, Green LS, Beeson J, Waugh S, Gillette WL, et al. 1998. *J. Biol. Chem.* 273:20556–67
21. Ng EW, Shima DT, Calias P, Cunningham ETJ, Guyer DR, Adamis AP. 2006. *Nat. Rev. Drug Discov.* 5:123–32
22. Tarasow TM, Tarasow SL, Eaton BE. 1997. *Nature* 389:54–57
23. Seelig B, Jaschke A. 1999. *Chem. Biol.* 6:167–76
24. Santoro SW, Joyce GF, Sakthivel K, Gramatikova S, Barbas CF. 2000. *J. Am. Chem. Soc.* 122:2433–39
25. Perrin DM, Garestier T, Helene C. 2001. *J. Am. Chem. Soc.* 123:1556–63
26. Thum O, Jager S, Famulok M. 2001. *Angew. Chem. Int. Ed. Engl.* 40:3990–93
27. Ichida JK, Zou K, Horhota A, Yu B, McLaughlin LW, Szostak JW. 2005. *J. Am. Chem. Soc.* 127:2802–3
28. Sieber SA, Marahiel MA. 2003. *J. Bacteriol.* 185:7036–43
29. Noren CJ, Anthony-Cahill SJ, Griffith MC, Schultz PG. 1989. *Science* 244:182–88
30. Li SW, Millward S, Roberts R. 2002. *J. Am. Chem. Soc.* 124:9972–73
31. Muranaka N, Hohsaka T, Sisido M. 2006. *Nucleic Acids Res.* 34:e7
32. Frankel A, Millward SW, Roberts RW. 2003. *Chem. Biol.* 10:1043–50
33. Forster AC, Tan ZP, Nalam MN, Lin HN, Qu H, et al. 2003. *Proc. Natl. Acad. Sci. USA* 100:6353–57
34. Forster AC, Cornish VW, Blacklow SC. 2004. *Anal. Biochem.* 333:358–64
35. Merryman C, Green R. 2004. *Chem. Biol.* 11:575–82
36. Josephson K, Hartman MC, Szostak JW. 2005. *J. Am. Chem. Soc.* 127:11727–35
37. Hartman MC, Josephson K, Szostak JW. 2006. *Proc. Natl. Acad. Sci. USA* 103:4356–61
38. Brenner S, Lerner RA. 1992. *Proc. Natl. Acad. Sci. USA* 89:5381–83
39. Orgel LE. 1995. *Acc. Chem. Res.* 28:109–18
40. Halpin DR, Harbury PB. 2000. *U.S. Patent Application WO 99-US24494*
41. Halpin DR, Harbury PB. 2004. *PLoS Biol.* 2:E173
42. Compton J. 1991. *Nature* 350:91–92
43. Halpin DR, Lee JA, Wrenn SJ, Harbury PB. 2004. *PLoS Biol.* 2:E175
44. Halpin DR, Harbury PB. 2004. *PLoS Biol.* 2:E174
45. Calderone CT, Liu DR. 2004. *Curr. Opin. Chem. Biol.* 8:645–53
46. Gat Y, Lynn DG. 1998. *Biopolymers* 48:19–28
47. Kozlov IA, De Bouvere B, Van Aerschot A, Herdewijn P, Orgel LE. 1999. *J. Am. Chem. Soc.* 121:5856–59
48. Li X, Zhan ZY, Knipe R, Lynn DG. 2002. *J. Am. Chem. Soc.* 124:746–47
49. Rosenbaum DM, Liu DR. 2003. *J. Am. Chem. Soc.* 125:13924–25

50. Gartner ZJ, Liu DR. 2001. *J. Am. Chem. Soc.* 123:6961–63
51. Gartner ZJ, Kanan MW, Liu DR. 2002. *Angew. Chem. Int. Ed. Engl.* 41:1796–800
52. Li X, Gartner ZJ, Tse BN, Liu DR. 2004. *J. Am. Chem. Soc.* 126:5090–92
53. Sakurai K, Snyder TM, Liu DR. 2005. *J. Am. Chem. Soc.* 127:1660–61
54. Kanan MW, Rozenman MM, Sakurai K, Snyder TM, Liu DR. 2004. *Nature* 431:545–49
55. Calderone CT, Puckett JW, Gartner ZJ, Liu DR. 2002. *Angew. Chem. Int. Ed. Engl.* 41:4104–8
56. Gartner ZJ, Kanan MW, Liu DR. 2002. *J. Am. Chem. Soc.* 124:10304–6
57. Calderone CT, Liu DR. 2005. *Angew. Chem. Int. Ed. Engl.* 44:7383–86
58. Snyder TM, Liu DR. 2005. *Angew. Chem. Int. Ed. Engl.* 44:7379–82
59. Gartner ZJ, Grubina R, Calderone CT, Liu DR. 2003. *Angew. Chem. Int. Ed. Engl.* 42:1370–75
60. Gartner ZJ, Tse BN, Grubina R, Doyon JB, Snyder TM, Liu DR. 2004. *Science* 305:1601–5
61. Liu DR. 2004. *PLoS Biol.* 2:E223
62. Rozenman MM, Liu DR. 2006. *ChemBioChem* 7:253–56
63. Chelliserrykattil J, Ellington AD. 2004. *Nat. Biotechnol.* 22:1155–60
64. Xia G, Chen LJ, Sera T, Fa M, Schultz PG, Romesberg FE. 2002. *Proc. Natl. Acad. Sci. USA* 99:6597–602
65. Xie JM, Schultz PG. 2005. *Methods* 36:227–38
66. Azerad R. 2005. *Adv. Org. Synth.* 1:455–518
67. Hoffmeister D, Yang J, Liu L, Thorson JS. 2003. *Proc. Natl. Acad. Sci. USA* 100:13184–89
68. Reetz MT. 2002. *Tetrahedron* 58:6595–602
69. Khosla C, Harbury PB. 2001. *Nature* 409:247–52
70. Kadereit D, Waldmann H. 2001. *Chem. Rev.* 101:3367–96
71. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. 2001. *Adv. Drug Deliv. Rev.* 46:3–26
72. Thompson LA, Ellman JA. 1996. *Chem. Rev.* 96:555–600
73. Jermutus L, Honegger A, Schwesinger F, Hanes J, Pluckthun A. 2001. *Proc. Natl. Acad. Sci. USA* 98:75–80
74. Geiger A, Burgstaller P, von der Eltz H, Roeder A, Famulok M. 1996. *Nucleic Acids Res.* 24:1029–36
75. Ruoslahti E. 2004. *Biochem. Soc. Trans.* 32:397–402
76. Cerchia L, Duconge F, Pestourie C, Boulay J, Aissouni Y, et al. 2005. *PLoS Biol.* 3:e123
77. Kerns EH. 2001. *J. Pharm. Sci.* 90:1838–58
78. Pidgeon C, Ong S, Liu H, Qiu X, Pidgeon M, et al. 1995. *J. Med. Chem.* 38:590–94
79. Di L, Kerns EH, Gao N, Li SQ, Huang Y, et al. 2004. *J. Pharm. Sci.* 93:1537–44
80. Aharoni A, Amitai G, Bernath K, Magdassi S, Tawfik DS. 2005. *Chem. Biol.* 12:1281–89
81. Mastrobattista E, Taly V, Chanudet E, Treacy P, Kelly BT, Griffiths AD. 2005. *Chem. Biol.* 12:1291–300
82. Reddy MM, Bachhawat-Sikder K, Kodadek T. 2004. *Chem. Biol.* 11:1127–37
83. Ekland EH, Szostak JW, Bartel DP. 1995. *Science* 269:364–70

Annual Review of Biochemistry

Volume 76, 2007

Contents

Mitochondrial Theme

Indexes

Errata

An online log of corrections to *Annual Review of Biochemistry* chapters (if any, 1997 to the present) may be found at http://biochem.annualreviews.org/errata.shtml