E Volutionary Principles Applied to the Discovery and Study of Functional Biological and Synthetic Molecules

Professor David R. Liu, Department of Chemistry and Chemical Biology, Harvard University
drliu@fas.harvard.edu (617) 496-1067 http://evolve.harvard.edu

Introduction

Two of the most common challenges in chemistry and chemical biology are (i) how to control chemical reactivity and (ii) how to discover molecules with desired properties from many possible structures. My laboratory has initiated a program to develop a new approach to addressing these two challenges that differs fundamentally from the approaches most frequently taken by chemists. Our approach is based on nature's effective molarity-based control of reactivity and nature's evolution-based discovery of functional molecules, both of which offer advantages compared with traditional approaches to synthesis and discovery. Our group uses the principles of biological evolution in conjunction with synthetic organic chemistry and molecular biology to discover and study small molecules, macromolecules, and new chemical reactions.

DNA-Templated Organic Synthesis for the Creation and Discovery of Synthetic Small Molecules and Synthetic Polymers

We recently discovered that DNA duplex formation exerts remarkable control over the effective molarity of DNA-linked reactants without requiring structural mimicry of the DNA backbone (1, 5, 16). As a result, DNA-templated organic synthesis (DTS) is a surprisingly general phenomenon that can direct a wide range of chemical reactions including carbon-carbon bond forming reactions and organometallic coupling reactions, even if the structures of the reactants or products do not resemble the DNA backbone (1, 5, 7-9, 11, 13, 14, 16). For many DNA-templated reactions, products form efficiently even when reactive groups are separated by large distances on the template ("distance-independent synthesis") (1, 5, 7, 9, 14, 16). DTS is sufficiently sequence-specific that a single DNA-linked template can react primarily with its sequence-programmed partner reagent in a single solution containing a 1,000-fold excess of non-partner, sequence-mismatched reagents (1).
Since our initial findings, we have developed a suite of linker and purification strategies that have enabled DNA templates to be translated into small-molecule products of multistep DNA-templated syntheses. For example, a 5'-amine-terminated DNA template was sequence-specifically translated into a non-natural tripeptide or a branched thioether product through three DNA-templated steps, with each step encoded by a different region of a 30-base DNA template (7). More recently, we translated DNA templates into N-acyloxazolidines and macrocyclic N-acyloxazolidines using multistep DTS (14). To date multistep DTS has been used to create more than a dozen classes of diverse organic small-molecule structures.

In addition to exploring and expanding the synthetic capabilities of DTS, we have also shown that DNA-templated synthesis enables new modes of controlling reactivity that are not possible using current synthetic approaches. For example, in a DNA-templated format many starting materials can undergo multiple otherwise incompatible reaction types in a single solution to generate exclusively a set of sequence-programmed products, while the analogous experiment in a traditional reaction format would generate an uncontrolled mixture of all possible products (8). This reaction mode has been used to diversify synthetic small molecule libraries using iterated branching reaction pathways in a single solution (28) in contrast to the more common diversification approach of using different building blocks in one type of reaction. In addition, we have also found that DTS enables heterocoupling reactions to take place efficiently between reactants that preferentially homocouple in a conventional synthesis format (20), and also allows multistep ordered small-molecule synthesis to take place in one pot between multiple reactants of comparable reactivity. For example, a orchestrated series of changes in template secondary structure was used to synthesize an ordered triolefin or an ordered tripeptide in a single solution in which all reactants are initially present (27). If combined under conventional synthesis conditions, these reactants would instead generate a vast mixture of predominantly non-ordered products.

Our other advances in this area include the development of two new template architectures that expand the synthetic capabilities of DTS by (i)
allowing virtually any DNA-template reaction to be encoded by any region of a DNA template, and (ii) by enabling two reactions to take place on a single DNA template in one step (9). The use of both of these architectures together with more recently developed DNA-template synthetic reactions proved crucial in the DNA-template N-acyloxazolidines syntheses mentioned above (14). In addition, we discovered the ability of a DNA template to induce stereoselectivity in a DNA-template reaction that generates products unrelated to the DNA backbone, and have traced the origins of this stereoselectivity to the macromolecular conformation of the templates (11). We used this stereoselectivity as a sensitive measure of the conditions under which the DNA templates can directly influence a reaction beyond simple modulation of the effective molarity of the reactants, and found that even a small number of rotatable bonds abrogates observed template-induced effects.

In some cases it may not be possible or convenient to link reactants to oligonucleotides. To develop an alternative strategy for DNA-programmed synthesis that enables the participation of non-DNA-linked reagents, we developed DNA-template functional group transformations that convert azide groups into amines, thiols, or carboxylic acids in a sequence-specific manner. These functional group transformations were used in conjunction with four non-DNA-tethered electrophilic reagents to convert four template-linked azides in a single solution into four sequence-programmed sulfonamide, carbamate, urea, and thiourea products (24).

We have also developed highly sensitive *in vitro* selections for DNA-linked synthetic small molecules (such as the products of DNA-template library synthesis) with protein binding affinity and specificity (12). These selections can be iterated to achieve enormous enrichments for functional DNA-linked synthetic small molecules.

Integrating many of the above concepts, we recently translated a library of 65 DNA templates into a pilot library of complex synthetic small-molecule macrocycles using a "genetic code" that dictates which reactants are recruited by each 10-base coding sequence. The resulting library of DNA-linked macrocycles was selected for binding to a target protein, and the DNA templates encoding macrocycles with target protein affinity were amplified by PCR and characterized by DNA sequencing (18). A single template of the pilot library that encodes a synthetic macrocycle with affinity for the target
protein was successfully enriched in this manner. This work represents the translation, selection, and amplification of a library of DNA sequences that encode synthetic small molecules, rather than proteins (18). Encouraged by these developments, we and others are now applying this approach to small molecule synthesis and discovery on libraries of much larger complexities and structural diversities. For example, we are currently characterizing a DNA-templated library of 1,000 N-acyloxazolidine heterocycles for use in functional selections.

We have begun to apply these principles to synthetic polymers in addition to small molecules. Based on the distance dependence of DNA-templated reductive amination and on the previous findings of David Lynn and co-workers, we have translated DNA templates into synthetic sequence-defined peptide nucleic acid (PNA) polymers using DNA-templated polymerization of PNA aldehydes (13). This polymerization proceeds with remarkable efficiency, excellent sequence-specificity, and can generate synthetic polymers of length similar to that of proteins and nucleic acids known to possess functional binding or catalytic properties (13). These findings are the basis of our ongoing efforts to evolve sequence-defined synthetic heteropolymers through processes of translation, selection, amplification, and diversification previously available only to natural biopolymers.

This novel approach to creating and discovering functional molecules offers significant advantages compared with existing methods. DNA-templated libraries of synthetic molecules can be subjected to true in vitro selections (as opposed to screens) for desired binding or catalytic activities, obviating the need to spatially separate each library member or to spend effort characterizing uninteresting molecules. Only minute quantities of material (~1,000 molecules of each different library member) are required for these selections because the information that directs each member's synthesis can be amplified by PCR; indeed the syntheses and selections described above were typically executed on a nanomole to sub-femtomole scale. The small amount of material required coupled with the suitability of these molecules to undergo selection in theory enables libraries of unprecedented complexity (much larger than the current...
A New Approach to Reaction Discovery

Unique features of DNA-templated organic synthesis have also led to a new approach for the discovery of bond-forming chemical reactions (20). In contrast with traditional reaction discovery methods, our approach does not focus on a specific combination of substrates or on the formation of one type of product structure. Instead, we combine pools of many DNA-linked substrates in one solution and select all possible pairwise combinations of substrates simultaneously for bond-forming combinations in a single experiment. The identity of bond-forming reactant pairs is revealed by exposing DNA sequences that survive the in vitro selection to DNA microarrays containing sequences that represent every possible combination of substrates. Because the results of this reaction discovery selection can be amplified by PCR, we perform this process on a femtomole scale that is unprecedented for reaction discovery.

We validated this approach to reaction discovery by "rediscovering" several known reactions mediated by transition metals or organic reagents. We have since used this system in a 96- and 168-reaction matrix format to discover several new transition metal-catalyzed bond-forming reactions that have been confirmed in a DNA-templated format. One of the discovered reactions, a carbon-carbon bond-forming macrocyclization between a simple alkyne and alkene mediated by catalytic quantities of Pd(II) in neutral water or mixed organic solvent at room temperature to form a macrocyclic trans-enone in high yield, has also been confirmed by extensive characterization in a non-DNA-templated, conventional synthesis format (20). Our exploration of this discovered enone-forming reaction has recently led to its successful use in an intermolecular (rather than macrocyclization) format. This approach enables a broad and unbiased search of functional group space for new reactions at a rate of thousands of combinations of reactants and reaction conditions per two-day experiment.

The development of these new areas lies at the heart of merging the creativity of the chemist with the powerful principles underlying the evolution of living systems (16).

Expanding the Scope of Protein and Nucleic Acid Evolution

We are also interested in developing and applying new methods for evolving biological macromolecules. We developed a new method for
diversifying nucleic acid libraries by nonhomologous random recombination (NRR), and used this new method to evolve DNA aptamers with significantly higher affinities than those evolved using error-prone PCR under identical selection conditions (4, 6). NRR has also proven to be a valuable strategy for minimizing a functional nucleic acid and for rapidly identifying structure-function relationships among evolved nucleic acids (6). We recently developed a modified version of NRR that enables protein evolution to access structures not previously accessible using existing protein evolution methods. The functional requirements of chorismate mutase, a natural protein enzyme, were explored in a broad and unbiased manner using protein NRR. Functional chorismate mutases emerging from protein NRR-diversified libraries included enzymes containing major rearrangements of secondary structural elements (15). When applied to the study of natural functional RNAs or DNA such as sRNA translational regulators or promoter elements that render gene expression dependent on specific cellular responses, NRR coupled with selection has served as an efficient and comprehensive method for revealing essential and non-essential elements (25).

We have also developed methods for evolving functional RNA molecules in vivo from random RNA libraries. For example, we evolved from random RNA sequences protein-RNA complexes capable of activating gene transcription to a degree comparable that of the most potent natural protein transcriptional activators (10). We have used site-directed mutagenesis in conjunction with chemical studies to characterize the nature of these artificially evolved RNA-protein transcription factors (10). The ability of RNA to be efficiently engineered and evolved enabled us to develop a variant of the evolved RNA transcriptional activator that is dependent on the presence of a cell-permeable synthetic small molecule (19). In other nucleic acid evolution studies, we seek to define the scope of biopolymeric catalysis and expand our ability to generate protein- and RNA-based catalysts by the stepwise evolution of new ribozymes from existing protein enzymes ("biopolymeric alchemy").

Our protein evolution efforts have focused on the evolution of inteins (protein splicing domains) that are only active in the presence of a cell-permeable synthetic small molecule (17, 33), the evolution of nucleases with tailor-made DNA cleavage specificities (31), the evolution of protein-protein interfaces, and the evolution of non-ribosomal peptide synthetases (32) with altered activities. In all of these cases, we have successfully linked protein activity both positively and negatively with the survival of a bacterial cell (3). Coupled
positive and negative in vivo selections enabled us to evolve evolve homing endonuclease with altered (rather than merely broadened) substrate specificity (31).

We have integrated a variety of directed evolution methods to successfully evolve inteins that undergo protein splicing only in the presence of a cell-permeable synthetic small molecule (called 4-HT), and have shown that these inteins enable the function of several unrelated proteins to be controlled in living cells in a rapid, dose-dependent, and post-translational manner (17). Recently we demonstrated in collaboration with Andrew McMahon’s group that these evolved inteins render the localization and activity of Gli1 and Gli3, two transcription factors involved in embryogenesis, dependent on 4-HT in living mammalian cells (33). Mammalian cells expressing evolved intein-inserted Gli1 differentiate into osteoblasts in a 4-HT-dependent manner (33), demonstrating that a complex biological signaling pathway can be regulated by a small molecule using the evolved intein. Small-molecule-activated inteins may serve as powerful and general tools for rapidly perturbing virtually any protein’s activity in living cells with a degree of temporal control, spatial resolution, and dose-dependence not achievable by intervening at the DNA or RNA levels. Although the use of evolved ligand-activated inteins requires genetic intervention, this approach does not require the discovery of a specific small molecule modulator for each protein of interest and therefore may serve as a complementary approach to chemical genetics.

References


3. "An In Vivo Selection System for Homing Endonuclease Activity" Gruen, M.; Chang, K.; Serbanescu, I.; Liu, D. R. Nucleic Acids Research 30, e29 (2002). This paper describes a selection system that is the basis of ongoing efforts to evolve and understand homing endonucleases with novel substrate specificities.


6. "Nucleic Acid Evolution and Minimization by Nonhomologous Random Recombination" Bittker, J. A.; Le, B. V.; Liu, D. R. *Nature Biotechnology* 20, 1024-1029 (2002). This work describes a new method (NRR) for the diversification of nucleic acids that allows random recombination to take place between multiple unrelated DNA sequences, and demonstrates that NRR enables nucleic acid evolution, functional dissection, and minimization in ways not possible using existing diversification methods.


10. "In Vivo Evolution of an RNA Transcriptional Activator" Buskirk, A. R.; Kehayova, P. D.; Landrigan, A.; Liu, D.R. *Chem. Biol.* 10, 533-540 (2003). This paper is noted as an article of interest in *Cell* 113 (7) and is previewed in a separate article in *Chem. Biol.* 10, 584-585 (2003). This work describes the evolution from random RNA libraries expressed in yeast cells of a short RNA sequence that strongly activates gene transcription. The potency of this RNA rivals that of the most active known natural transcriptional activators (all of which are proteins). Its function was dissected using a combination of site-directed mutagenesis, additional evolution, and secondary structure prediction.

11. "Stereoselectivity in DNA-Templated Organic Synthesis and Its Origins" Li, X. and Liu, D. R. *J. Am. Chem. Soc.* 125, 10188-10189 (2003). This paper reports the ability of DNA templates to induce stereoselectivity in DNA-templated reactions and traces the origins of this stereoselectivity to the macromolecular conformation of the DNA templates. These findings reveal the conditions under which DNA templates can perturb reaction outcomes beyond simple modulation of effective molarity.

12. "Highly Sensitive In Vitro Selections for DNA-Linked Synthetic Small Molecules with Protein Binding Affinity and Specificity" Doyon, J. B.; Snyder, T. M.; Liu, D. R. *J. Am. Chem. Soc.* 125, 12372-12373 (2003). This work describes the development of affinity and specificity selections for DNA-linked synthetic molecules such as those emerging from DNA-templated small molecule library synthesis.


14. “Translation of DNA into Synthetic N-Acyloxazolidines” Li, X.; Gartner, Z. J.; Tse, B. N.; Liu, D. R. *J. Am. Chem. Soc.* 126, 5090-5092 (2004). Multistep DNA-templated synthesis was used to generate N-acyloxazolidines and macrocyclic N-acyloxazolidines. These structures represent the most complex synthetic small molecules to date translated from DNA.

random recombination (protein NRR) to evolve active chorismate mutase protein enzymes with rearranged secondary structures are described in this work.


17. “Directed Evolution of Ligand Dependence: Small Molecule-Dependent Protein Splicing” Buskirk, A. R.; Ong, Y.-C.; Gartner, Z. J.; Liu, D. R. Proc. Natl. Acad. Sci. USA 101, 10505-10510 (2004). The laboratory evolution of an intein that undergoes splicing only in the presence of a cell-permeable synthetic small molecule is described in this work. The evolved intein was used in living cells to render the function of four unrelated proteins dependent on the small molecule in a post-translational and dose-dependent manner.


