J|A|C|S

A R T I C L E S

Development and Initial Application of a Hybridization-Independent, DNA-Encoded Reaction Discovery System Compatible with Organic Solvents

Mary M. Rozenman, Matthew W. Kanan, and David R. Liu*

Contribution from the Howard Hughes Medical Institute and the Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138

Received June 7, 2007; E-mail: drliu@fas.harvard.edu

Abstract: We have developed and applied an approach to reaction discovery that takes advantage of DNA encoding, DNA-programmed assembly of substrate pairs, in vitro selection, and PCR amplification, yet does not require reaction conditions that support DNA hybridization. This system allows the simultaneous evaluation of >200 potential bond-forming combinations of substrates in a single experiment and can be applied in a range of solvent and temperature conditions. In an initial application, we applied this system to explore Au(III)-mediated chemistry and uncovered a simple, mild method for the selective Markovnikov-type hydroarylation of vinyl arenes and trisubstituted olefins with indoles.

Introduction

New functional molecules emerge in nature through iterated cycles of translation, selection, amplification, and diversification of genetic material. In the laboratory, researchers can carry out the same evolutionary process in a directed fashion to access molecules possessing desired properties. Evolution-based approaches have primarily been applied to the discovery of biomolecules with a broad range of function.[1,2] More recently, the techniques of molecular evolution have been applied to problems in the chemical sciences, including the synthesis and discovery of functional small molecules and the discovery of new chemical reactions.[3,4] Such approaches can be particularly powerful because they are compatible with a selection, a process that simultaneously evaluates all members of an arbitrarily large population of molecules and separates functional molecules from inactive variants. Selections can be more efficient than conventional screens in which molecules or reactions are individually evaluated in a low- or high-throughput manner because selections allow for en masse evaluation without requiring spatial separation of candidate molecules. Selections have proven especially effective when the molecules under selection are associated with nucleic acids that encode each molecule's identity[4] because nucleic acids can be readily amplified and decoded. As a result, selections carried out on nucleic acids or nucleic acid−small molecule conjugates require only minute quantities of material (typically, sub-nanomolar) and can be iterated to multiply their net effectiveness.[5]

We recently implemented a selection-based approach to the discovery of bond-forming reactions.[6] Reaction discovery is a central endeavor in chemistry because it provides new tools for chemical synthesis, facilitates the discovery of functional synthetic molecules, and can reveal new principles of reactivity when coupled with mechanistic investigation. Most methods for reaction discovery search for conditions that enable a specific desired product structure to be formed from potential precursors. Our approach is complementary because it does not focus on one particular product, but instead simultaneously evaluates bond formation between any two members of a large collection of substrates under one of many different reaction conditions. Because this approach uses a selection for bond formation that is independent of substrate or product structure, it does not rely on specific reactivity predictions and enables a broad search for reactivity among a range of substrates.[7]

Our first-generation reaction discovery system (Figure 1a)[6] used DNA hybridization to organize many potential bond-forming substrate combinations into discrete pairs. Following the exposure of DNA-duplex localized reactants to a given set of reaction conditions, DNA-templated bond formation between substrates in each reactive pair induced the transfer of a biotin group to the DNA strand encoding those two substrates. In vitro selection using immobilized streptavidin, PCR amplification, and DNA microarray analysis subsequently revealed the identities of reactive substrate pairs. This system uncovered a mild and efficient Pd(II)-mediated coupling reaction between alkynamides and alkenes to generate trans-α,β-unsaturated ketones.[6,8]

(1) Lin, H.; Cornish, V. W. *Angew. Chem., Int. Ed.* **2002**, *41*, 4402−25.
(2) Yuan, L.; Kurek, I.; English, J.; Keenan, R. *Microbiol. Mol. Biol. Rev.* **2005**, *69*, 373−92.
(3) Gartner, Z. J. *Pure Appl. Chem.* **2006**, *78*, 1−14.
(4) Rozenman, M. M.; McNaughton, B. R.; Liu, D. R. *Curr. Opin. Chem. Biol.* **2007**, *11*, 259−68.
(5) Doyon, J. B.; Snyder, T. M.; Liu, D. R. *J. Am. Chem. Soc.* **2003**, *125*, 12372−3.

(6) Kanan, M. W.; Rozenman, M. M.; Sakurai, K.; Snyder, T. M.; Liu, D. R. *Nature* **2004**, *431*, 545−9.
(7) Porco and coworkers have recently reported promising results from a LC// MS-based reaction discovery system that screens the ability of one substrate to undergo a bond-forming reaction with one of many different potential partners. See: Beeler, A. B.; Su, S.; Singleton, C. A.; Porco, J. A., Jr. *J. Am. Chem. Soc.* **2007**, *129*, 1413−9.
(8) Momiyama, N.; Kanan, M. W.; Liu, D. R. *J. Am. Chem. Soc.* **2007**, *129*, 2230−1.
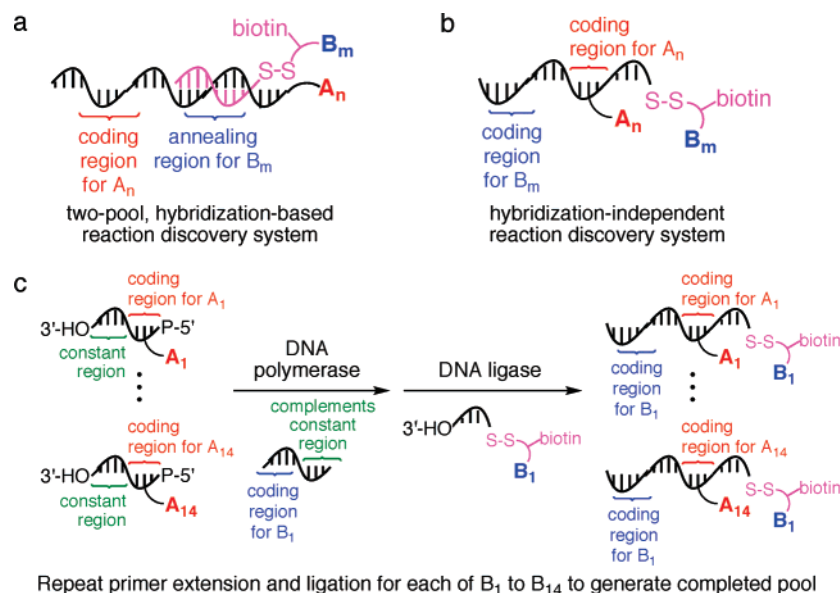
Figure 1. DNA hybridization-dependent (a) and DNA hybridization-independent (b) systems for reaction discovery. The modular assembly strategy for the hybridization-independent system is shown in (c).

Past applications of evolutionary principles to problems in the chemical sciences have largely been limited to contexts that mimic the biological milieu. For example, our first-generation reaction discovery system was limited to aqueous, high salt, low-temperature conditions that facilitate DNA hybridization. Here, we describe a hybridization-independent reaction discovery system (Figure 1b) that offers the advantages of a selection-based approach but allows for discovery in organic solvents, at high temperatures, and in the presence of additives that may preclude DNA base pairing. In addition, we report the use of this second-generation system to discover a Au(III)- or acid-mediated alkene hydroarylation reaction.

Results and Discussion

Preparation of DNA-Encoded Substrate Pools. To remove the requirement for DNA hybridization, we replaced the complementary strands in the first-generation system with a single strand attached to both substrates in a manner that enables selection for bond formation (Figure 1b). These single strands that each contain a substrate pair are assembled modularly by enzyme-catalyzed primer extension and ligation reactions that minimize the number of required starting components.

Briefly, small-molecule substrates are chosen for each of two substrate pools. Each of n substrates from pool A is linked through an internal adenine to an oligonucleotide that contains both a coding region for that substrate as well as a constant region (Figure 1c). The combined set of n pool A DNA−substrate conjugates is hybridized with one pool B primer, which contains a region complementary to the pool A constant region, as well as a unique coding sequence for one pool B substrate. Enzyme-catalyzed primer extension and ligation reactions result in a set of n finished substrates in which one pool B substrate is covalently linked through a biotinylated disulfide linker to an oligonucleotide bearing one of the n pool A substrates (Figure 1c and Supporting Information Figures S4−S6). These primer extension and ligation steps are repeated for each of m different pool B substrates. The resulting samples are combined to provide the completed pool of $n \times m$ substrate pairs, each linked to a single strand of DNA that uniquely identifies its two attached substrates.

Selection Design. Selection for bond formation in the hybridization-independent reaction discovery system (Figure 2) uses concepts developed by the molecular evolution community.[9] A single solution containing ∼1 pmol total of the completed substrate pool is incubated under a chosen set of reaction conditions. After cleavage of the disulfide bond, only those oligonucleotides encoding a productive bond-forming substrate combination retain a covalently attached biotin group (Figure 2). DNA strands encoding bond-forming combinations are selected using streptavidin-linked magnetic beads. PCR amplification and DNA microarray analysis reveals the identity of substrate pairs that undergo bond formation under the chosen reaction conditions as green microarray spots (Figure 2).[6]

A distinguishing feature of the current and former DNA-encoded reaction discovery systems is the ability to detect bond-forming reactions between substrates that would preferentially homocouple under the reaction conditions. When the substrate pool is exposed to reaction conditions at extremely low concentrations (∼nanomolar), random intermolecular reactions including substrate homocoupling do not take place at an appreciable rate. In contrast, DNA-encoded substrate pairs experience a much higher effective molarity (∼millimolar), enabling them to react and survive the selection for bond formation.

For our initial construction of this system, we chose 14 pool A and 14 pool B substrates (Figure 3) to represent simple, readily accessible organic functional groups. We assembled the system starting with 4 nmol of A-linked oligonucleotides (Figure 1c, left), resulting in >300 pmol of the 196 heterocoupling species and 28 homocoupling species that comprise the fully assembled pool of 224 substrate combinations. This quantity of material is sufficient to assess the reactivity of these substrate combinations under several hundred different reaction conditions, representing the evaluation of >50,000 potential reactions.

(9) Wilson, D. S.; Szostak, J. W. *Annu. Rev. Biochem.* **1999**, *68*, 611−47.
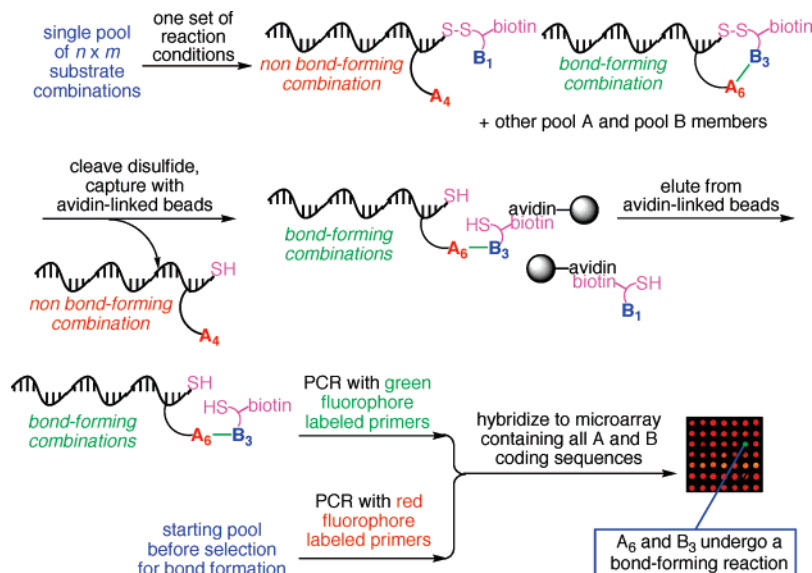
Figure 2. A one-pot selection and analysis method for the detection of bond-forming reactions between DNA-linked small-molecule substrates.

Selection for Bond Formation in Organic Solvents and at High Temperatures. To validate the ability of the system to detect bond-forming events under conditions that do not support DNA hybridization, we performed several selections for known reactions in a variety of organic solvents and at a range of temperatures. We exposed 1.5 pmol of substrate pool to 1 mM Cu(I) for 10 min at 25 °C in acetonitrile, DMF, methanol, and dioxane; in all cases, the final solvent composition was 90% organic solvent and 10% H_2O. In all four cases, selection for bond formation, PCR, and microarray analysis resulted exclusively in a strong signal from the substrate combination corresponding to the Cu(I)-catalyzed cycloaddition[10] between an azide (A7) and alkyne (B5) (Figure 3 and Supporting Information Figure S8). To test for the ability of the system to detect a less efficient reaction, we exposed the substrate pool to NaBH$_3$CN in 9:1 acetonitrile:water. Microarray analysis reflected the expected reductive amination reaction between the aryl aldehyde (A2) and allyl amine (B14) (Figure 3). Finally, we exposed the system to 1 mM Na$_2$PdCl$_4$ in aqueous solvent for 20 min at 25, 50, 65, or 95 °C. Microarray analysis reported expected Pd-mediated reactions at all four temperatures, including those that do not support DNA hybridization (Supporting Information Figure S9). Taken together, these findings validate the hybridization-independent reaction discovery system as a means of revealing bond-forming reactivity in organic solvents and across a range of temperatures.

We then applied this system to explore the reactivity of Au-(III) salts. Recent reports suggest that Au(III) can activate a broad range of organic functional groups.[11,12] We exposed 1 pmol of the substrate pool to 10 mM AuCl$_3$ in 9:1 acetonitrile:water. After 2 h at 25 °C, selection, PCR amplification, and DNA microarray analysis were carried out as described above. The strongest green spots on the microarray suggested bond formation between the indole (B9) and substrates including styrene (A14) and a terminal alkyne (A13) (Figure 3).

The B9+A13 indole–alkyne coupling signal is consistent with previous reports.[13-15] To the best of our knowledge, however, a gold-mediated coupling between indoles and styrenes has not been previously reported, although gold has been observed to mediate the addition of activated methylenes to styrene[16] and the intramolecular hydroarylation of allenes with indoles.[17]

Characterization of the Indole–Styrene Coupling Reaction. The putative indole–styrene reaction was characterized both in a DNA-linked format and in a conventional flask-based format. To mimic the reaction discovery environment, we exposed 100 pmol of DNA-linked indole to 50 mM *N*-propyl-4-vinylbenzamide and 10 mM AuCl$_3$ in 9:1 acetonitrile:water. After 2 h at 25 °C, the DNA-linked material was treated with S1 nuclease to cleave the DNA strand into mononucleotide adducts (Supporting Information). High-resolution mass spectrometry of the resulting material was consistent with the formation of a redox-neutral coupling product with a molecular weight corresponding to the sum of the indole and styrene substrates (Supporting Information Figure S10). These findings validated the B9+A14 selection result and indicated that all atoms present in the starting material were present in the reaction product.

We subsequently optimized reaction conditions using non-DNA-linked substrates. Although product formation was inefficient in MeCN, slow addition of styrene to *N*-phenylsulfonyl-(Bs)-protected indole (**1**) in the presence of stoichiometric AuCl$_3$ in methylene chloride resulted in 82% isolated yield of the Markovnikov-type hydroarylation product (**2**) without the exclusion of air or moisture from the system (Table 1, entries 1–3). High yields of product were also observed with the use of 10 mol % Au(III) in the presence of 30 mol % AgOTf[15] (Table 1, entries 4 and 5). While changing the gold counterion did not affect product formation (Table 1, entry 5), the silver counterion proved to be an important determinant of reaction efficiency

(10) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem., Int. Ed.* **2002**, *41*, 2596–9.
(11) Gorin, D. J.; Toste, F. D. *Nature* **2007**, *446*, 395–403.
(12) Hashmi, A. S.; Hutchings, G. J. *Angew. Chem., Int. Ed.* **2006**, *45*, 7896–936.
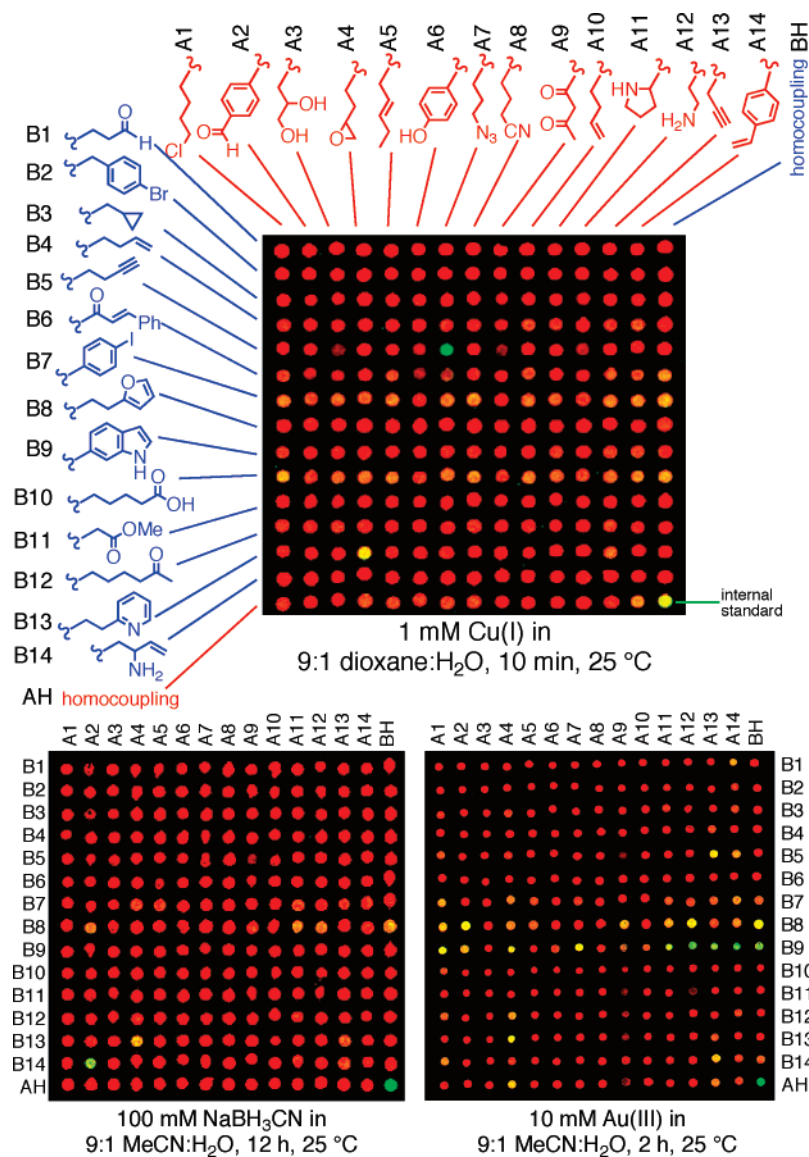
(13) Ferrer, C.; Echavarren, A. M. *Angew. Chem., Int. Ed.* **2006**, *45*, 1105–9.
(14) Li, Z.; Shi, Z.; He, C. *J. Organomet. Chem.* **2005**, *690*, 5049–5054.
(15) Sommer, K.; Reetz, M. T. *Eur. J. Org. Chem.* **2003**, 3485–3496.
(16) Yao, X.; Li, C. J. *J. Am. Chem. Soc.* **2004**, *126*, 6884–5.
(17) Liu, C.; Widenhoefer, R. A. *Org. Lett.* **2007**, *9*, 1935–8.

Figure 3. Substrates used in this work and microarray analysis of selection experiments carried out after exposure to reactions conditions listed under each array. Green spots suggest bond formation between corresponding substrates.

(Table 1, entries 4−7). On the basis of this observation[18] and recent literature reporting parallels in reactivity between metal triflates and triflic acid,[19,20] we investigated triflic acid as a potential catalyst of this olefin hydroarylation reaction. We found that 5 mol % triflic acid generated product in 91% yield (Table 1, entry 9). Silver triflate alone (Table 1, entry 8) did not mediate the reaction. Exposure of the reaction substrates to 5 mol % HCl or to 1 equiv of HCl (Table 1, entries 10 and 11) also failed to generate product, indicating that the reaction can be mediated either by triflic acid or by AuCl$_3$.

The prospect of accessing aryl-functionalized indole scaffolds under very mild conditions prompted further exploration of the substrate scope of the triflic acid-catalyzed hydroarylation reaction (Table 2). In all cases, reactions were run open to the air with no precautions taken to exclude moisture. Moderate to

(18) Anderson, L. L.; Arnold, J.; Bergman, R. G. *J. Am. Chem. Soc.* **2005**, *127*, 14542−3.
(19) Li, Z.; Zhang, J.; Brouwer, C.; Yang, C. G.; Reich, N. W.; He, C. *Org. Lett.* **2006**, *8*, 4175−8.
(20) Rosenfeld, D. C.; Shekhar, S.; Takemiya, A.; Utsunomiya, M.; Hartwig, J. F. *Org. Lett.* **2006**, *8*, 4179−82.

Table 1. Optimization of Indole-Styene Coupling[a]

entry	additive	solvent	% yield
1	1 equiv of AuCl$_3$	CH$_3$CN	<1
2	1 equiv fo AuCl$_3$	1,2-C$_2$H$_4$Cl$_2$	74
3	1 equiv of AuCl$_3$	CH$_2$Cl$_2$	82
4	10 mol % AuCl$_3$/30 mol % AgOTf	CH$_2$Cl$_2$	93
5	10 mol % AuBr$_3$/30 mol % AgOTf	CH$_2$Cl$_2$	90
6	10 mol % AuCl$_3$/30 mol % AgBF$_4$	CH$_2$Cl$_2$	52
7	10 mol % AuCl$_3$/30 mol % Ag(O$_2$CCF$_3$)	CH$_2$Cl$_2$	0
8	1 equiv of AgOTf	CH$_2$Cl$_2$	0
9	5 mol % TfOH	CH$_2$Cl$_2$	91
10	5 mol % HCl	CH$_2$Cl$_2$	0
11	1 equiv of HCl	CH$_2$Cl$_2$	0

[a] Reactions were carried out with 1 equiv of styrene added slowly to a solution of **1** and the specified additive. No care was taken to exclude air or moisture from the system. The yields shown are isolated yields of pure product.
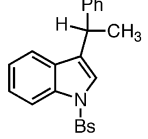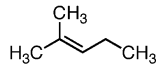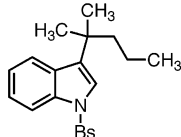
Table 2. Acid-Catalyzed Indole Hydroarylations[a]

entry	arene (3)	alkene (4)	product (5)	% yield
1				91
2				77
3				68
4				60
5			---	0

[a] Reactions were carried out with 1 equiv of **4** added slowly to a solution of **3** in the presence of 5 mol % TfOH. No care was taken to exclude air or moisture from the system. The yields shown are isolated yields of pure product.

good yields of Markovnikov-type hydroarylation products were obtained with substituted styrene variants as well as with 2-methyl-2-pentene (Table 2, entries 2−4). However, mono- and disubstituted olefins such as 1-pentene, *trans*-2-heptene, or cyclohexene are not productive substrates for this transformation (Table 2, entry 5 and Supporting Information).

The reactions described above represent a significant addition to Friedel−Crafts-type indole hydroarylation chemistry.[21] Researchers have previously developed organocatalytic, Brønsted-acid-mediated, and Lewis-acid-mediated hydroarylation reactions that couple indoles to electron-deficient olefins such as nitroalkenes[22] and α,β-unsaturated carbonyls,[21,23] or to electron-rich olefins such as enamines.[24,25] Although methods for the hydroarylation of styrenes and unactivated alkyl-substituted

olefins with indole nucleophiles are also of significant interest, they have remained scarce. Widenhoefer and co-workers have reported a Pt-catalyzed intramolecular hydroarylation of unactivated alkenes.[26] While this chemistry was recently extended to include the intermolecular hydroarylation of styrenes and simple α-olefins with 1,2-disubstituted indoles,[27] it requires high temperatures and exhibits ≤2:1 selectivity for the Markovnikov product when styryl olefins are used. Friedel−Crafts-type hydroarylations of styrenes with anisole, thiophene, and xylene nucleophiles under Lewis acid catalysis have also been recently reported,[28−30] but require the use of superstoichiometric arene (≥4 equiv) and have not been extended to indole nucleophiles.

The Au or acid-mediated regioselective hydroarylation of styrenes with Bs-protected indole reported here therefore provides an efficient route to indole-containing diaryl scaffolds not readily accessed by other hydroarylation reactions.[31] In addition, the product scaffolds accessed by these reactions are of particular interest because they are featured prominently in a wide array of broadly applied pharmacophores including peroxisome proliferator-activated receptor gamma (PPARγ) agonists, endothelin (ET$_A$) receptor agonists, and monoamine reuptake inhibitors.

Conclusion

The application of selection-based principles to address chemical problems is a relatively new but promising area.[4] Here, we have implemented an approach to reaction discovery that takes advantage of DNA encoding, DNA-programmed assembly of substrate pairs, in vitro selection, and PCR amplification, yet does not require reaction conditions that support DNA hybridization. This hybridization-independent reaction discovery system allows the simultaneous evaluation of >200 potential bond-forming combinations of substrates in a single experiment and has led to the discovery of a simple, mild method for the selective Markovnikov-type hydroarylation of olefins with indoles. Although the discovered transformation is consistent with previous notions of indole and styrene reactivity, its selectivity, efficiency, and mildness make it a potentially useful addition to known methods for accessing diaryl scaffolds. The scope and mechanism of this chemistry are under further investigation in our laboratory. Furthermore, we are in the process of applying the hybridization-independent reaction discovery system to a broad-scale exploration of transition metal-mediated and organocatalyst-mediated reactivity of simple organic functional groups in aqueous and organic solvent.

(21) Bandini, M.; Melloni, A.; Tommasi, S.; Umani-Ronchi, A. *Synlett* **2005**, 1199−1222.
(22) Herrera, R. P.; Sgarzani, V.; Bernardi, L.; Ricci, A. *Angew. Chem., Int. Ed.* **2005**, *44*, 6576−9.
(23) Evans, D. A.; Fandrick, K. R.; Song, H. J.; Scheidt, K. A.; Xu, R. *J. Am. Chem. Soc.* **2007**, *129*, 10029−41.
(24) Terada, M.; Sorimachi, K. *J. Am. Chem. Soc.* **2007**, *129*, 292−3.
(25) Wang, Y. Q.; Song, J.; Hong, R.; Li, H.; Deng, L. *J. Am. Chem. Soc.* **2006**, *128*, 8156−7.

(26) Liu, C.; Han, X.; Wang, X.; Widenhoefer, R. A. *J. Am. Chem. Soc.* **2004**, *126*, 3700−1.
(27) Zhang, Z.; Wang, X.; Widenhoefer, R. A. *Chem. Commun.* **2006**, 3717−9.
(28) Kischel, J.; Jovel, I.; Mertins, K.; Zapf, A.; Beller, M. *Org. Lett.* **2006**, *8*, 19−22.
(29) Rueping, M.; Nachtsheim, B. J.; Scheidt, T. *Org. Lett.* **2006**, *8*, 3717−9.
(30) Sun, H.-B.; Li, B.; Hua, R.; Yin, Y. *Eur. J. Org. Chem.* **2006**, 4231−4236.
(31) Kobayashi and coworkers have reported a Bronsted acid-catalyzed dehydrative coupling of l-methylindole with activated benzyl alcohols that provides access to diaryl scaffolds similar to the diaryl hydroarylation products reported here. See: Shirakawa, S.; Kobayashi, S. *Org. Lett.* **2007**, *9*, 311−4.

Supporting Information Available: Experimental details and supporting data. This material is available free of charge via the Internet at http://pubs.acs.org.

JA074155J