Combinatorial Chemistry

DOI: 10.1002/ange.200502899

Small-Molecule Diversification from Iterated Branching Reaction Pathways Enabled by DNA-Templated Synthesis**

*Christopher T. Calderone and David R. Liu**

As synthetic small molecules continue to interface with an increasingly large swath of biology,[1,2] researchers have sought new approaches to rapidly create collections of highly diverse small molecules.[3] Early small-molecule libraries generally contained a single core scaffold decorated by different chemical groups.[4] Later efforts introduced modest structural variation in the central scaffold.[5] More recently, researchers have introduced libraries based on "folding pathways", in which library intermediates undergo diverse rearrangements, and "differentiation pathways", in which intermediates are treated with different reagents to yield different scaffolds.[3] Incorporating such branching reaction pathways into library syntheses is a promising strategy for the creation of small-molecule diversity beyond what can be accessed by using nonbranched synthetic routes.[6]

The vast majority of current library synthesis efforts are based on solid-phase, split-pool methodologies.[7,8] Although these methods offer technical advantages, they also limit the diversity of structures that can be created. With relatively few exceptions,[5,9–11] it is generally not possible to direct a specific subset within a mixture of beads (for example, those beads containing molecules with a primary amine group) to one of several possible subsequent reaction conditions (for example, exposure to an acylating agent). As a result, every intermediate in a split-pool library synthesis must be reactive toward any reactant it may encounter in a subsequent step. Moreover, it is usually not possible to purify unreacted support-bound starting material away from desired products after each library synthesis step, so only highly efficient reactions can be used. As a result, the diversification of scaffolds during library synthesis has thus far been limited to a single (predominantly terminal) step.[12–15]

We recently demonstrated the ability of DNA-templated organic synthesis (DTS) to direct multiple, otherwise incompatible reactions in a single vessel.[16] We speculated that this feature and other properties of DTS,[17] such as the ability to purify library intermediates away from starting materials independent of product structure in a single manipulation,[18,19] would enable DTS-driven library synthesis to exploit branching reaction pathways at multiple steps, thus creating a network of reaction sequences that generates highly diverse families of molecules within a single library.[16] To implement the concept of iterated branching reaction pathways in a DNA-templated format, we set out to transform a single solution containing five DNA templates into five diverse types of product structures, each arising from a unique sequence of reactions. We designed a synthetic route (Scheme 1) that combines base-labile sulfone[18] and thioester[20] linkers, previously developed for multistep DTS, together with amine acylation,[21] aldehyde thiazolidination, Wittig olefination,[21] carboxylate amidation, and palladium-catalyzed cross-coupling chemistries.[21,22]

In previous DNA-templated libraries,[19] split-pool oligonucleotide synthesis was used to generate starting-template pools in which every combination of reagent annealing sequence ("codons") was represented. In the case of libraries constructed from branching reaction pathways, however, only a fraction of such a collection of templates would encode productive reaction sequences. For example, only 36% of the possible templates of the format (reaction 1 codon)–(reaction 2 codon)–(reaction 3 codon) encode three-step products in the branching library synthesis shown in Scheme 1. To address this problem, we developed a cassette ligation strategy to generate starting-template pools. In this strategy, single-stranded DNA overhangs at the ends of each cassette were designed such that ligation is only possible between adjacent cassettes that encode productive reaction sequences (see the Supporting Information).

To validate each three-step pathway, we optimized the five syntheses individually such that all step-1 reactions, all step-2 reactions, and all step-3 reactions are carried out under reaction conditions (a), (b), or (c), respectively (Figure 1). Each pathway was then individually executed starting with amine-terminated oligonucleotide **1**. After each step, the reaction mixture was captured on immobilized streptavidin and washed to remove any starting material template. The desired template-linked product eluted from the beads upon base-induced linker cleavage.[18]

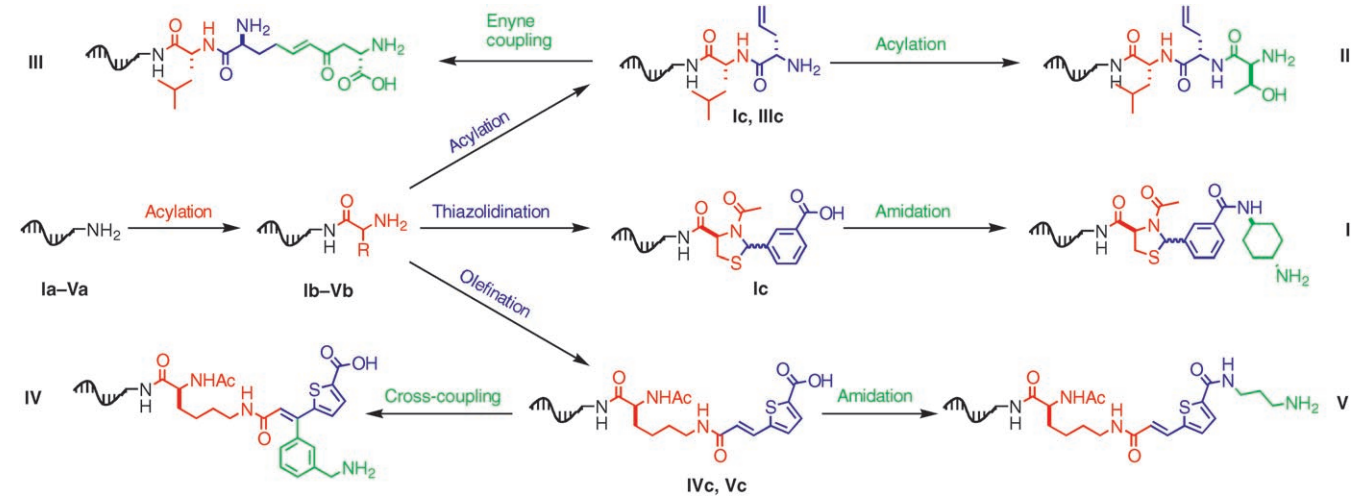
All five pathways begin with amine acylation (step 1). In step 2—the first branching step—the step-1 products participate in one of three reactions (acylation, thiazolidine formation, or Wittig olefination) dictated by the identity of the previous reaction partner of each molecule. For example, the step-1 products containing phosphorane groups exclusively participate in step-2 Wittig olefination reactions. Step 3 represents a second branch point, in which each reaction type is dictated by the structure established in step 2.

Pathway I consists of amine acylation followed by thiazolidine formation and amidation. Disulfide-protected L-cysteine was sequence-specifically delivered to the template **1** upon addition of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) and *N*-hydroxysulfosuccinimide (*s*-NHS) in 26% yield of isolated product. After disulfide reduction,

[*] C. T. Calderone, Prof. D. R. Liu
Harvard University
Department of Chemistry and Chemical Biology
12 Oxford Street, Cambridge, MA 02138 (USA)
Fax: (+1) 617-496-5688
E-mail: drliu@fas.harvard.edu

[**] This work was supported by the NIH/NIGMS (R01GM065865), the Office of Naval Research (No. N00014-03-1-0749), and an ACS Division of Organic Chemistry Graduate Research Fellowship to C.T.C. We thank Thomas M. Snyder for assistance with codon design.

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.



Scheme 1. Synthetic strategy for iterated branching reaction pathways.

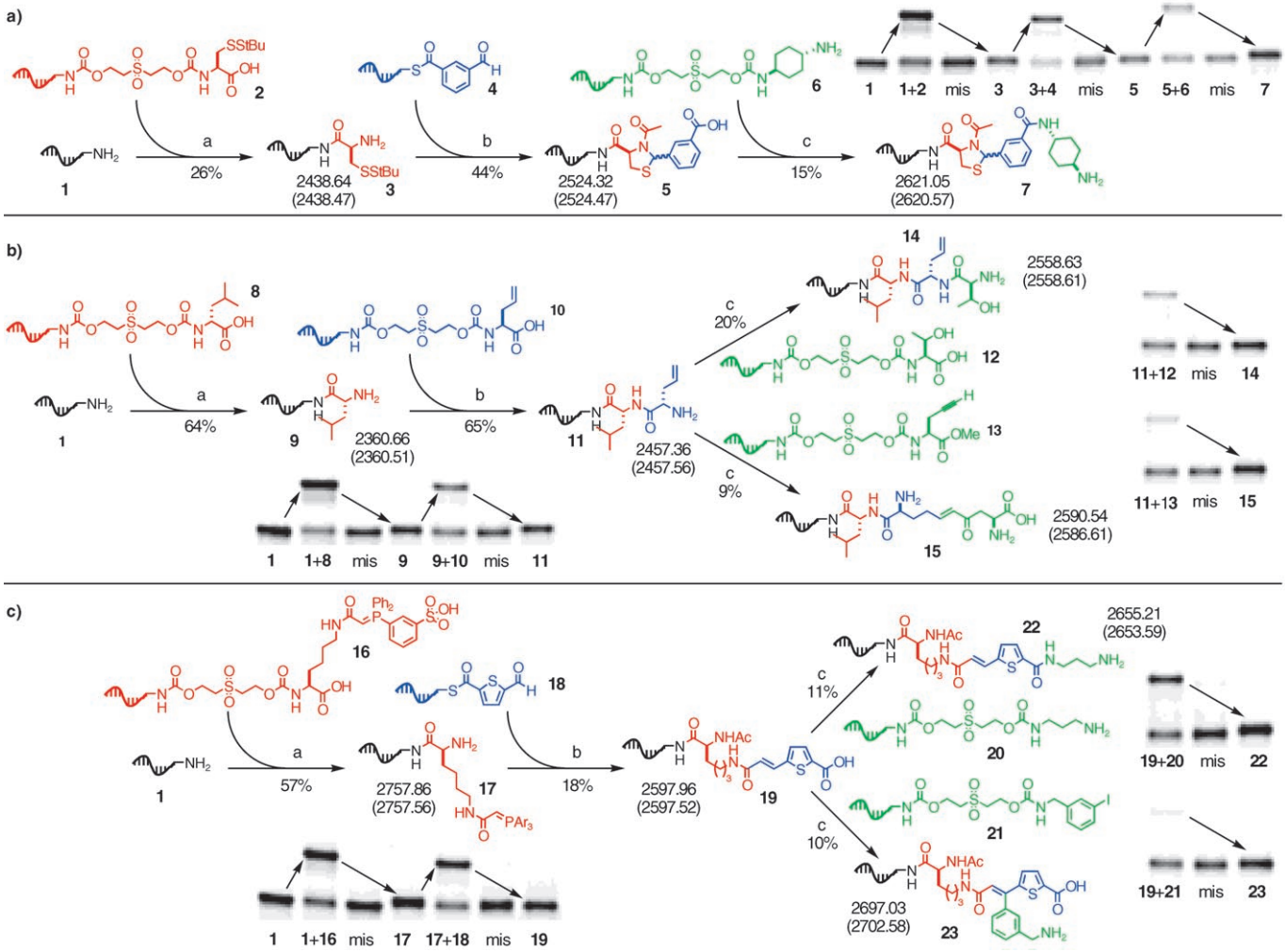


Figure 1. Individual validation of the five reaction pathways. Template amine **1** was separately treated with the specified oligonucleotide-linked reagents to yield **7**, **14**, **15**, **22**, and **23**. In each denaturing polyacrylamide gel electrophoresis (PAGE) image, lower bands represent template **1** or functionalized derivatives thereof, and upper bands represent covalently linked template–reagent pairs. Each cleaved intermediate was digested with *Eco*RI and subjected to mass-spectrometric analysis as shown, with calculated masses given in parentheses. To verify sequence specificity of each DNA-templated reaction, a sample of each template intermediate was treated with a reagent oligonucleotide containing a scrambled DNA sequence ("mis"). Reaction conditions: a) 20 mM EDC, 15 mM *s*-NHS, pH 6.0, 12 h, 23 °C, then pH 10.0 buffer, 2 h, 37 °C; b) 25 mM 1,4-dithiothreitol (DTT), pH 8.5, 1 h, 37 °C, then pH 7.0 buffer, 90 min, then 20 mM EDC, 15 mM *s*-NHS, 12 h, 23 °C, then 1:200 volume Ac_2O, 1 h, then pH 12.0 buffer, 10 mM 2-mercaptoethanol, 1 h, 37 °C; c) 1 M NaBr, 50 μM $Na_2[PdCl_4]$, pH 6.0, 4 h, 37 °C, then 25 mM DTT, 10 min, 85 °C, then 20 mM EDC, 15 mM *s*-NHS, pH 6.0, 12 h, 23 °C, then pH 12.0 buffer, 30 min, 23 °C.

treatment with a complementary DNA-linked thiol bearing 3-formylbenzoic acid for 90 min before the addition of EDC/*s*-NHS and acetylation yielded thiazolidine **5** in 44% yield. Treatment with Pd^{II} at 37°C for 4 h and addition of EDC/*s*-NHS yielded product **7** in 15% yield (2% overall yield of isolated product over three steps, Figure 1a).

Pathways II and III use identical reaction conditions as pathway I but effect two consecutive amine acylations in steps 1 and 2. The third step of pathway II consists of a final amine acylation to yield a tripeptide, whereas the third step of pathway III uses a palladium-catalyzed alkene–alkyne coupling to generate an enone.[22] To validate these pathways, D-leucine and L-allylglycine were sequentially added to template **1** (steps 1 and 2) in 64 and 65% yields of the isolated products, respectively, to form **11**. In step 3, the allylglycine was either acylated with L-threonine to yield tripeptide **14** in 20% yield (pathway III, 8.3% yield of the isolated product over three steps) or was cross-coupled[22] with the alkyne group in L-propargylglycine to yield enone **15** in 8% yield (pathway IV, 3.6% yield of the isolated product over three steps; Figure 1b).

Finally, pathways IV and V begin with amine acylation followed by Wittig olefination. The third step of pathway IV is carboxylate amidation, whereas the third step in pathway V is a palladium-catalyzed Heck reaction. Thus, a lysine derivative bearing a side-chain phosphorane **16** was coupled to **1** (step 1) in 57% yield of the isolated product followed by incubation with the DNA thioester of formylthiophene carboxylic acid (step 2) to yield **19** in 18% yield of isolated product. In the third step, treatment with amine **20** yielded disubstituted olefin **22** in 11% yield of isolated product (1.7% yield over three steps), whereas the Heck coupling with aryl iodide **21** yielded trisubstituted olefin **23** in 10% yield of isolated product (0.8% yield over three steps; Figure 1c). All five pathways were monitored by gel electrophoresis and MALDI-TOF mass-spectrometric analysis. Each pathway yielded intermediates and final products whose electrophoretic behavior and mass-spectrometric characterization were consistent with the desired structures. The reaction sequences resulted in product yields similar to those reported for previous three-step DNA-templated synthetic sequences, despite requiring compromises among the different optimal reaction conditions in steps 2 and 3.[18–20]

Having developed each of the pathways individually, we next performed the simultaneous synthesis of **I**–**V** in a single solution per step (Figure 2). The ligated and gel-purified pool of templates was treated with 1.5 equivalents of each functionalized-reagent DNA per step and was incubated under the same conditions used to verify each pathway individually. After each step, we observed only the presence of desired products by nuclease digestion followed by MALDI-TOF-MS. To evaluate the sequence specificity of each reaction, each step was repeated using reagents linked to noncomplementary (mismatched) oligonucleotides. No significant product formation (< 5%) was observed (Figure 2). Because not all five templates yield step-1 and -2 products that are distinguishable by mass, we also treated an aliquot of material at each step with a reactive oligonucleotide uniquely complementary to each template (**I′**–**V′**), thus converting a portion of each pool into a higher-molecular-weight species

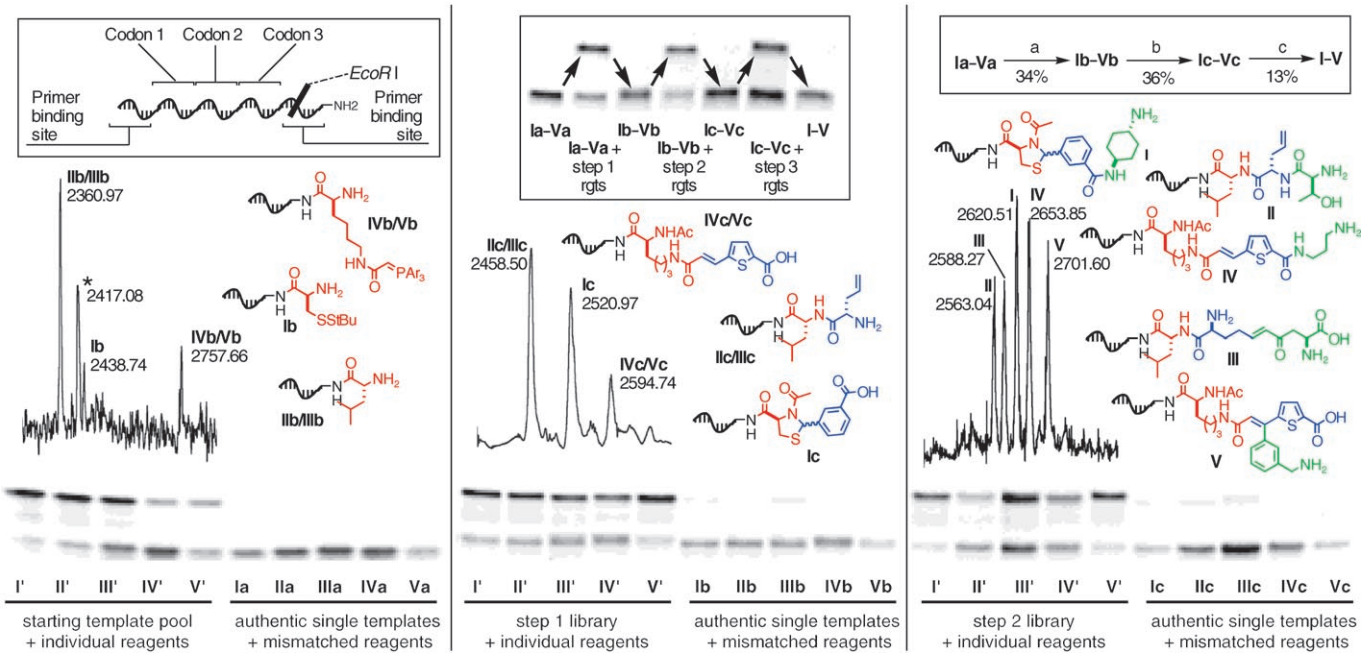


Figure 2. Characterization of the single solution per step (pooled) synthesis of **I**–**V**. PAGE analysis of the three pooled steps is shown in the upper gel image. In the mass spectrum of step 1, the peak denoted with an asterisk arises from the decomposition of **IVb** and **Vb** during mass-spectrometric analysis (see the Supporting Information). In each of the lower gels, the five lanes furthest on the left show the reaction of the starting-template pool, step-1 product pool, or step-2 product pool with each of the reagents **I′**–**V′**, which are reactive with and uniquely complementary to templates **Ia**–**Va**, respectively. The five lanes furthest on the right of each lower gel image show the reaction of individual starting materials for each step under conditions identical to those of the actual library synthesis (see Figure 1), but with sequence-mismatched reagents.

and confirming that all five templates were present at every step of the pooled synthesis (see the Supporting Information).

From 8 nmol of template, we obtained 125 pmol of **I**–**V** which corresponds to an average overall yield of 1.6% of isolated product. This quantity of material, even if hypothetically divided among thousands of library members, is sufficient for >1000 in vitro selections and PCR-based amplifications needed to identify library members with desired properties.[19,23] Our initial use of iterated branching reaction pathways demonstrates the promise of such an approach in the synthesis of diverse small-molecule libraries that would be challenging to create by using current library synthesis methods. The structures created in this work also include several of the most complex synthetic small-molecule products translated from DNA to date.

Received: August 15, 2005
Published online: October 20, 2005

Keywords: bioorganic chemistry · combinatorial chemistry · diversity-oriented synthesis · DNA · template synthesis

[1] B. R. Stockwell, *Nature* **2004**, *432*, 846.
[2] S. L. Schreiber, *Bioorg. Med. Chem.* **1998**, *6*, 1127.
[3] M. D. Burke, S. L. Schreiber, *Angew. Chem.* **2004**, *116*, 48; *Angew. Chem. Int. Ed.* **2004**, *43*, 46.
[4] B. A. Bunin, M. J. Plunkett, J. A. Ellman, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 4708.
[5] S. Ding, N. S. Gray, X. Wu, Q. Ding, P. G. Schultz, *J. Am. Chem. Soc.* **2002**, *124*, 1594.
[6] G. C. Micalizio, S. L. Schreiber, *Angew. Chem.* **2002**, *114*, 3406; *Angew. Chem. Int. Ed.* **2002**, *41*, 3272.
[7] A. Furka, F. Sebestyen, M. Asgedom, G. Dibo, *Int. J. Pept. Protein Res.* **1991**, *37*, 487.
[8] R. B. Merrifield, *J. Am. Chem. Soc.* **1963**, *85*, 2149.
[9] K. C. Nicolaou, J. A. Pfefferkorn, S. Barluenga, H. J. Mitchell, A. J. Roecker, G. Q. Cao, *J. Am. Chem. Soc.* **2000**, *122*, 9968.
[10] K. C. Nicolaou, J. A. Pfefferkorn, H. J. Mitchell, A. J. Roecker, S. Barluenga, G. Q. Cao, R. L. Affleck, J. E. Lillig, *J. Am. Chem. Soc.* **2000**, *122*, 9954.
[11] K. C. Nicolaou, J. A. Pfefferkorn, A. J. Roecker, G. Q. Cao, S. Barluenga, H. J. Mitchell, *J. Am. Chem. Soc.* **2000**, *122*, 9939.
[12] M. D. Burke, E. M. Berger, S. L. Schreiber, *Science* **2003**, *302*, 613.
[13] M. D. Burke, E. M. Berger, S. L. Schreiber, *J. Am. Chem. Soc.* **2004**, *126*, 14095.
[14] H. Oguri, S. L. Schreiber, *Org. Lett.* **2005**, *7*, 47.
[15] S. J. Taylor, A. M. Taylor, S. L. Schreiber, *Angew. Chem.* **2004**, *116*, 1713; *Angew. Chem. Int. Ed.* **2004**, *43*, 1681.
[16] C. T. Calderone, J. W. Puckett, Z. J. Gartner, D. R. Liu, *Angew. Chem.* **2002**, *114*, 4278; *Angew. Chem. Int. Ed.* **2002**, *41*, 4104.
[17] Z. J. Gartner, D. R. Liu, *J. Am. Chem. Soc.* **2001**, *123*, 6961.
[18] Z. J. Gartner, M. W. Kanan, D. R. Liu, *J. Am. Chem. Soc.* **2002**, *124*, 10304.
[19] Z. J. Gartner, B. N. Tse, R. Grubina, J. B. Doyon, T. M. Snyder, D. R. Liu, *Science* **2004**, *305*, 1601.
[20] X. Y. Li, Z. J. Gartner, B. N. Tse, D. R. Liu, *J. Am. Chem. Soc.* **2004**, *126*, 5090.
[21] Z. J. Gartner, M. W. Kanan, D. R. Liu, *Angew. Chem.* **2002**, *114*, 1874; *Angew. Chem. Int. Ed.* **2002**, *41*, 1796.
[22] M. W. Kanan, M. M. Rozenman, K. Sakurai, T. M. Snyder, D. R. Liu, *Nature* **2004**, *431*, 545.
[23] J. B. Doyon, T. M. Snyder, D. R. Liu, *J. Am. Chem. Soc.* **2003**, *125*, 12372.

