Stereoselective Synthesis of over Two Million Compounds Having Structural Features Both Reminiscent of Natural Products and Compatible with Miniaturized Cell-Based Assays

Derek S. Tan, Michael A. Foley, Matthew D. Shair, and Stuart L. Schreiber*

Department of Chemistry & Chemical Biology
Harvard Institute of Chemistry & Cell Biology
Howard Hughes Medical Institute
Harvard University, 12 Oxford St.
Cambridge, Massachusetts 02138

Received May 19, 1998

Understanding the cellular function of a protein usually requires a means to alter the function. This is commonly done by mutating the gene encoding the protein (genetic approach). It can also be done by binding the protein directly with a small molecule ligand (chemical genetic approach). Such ligands, primarily natural products or synthetic variants of them, can either inactivate[1] or activate[2] protein function. For the chemical genetic approach to have its maximal impact, efficient methods of ligand discovery will be required to provide, in the limit, a small molecule partner for every gene product.

Our laboratory has recently described miniaturized cell culture assays for screening large numbers of small molecule ligands, potentially on a genome-wide basis.[3] However, despite its success in natural products synthesis, synthetic chemistry has not yet been applied to the synthesis of vast numbers of compounds with structures both reminiscent of natural products and compatible with miniaturized assays. These features of a synthesis will likely be required in order to discover nonnatural compounds having the binding affinities and specificities characteristic of natural products.

The synthetic strategy we have undertaken is to develop highly efficient multistep syntheses of natural product-like compounds that include several coupling steps and to use split-pool techniques[4] at these steps in order to generate diverse outcomes. This technique requires that one of the coupling substrates be immobilized on a solid support. Our syntheses are further constrained by the water-compatible photolabile supports required when carrying out large numbers ($>10^5$) of miniaturized "nanodroplet" cell culture assays aimed at identifying cell permeable ligands.[3] This presents a considerable synthetic challenge since the structural elements that provide these properties (e.g., poly(ethylene glycol), nitrobenzyl moieties) are incompatible with numerous reagents used in conventional synthetic chemistry.

With these considerations in mind, we first converted shikimic acid, **1**, into both enantiomers of epoxycyclohexenol carboxylic acid **2**,[5] which were then coupled to a photocleavable linker on solid support by standard methods (Figure 1).[6] Treatment of the resin-bound epoxycyclohexenol, **3**, with various nitrone carboxylic acids,[7] **7a**−**f** and **9b**−**d**, under esterification conditions yielded tetracyclic compounds **4a**−**f** and **5b**−**d** with complete regio- and

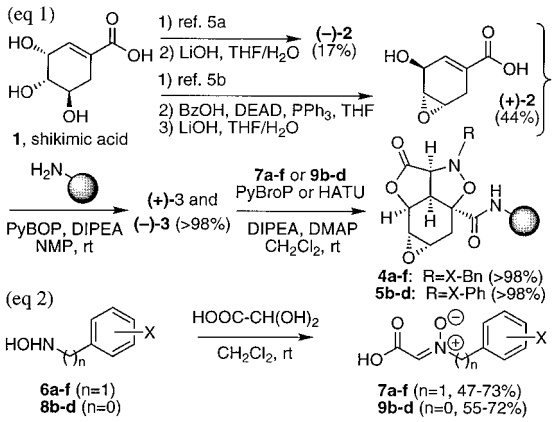


Figure 1. Synthesis of tetracycles and nitrones. **a**: X = H, **b**: X = 2-I, **c**: X = 3-I, **d**: X = 4-I, **e**: X = 4-CF_3, **f**: X = 3,4-$(OMe)_2$.

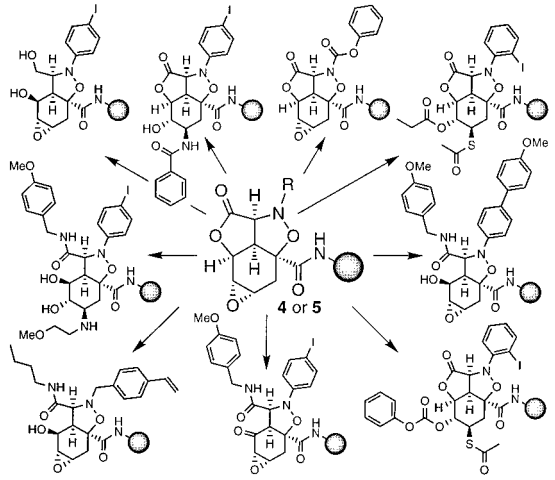


Figure 2. Compounds derived from templates **4** and **5**.

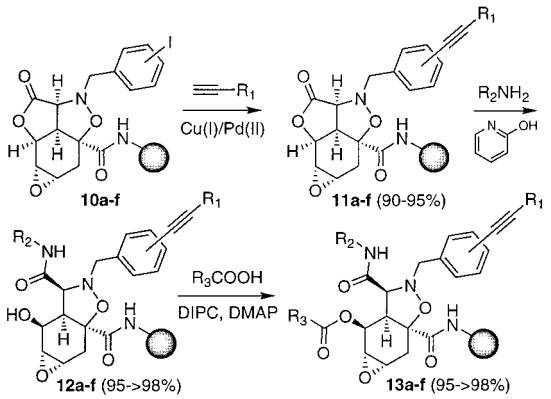
stereoselectivity, presumably via tandem acylation/1,3-dipolar cycloaddition.[8] No intermediate nitrone ester or carboxy isoxazolidine structures were observed. The iodoaryl tetracycles **4b**−**d** and **5b**−**d** were selected for further study since the support-bound aryl iodides could be modified with commercially available reagents, avoiding the need for solution-phase syntheses of numerous nitrone acids.

Tetracycles **4b**−**d** and **5b**−**d** are rigid, densely functionalized compounds that can undergo further reactions to introduce a variety of functional groups around the central octahydrobenzisoxazole structure, notably, without the use of protecting groups. The iodoaryl groups can serve as substrates for palladium cross-coupling reactions. The electrophilic lactone and epoxide can react with nucleophiles while simultaneously unmasking alcohols for subsequent reactions. Furthermore, reductive N−O bond cleavage would provide two additional handles for functionalization. We developed several such reactions (Figure 2) with product purity ranging from ∼50 to ≥98% following photocleavage.[9] Since purification is not practical in a large split-pool synthesis, only reactions generating products in ≥90% purity were acceptable.

The most promising reactions were studied in detail to define the scope, limitations, and optimal conditions for each (Figure 3). Iodobenzyl tetracycles **4b**−**d** were selected as scaffolds since

(1) See http://www-schreiber.chem.harvard.edu on the Web.
(2) Crabtree, G. R.; Schreiber, S. L. *Trends in Biochem. Sci.* **1996**, *21*, 418−422.
(3) (a) Borchardt, A.; Liberles, S. D.; Biggar, S. R.; Crabtree, G. R.; Schreiber, S. L. *Chem. Biol.* **1997**, *4*, 961−968. (b) You, A. J.; Jackman, R. J.; Whitesides, G. M.; Schreiber, S. L. *Chem. Biol.* **1997**, *4*, 969−975.
(4) (a) Furka, A.; Sebestyén, F.; Asgedom, M.; Dibó, G. *Int. J. Pept. Protein Res.* **1991**, *37*, 487−493. (b) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. *Nature* **1991**, *354*, 82−84.
(5) (a) Wood, H. B.; Ganem, B. *J. Am. Chem. Soc.* **1990**, *112*, 8907−8909. (b) McGowan, D. A.; Berchtold, G. A. *J. Org. Chem.* **1981**, *46*, 2381−2383. (c) Mitsunobu, O. *Synthesis* **1981**, 1−28.
(6) The resin structure shown represents Tentagel S NH_2, a poly(ethylene glycol)-polystyrene copolymer, loaded with a photocleavable 3-amino-3-*o*-nitrophenylpropionic acid linker. (Brown, B. B.; Wagner, D. S.; Geysen, H. M. *Mol. Div.* **1995**, *1*, 4−12.).
(7) Keirs, D.; Overton, K. *Heterocycles* **1989**, *28*, 841−848.
(8) Tamura, O.; Okabe, T.; Yamaguchi, T.; Gotanda, K.; Noe, K.; Sakamoto, M. *Tetrahedron* **1995**, *51*, 107−118.
(9) Photocleavage products were analyzed by ^{1}H NMR, FAB-MS, HPLC, and TLC.

S0002-7863(98)01746-6 CCC: $15.00
Published on Web 08/08/1998



with ε-aminocaproic acid spacer

a: R_1=(*ortho*)-*t*-butyl, R_2=4-methoxybenzyl, R_3=isobutyl
b: R_1=(*meta*)-diethoxymethyl, R_2=2,2-dimethoxyethyl, R_3=1-propenyl
c: R_1=(*para*)-*n*-propyl, R_2=4-methoxyphenethyl, R_3=ethyl

without spacer

d: R_1=(*ortho*)-benzyl, R_2=cyclobutyl, R_3=4-methoxybenzyl
e: R_1=(*meta*)-2-propenyl, R_2=allyl, R_3=phenyl
f: R_1=(*para*)-4-chlorophenyl, R_2=2-methoxybenzyl, R_3=isopropyl

Figure 3. Synthesis of 24 demonstration compounds and a 2.18 million compound library.

their palladium-mediated cross-coupling reactions were more efficient than those involving iodophenyl tetracycles **5b**−**d**. A carefully controlled alkyne coupling reaction was used to convert the aryl iodides (**10**) into aryl alkynes (**11**).[10] Cycloreversion was a persistent problem in uncatalyzed lactone aminolysis; however, 2-hydroxypyridine mediated the efficient reaction of primary amines with the γ-butyrolactone to afford γ-hydroxyamides (**12**).[11] After testing numerous acylating conditions, we found that the unmasked secondary alcohol could be capped reliably with DIPC/DMAP-activated carboxylic acids to give γ-acyloxy amides (**13**).[12]

Split-pool synthesis provides the theoretical means to synthesize the large numbers of molecules likely required for chemical genetic screens, where molecules replace mutations. However, such syntheses present enormous analytical challenges. We have developed a four-step protocol in order to provide maximum confidence that a complex split-pool synthesis of encoded molecules yields the anticipated products in high purity and efficiency. First, to demonstrate the suitability of the entire reaction sequence for library synthesis, we synthesized and fully characterized 24 compounds,[13] **10a**−**f** through **13a**−**f**, with the latter providing photocleavage products in acceptable 80−90% purity. Second, potential building blocks at each step were tested by reaction with a single selected substrate. Thus, 50 alkynes were tested by reaction with **10a**, 87 amines with **11a**, and 98 acids with **12a**. Of these, 23 alkynes, 54 amines, and 44 acids reacted with ≥90% conversion and purity (LC-MS). These building blocks, along with a limited number of less optimal candidates (generally ≥70% conversion), were selected for inclusion in library synthesis. Third, to verify effective synthesis in a split-pool format, we generated a small test library with building blocks carefully selected such that the products within each final pool would have unique masses, allowing analysis by LC−MS. Indeed, all of the expected 456 masses were detected.

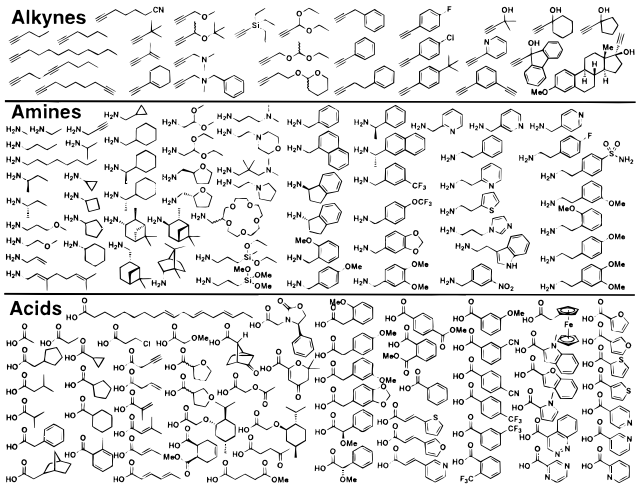


Figure 4. Building blocks used in library synthesis.

Fourth, during the encoded library synthesis below,[14] the effectiveness of carbene tagging was verified by analyzing beads from every pool at each step of the synthesis.

A large encoded library was constructed by split-pool synthesis beginning with attachment of two spacers, ϵ-aminocaproic acid and glycine, to the photolinker on resin. A third portion of the resin was left without a spacer. The resin was pooled, divided into two portions, and one enantiomer of epoxycyclohexenol carboxylic acid **2** was coupled to each pool. After the resin was pooled and divided into three portions, iodobenzyl nitrone acids **7b**−**d** were coupled, resulting in a total of 18 tetracyclic scaffolds. The synthesis was completed by reaction with 30 terminal alkynes, then with 62 primary amines, and finally with 62 carboxylic acids (Figures 3 and 4) with an additional skip codon at each step.[15]

This six-step reaction sequence resulted in a collection of compounds calculated to contain 2.18 million distinct, spatially separated, and encoded chemical entities. These synthetic compounds are rigid, stereochemically defined, and structurally diverse, characteristics common to many natural products. Furthermore, the synthesis allows the controlled release of compounds from the individual 90 μ supports into nanodroplets containing engineered cells, features critical to the miniaturized cell-based assays now being used to screen this library for cell permeable, protein-binding ligands.[3] Encouragingly, we have already found that several members of this library activate a reporter gene in mink lung cells.[16] We note, however, that this synthesis relies in part on simple acylation chemistry. Syntheses benefiting from a greater range of mild and selective synthetic methods may facilitate the routine discovery of compounds with protein-binding properties rivaling those of natural products.

Acknowledgment. We thank Dr. A. Tyler, J. Athanasopoulos, and J. Lynch for expert mass spectral support.

Supporting Information Available: Acknowledgments, additional discussion, complete experimental procedures, and analytical data for **2**, **3**, **6b**−**d**, **7b**−**d**, **10**−**13a**−**f**, building block testing, test library synthesis, large library synthesis, binary encoding (77 pages, print/PDF). See any current masthead page for ordering information and Web access instructions.

JA981746O

(10) (a) Sonogashira, K.; Tohda, Y.; Hagihara, N. *Tetrahedron Lett*. **1975**, 4467−4470. (b) Young, J. K.; Nelson, J. C.; More, J. S. *J. Am. Chem. Soc*. **1994**, *116*, 10841−10842. (c) Collini, M. D.; Ellingboe, J. W. *Tetrahedron Lett*. **1997**, *38*, 7963−7966. (d) Odingo, J.; Sharpe, B. A.; Oare, D., Presented at the 213th National Meeting of the American Chemical Society, San Francisco, CA, April, 1997; Paper ORGN 574.

(11) (a) Rony, P. R. *J. Am. Chem. Soc*. **1969**, *91*, 6090−6096. (b) Openshaw, H. T.; Whittaker, N. *J. Chem. Soc. C* **1969**, 89−91.

(12) (a) Atuegbu, A.; Maclean, D.; Nguyen, C.; Gordon, E. M.; Jacobs, J. W. *Bioorg. Med. Chem*. **1996**, *4*, 1097−1106. (b) Hydroxyl groups introduced during the alkyne coupling step were also acylated efficiently under these conditions.

(13) ^{1}H NMR with complete peak assignments by extensive homonuclear decoupling and/or DQF-COSY, TOCSY, and NOESY experiments, FAB-MS, HRMS, HPLC, TLC.

(14) Binary encoding was performed by a modified version of the literature procedure (see Supporting Information): Nestler, H. P.; Bartlett, P. A.; Still, W. C. *J. Org. Chem*. **1994**, *59*, 4723−4724.

(15) Combs, A. P.; Kapoor, T. M.; Feng, S.; Chen, J. K.; Daudé-Snow, L. F.; Schreiber, S. L. *J. Am. Chem. Soc*. **1996**, *118*, 287−288.

(16) Compounds with the general structure **13** in Figure 3, with no spacer, R_1 = carbon-based substituents attached to an alkyne substituted at the meta position (see S46 in Supporting Information), R_2 = 3,4-dimethoxybenzyl, and R_3 = methoxymethyl, were found to activate the reporter plasmid p3TPLux in a stably transfected mink lung cell line, both independently and synergistically with TGF-β1 (B. R. Stockwell and S. L. S., unpublished results and *Curr. Biol*. **1998**, *8*, 761−770.).