# Synthesis and Preliminary Evaluation of a Library of Polycyclic Small Molecules for Use in Chemical Genetic Assays

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

*Contribution from the Howard Hughes Medical Institute, Department of Chemistry and Chemical Biology, and Harvard Institute of Chemistry and Cell Biology, Harvard University, 12 Oxford St., Cambridge, Massachusetts 02138*

*Received June 23, 1999*

**Abstract:** (−)-Shikimic acid, **3**, was converted into both enantiomers of epoxycyclohexenol carboxylic acid, **7**, which were attached to a solid support via a photocleavable linker. Tandem acylation−1,3-dipolar cycloaddition with nitrones **11a**−**g** yielded tetracyclic templates **12a**−**g**. After development of several efficient coupling reactions of iodobenzyl tetracycles **12b**−**d** and completion of extensive validation protocols, a split-pool synthesis yielded a binary encoded library calculated to contain 2.18 million polycyclic compounds. These compounds are compatible with miniaturized cell-based "forward" chemical genetic assays designed to explore biological pathways and "reverse" chemical genetic assays designed to explore protein function. As a simple illustration of the potential of these compounds, several were shown to activate a TGF-$\beta$-responsive reporter gene in mammalian cells.

## Introduction

Biological systems can be explored using genetic methods, where mutations in genes are introduced and the resulting biological effects are examined. In the classical or "forward genetic" approach, a large collection of randomly generated mutations is introduced and subsequently screened in search of a specific biological effect. The mutated gene is then identified and the roles of its cognate wild-type gene and encoded protein product in the biological pathway are investigated further.

Progress in whole genome sequencing[1] has facilitated a related method of exploring biological systems. Such sequencing efforts are uncovering myriad novel genes with unknown functions. In the "reverse genetic" approach, a deletion or "knockout" mutation is targeted to a known gene of unknown function. This is followed by a broad search for all resulting biological effects, allowing the function of the gene to be inferred.

The chemical genetic approach[2,3] is complementary to the genetic approach (Figure 1). Instead of mutations, small molecules are used to activate or inactivate proteins by direct interactions. The "forward chemical genetic" approach involves the search for small molecules, beginning with a large random collection, that modulate a specific biological pathway or process. The active small molecule and its protein partner are then identified, leading to an understanding of the protein's role in the pathway. In the "reverse chemical genetic" approach, a small molecule having a known and specific protein target is used to alter the function of its target. By determining the pathways and processes altered by the small molecule, the functions of its target can be inferred.

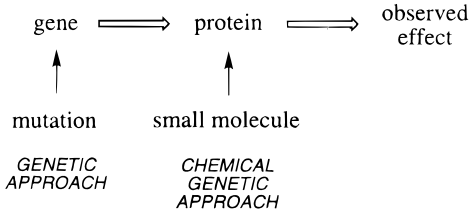



**Figure 1.** Genetic and chemical genetic approaches.

Chemical genetic approaches allow protein function to be altered rapidly and conditionally. They are also dependent upon the identification of small molecules that bind to proteins with high specificity. Natural products and natural product-derived compounds provide many of the most striking examples, particularly in terms of their specificity (Figure 2). Examples of interest to our laboratory include rapamycin, which led to the discovery of the immunophilin FKBP12 and the amino acid sensor FRAP,[4,5] and trapoxin, which led to the discovery of the chromatin-remodeling histone deacetylases (HDACs)[6,7] and nucleosome remodeling and deacetylating (NRD) complex.[8] The natural products Taxol and discodermolide bind and activate, rather than inactivate, the protein tubulin,[9] which itself was discovered as the protein target of colchicine.[10−12] Inactivating small molecules have also been converted into activating ones

(1) Collins, F. S.; Patrinos, A.; Jordan, E.; Chakravarti, A.; Gesteland, R.; Walters, L.; and the members of the DOE and NIH planning groups *Science* **1998**, *282*, 682−689.
(2) Schreiber Group Website. http://www-schreiber.chem.harvard.edu (accessed Jun 1999).
(3) Schreiber, S. L. *Bioorg. Med. Chem.* **1998**, *6*, 1127−1152.
(4) Brown, E. J.; Schreiber, S. L. *Cell* **1996**, *86*, 517−520.
(5) Brown, E. J.; Albers, M. W.; Shin, T. B.; Ichikawa, K.; Keith, C. T.; Lane, W. S.; Schreiber, S. L. *Nature* **1994**, *369*, 756−758.
(6) Hassig, C. A.; Schreiber, S. L. *Curr. Opin. Chem. Biol.* **1997**, *1*, 300−308.
(7) Taunton, J.; Hassig, C. A.; Schreiber, S. L. *Science* **1996**, *272*, 408−411.
(8) Tong, J. K.; Hassig, C. A.; Schnitzler, G. R.; Kingston, R. E.; Schreiber, S. L. *Nature* **1998**, *395*, 917−921.
(9) Hung, D. T.; Chen, J.; Schreiber, S. L. *Chem. Biol.* **1996**, *3*, 287−293.
(10) Borisy, G. G.; Taylor, E. W. *J. Cell Biol.* **1967**, *34*, 525−533.
(11) Borisy, G. G.; Taylor, E. W. *J. Cell Biol.* **1967**, *34*, 535−548.
(12) Shelanski, M. L.; Taylor, E. W. *J. Cell Biol.* **1967**, *34*, 549−554.

10.1021/ja992144n CCC: $18.00 
Published on Web 09/21/1999

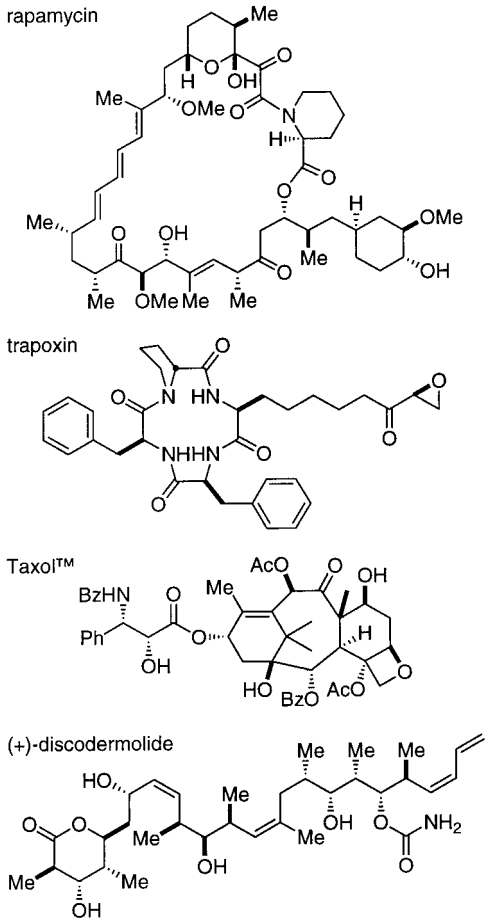



**Figure 2.** Natural products used in chemical genetic studies.

by combined genetic modification of target proteins and chemical modification of natural products to generate chemical inducers of dimerization (CIDs).[13]

It seems unlikely that natural products alone will provide the hypothetical "complete" set of small molecules that would allow the functions of all proteins, as well as their individual domains, to be determined. For chemistry to have its maximal effect on biology, efficient methods for discovering this set of small molecules are in great demand. Organic synthesis will play a central role, yet syntheses must be devised that will allow the products to be used effectively in both forward and reverse chemical genetic studies. The strategy of split-pool (also known as mix-and-split) synthesis, involving a simple mechanical intervention at strategic steps in a conventional synthesis, provided the first practical means of synthesizing vast numbers of spatially segregated small molecules.[14,15] This strategy ideally requires that an early building block of the synthetic compounds be immobilized onto small polymeric supports. Its most effective implementation will require the development of syntheses of small molecules having structural features optimal for protein binding and specificity. There are many challenges involved in the identification of these structural features and in the development of effective syntheses of molecules equipped with them. There are also many challenges associated with the development

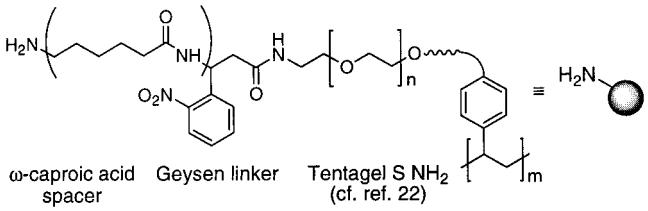



**Figure 3.** Abbreviated spacer, photolabile linker, and support.

of effective forward and reverse chemical genetic screens. Two that we are finding to be of great value are the cytoblot technique (for forward chemical genetics)[16,17] and small molecule printing (for reverse chemical genetics).[18]

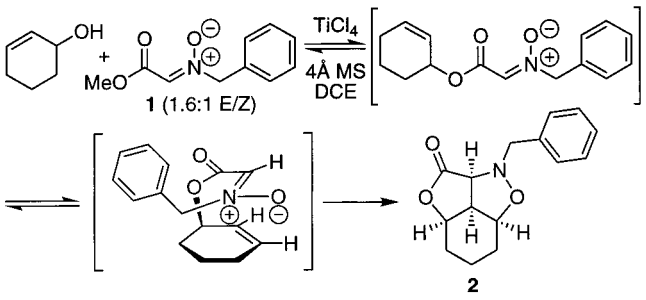
We are developing an integrated program aimed at advancing chemical genetics. At the front end of this effort is a series of binary encoded,[19] split-pool syntheses of complex small molecules having structural features reminiscent of natural products. A preliminary account of a first such synthesis was recently reported.[20] That study addressed numerous issues that we feel will be of relevance to future studies, especially the development of library validation protocols. Herein, we provide a full account of these investigations.

**Template Synthesis.** Decreasing the degrees of conformational freedom in small molecules can facilitate protein binding and specificity by reducing the entropic cost of binding. In natural products, this conformational restriction results from macrocyclic, polycyclic, or highly substituted linear structures (Figure 2). An appropriate constellation of hydrogen bonding and hydrophobic functional groups provides the enthalpic driving force for protein binding. We are synthesizing similarly constructed molecules using efficient multistep syntheses that include several coupling steps. The use of the split-pool strategy allows the efficient syntheses of collections of $10^4-10^7$ spatially segregated compounds, which can then be used in phenotype-based cytoblot assays or binding-based small molecule printing assays. Since purification is not practical in large split-pool syntheses, reactions generating products in ≥90% purity are of considerable value. Ideally, every reaction should increase the diversity of the library; thus, batch transformations and protecting group manipulations are undesirable.[21] Selection of a solid support and linker that allow release of the synthetic compound from a single bead into miniaturized wells is a critical aspect of the experimental design. Linkers and beads can also present significant synthetic challenges since their structural elements may be incompatible with numerous reagents used in conventional synthetic chemistry.

In our initial studies, we selected TentaGel S $NH_2$,[22] a poly(ethyleneglycol)−polystyrene copolymer, as a solid support with compounds attached via a photolabile linker developed by Geysen and co-workers (Figure 3).[23] This support-linker com-

(13) Crabtree, G. R.; Schreiber, S. L. *Trends Biochem. Sci.* **1996**, *21*, 418−422.
(14) Furka, A.; Sebestyén, F.; Asgedom, M.; Dibó, G. *Int. J. Pept. Protein Res.* **1991**, *37*, 487−493.
(15) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. *Nature* **1991**, *354*, 82−84.
(16) Stockwell, B. R.; Haggarty, S. J.; Schreiber, S. L. *Chem. Biol.* **1999**, *6*, 71−83.
(17) Mayer, T. U.; Kapoor, T. M.; Haggarty, S. J.; King, R. W.; Schreiber, S. L.; Mitchison, T. J. *Science*. In press.
(18) MacBeath, G.; Koehler, A. N.; Schreiber, S. L. *J. Am. Chem. Soc.* **1999**, *121*, 7967−7968.
(19) Ohlmeyer, M. H. J.; Swanson, R. N.; Dillard, L. W.; Reader, J. C.; Asouline, G.; Kobayashi, R.; Wigler, M.; Still, W. C. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10922−10926.
(20) Tan, D. S.; Foley, M. A.; Shair, M. D.; Schreiber, S. L. *J. Am. Chem. Soc.* **1998**, *120*, 8565−8566.
(21) Bartlett and co-workers have proposed related criteria for the synthetic design of split-pool libraries: Marx, M. A.; Grillot, A.-L.; Louer, C. T.; Beaver, K. A.; Bartlett, P. A. *J. Am. Chem. Soc.* **1997**, *119*, 6153−6167.
(22) Rapp Polymere Home Page. http://www.rapp-polymere.com (accessed Jun 1999).
(23) Brown, B. B.; Wagner, D. S.; Geysen, H. M. *Mol. Div.* **1995**, *1*, 4−12.

**Scheme 1**

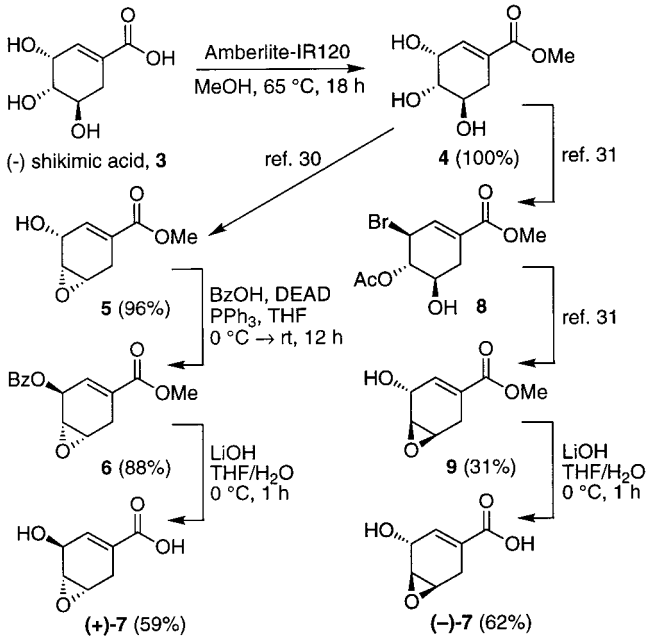



bination allows compounds to be released from the beads in the presence of living cells in aqueous media by exposure to long-wave (365 nm) ultraviolet light.[24−26] An $\omega$-aminocaproic acid (Aca) spacer was coupled after the linker to minimize steric interactions between compounds and the linker.

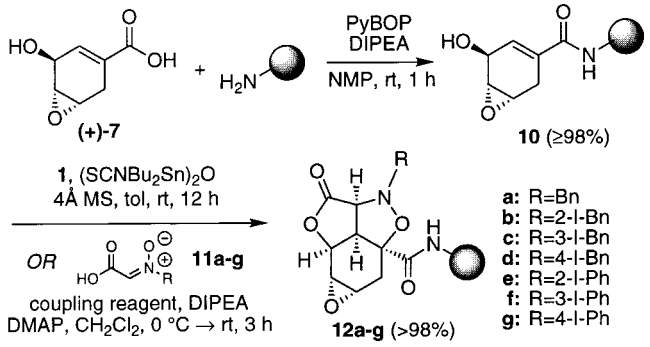
Tamura and co-workers have described the synthesis of a tricyclic compound, **2**, by a tandem transesterification−cycloaddition reaction of nitrone methyl ester **1** and cyclohexen-3-ol (Scheme 1).[27,28] Titanium tetrachloride catalyzes the initial transesterification to a nitrone cyclohexenyl ester intermediate. Presumably, isomerizations to an *s*-cis-like conformation of the ester C−O bond and to the *E* configuration of the nitrone C═N bond are required to allow the appropriate transition state geometry for the 1,3-dipolar cycloaddition reaction to form **2**.[29] The tricyclic structure, synthesized by a modified sequence, was selected as a core template for a library of small molecules.

Both enantiomers of epoxycyclohexenol **7** can be synthesized from the natural product (−)-shikimic acid, **3** (Scheme 2).[30−32] Introduction of an epoxide functionality in the allylic alcohol component of the Tamura tandem reaction was expected to provide a handle for diversification of the cycloadduct or, alternatively, a potentially reactive functional group for interaction with cellular proteins. Thus, **(+)-7** was attached to the solid support using standard coupling conditions to yield **10** (Scheme 3).[33] Because of the epoxide functionality, titanium tetrachloride was initially replaced with a milder organotin transesterification catalyst.[34] Tamura had demonstrated that this catalyst was also effective in mediating the tandem reaction. Exposure of **10** to nitrone ester **1** and the organotin catalyst generated tetracycle **12a** with complete stereo- and regioselectivity. Unfortunately, the solid-phase tandem reaction proceeded sluggishly upon scale-up. Replacement of the transesterification reaction by a direct acylation of **10** with nitrone carboxylic acid **11a**[35] in the presence of PyBroP, DIPEA, and DMAP[36] again generated the

**Scheme 2**




**Scheme 3**




desired tetracycle **12a**. This reaction proceeded satisfactorily upon scale-up although multiple couplings were generally required, presumably due to the sensitivity of the activated nitrone carboxylic acid species.

**Reaction Survey.** The tetracyclic template, **12**, is rigid and densely functionalized. A variety of building blocks[37] can be coupled to the central octahydrobenzisoxazole structure, potentially without the use of protecting groups (Figure 4). The isoxazoline nitrogen can be substituted with various functional groups, installed at either the nitrone or the tetracycle stage. The electrophilic lactone and epoxide can react with nucleophiles while simultaneously unmasking the C-6 and C-5 alcohols for subsequent reactions. Furthermore, reductive N−O bond

(24) Huang, J.; Schreiber, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 13396−13401.

(25) You, A. J.; Jackman, R. J.; Whitesides, G. M.; Schreiber, S. L. *Chem. Biol.* **1997**, *4*, 969−975.

(26) Borchardt, A.; Liberles, S. D.; Biggar, S. R.; Crabtree, G. R.; Schreiber, S. L. *Chem. Biol.* **1997**, *4*, 961−968.

(27) Tamura, O.; Okabe, T.; Yamaguchi, T.; Gotanda, K.; Noe, K.; Sakamoto, M. *Tetrahedron* **1995**, *51*, 107−118.

(28) Tamura, O.; Okabe, T.; Yamaguchi, T.; Kotani, J.; Gotanda, K.; Sakamoto, M. *Tetrahedron* **1995**, *51*, 119−128.

(29) Tamura et al. propose reaction of the *Z*-nitrone via an exo transition state. However, orbital overlap appears to be more favorable in the reaction of the *E*-nitrone via an endo transition state as indicated in Scheme 1.

(30) McGowan, D. A.; Berchtold, G. A. *J. Org. Chem.* **1981**, *46*, 2381−2383.

(31) Wood, H. B.; Ganem, B. *J. Am. Chem. Soc.* **1990**, *112*, 8907−8909.

(32) Mitsunobu, O. *Synthesis* **1981**, 1−28.

(33) All resin-bound products were photocleaved and analyzed by standard techniques (TLC, HPLC, NMR, MS).

(34) Otera, J.; Dan-oh, N.; Nozaki, H. *J. Org. Chem.* **1991**, *56*, 5307−5311.

(35) Keirs, D.; Overton, K. *Heterocycles* **1989**, *28*, 841−848.

(36) Abbreviations not cited in text: DMF = *N*,*N*-dimethylformamide, DMA = *N*,*N*-dimethylacetamide, NMP = *N*-methylpyrrolidinone, PyBroP = bromotripyrrolidinophosphonium hexafluorophosphate, HATU = *N*-[(dimethylamino)-1*H*-1,2,3-triazolo[4,5-*b*]pyridin-1-ylmethylene]-*N*-methylmethanaminium hexafluorophosphate, PyBOP = benzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate, TFFH = *N*,*N*,*N*′,*N*′-tetramethylfluoroformamidinium hexafluorophosphate, DIPC = 1,3-diisopropylcarbodiimide, CDI = 1,1′-carbonyldiimidazole, DIPEA = *N*,*N*-diisopropylethylamine, DMAP = 4-(dimethylamino)pyridine, cod = 1,5-cyclooctadiene, DPPF = 1,1′-bis(diphenylphosphino)ferrocene, DBN = 1,5-diazabicyclo[4.3.0]non-5-ene, LC-MS = tandem high-pressure liquid chromatography−mass spectrometry, HR-FAB-MS = high-resolution fast atom bombardment mass spectrometry, HR-TOF-ESI-MS = high-resolution time-of-flight electrospray induction mass spectrometry, EC-GC = electron capture detection gas chromatography.

(37) The term "building block" is preferred over the commonly used term "monomer" since the products are non-oligomeric.

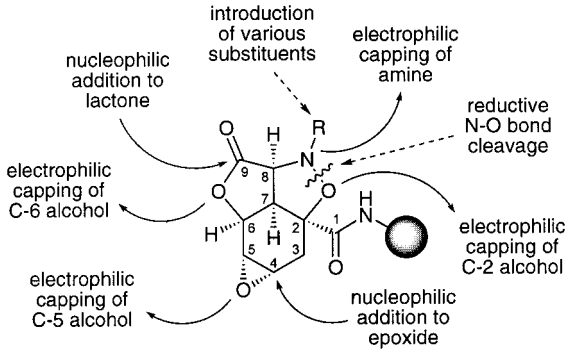
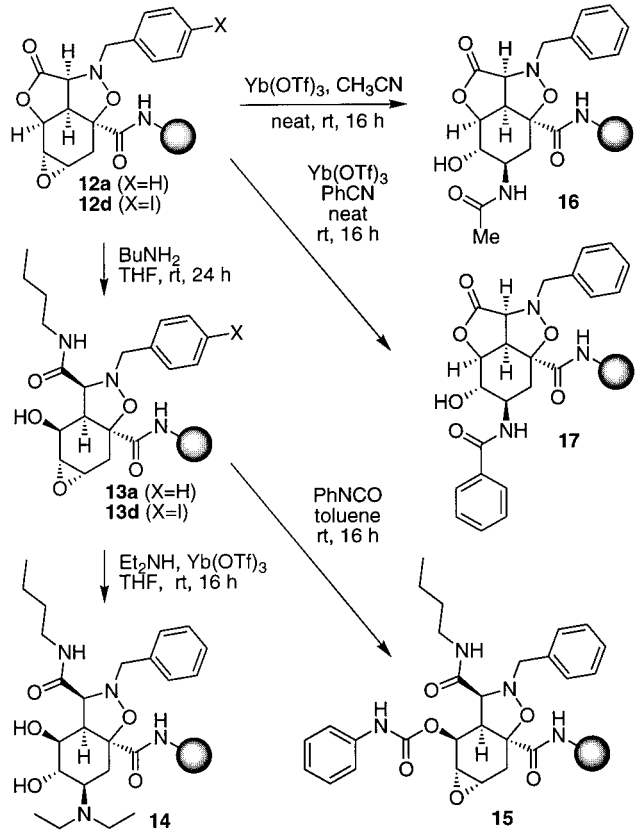



**Figure 4.** Potential building block coupling sites on **12**.

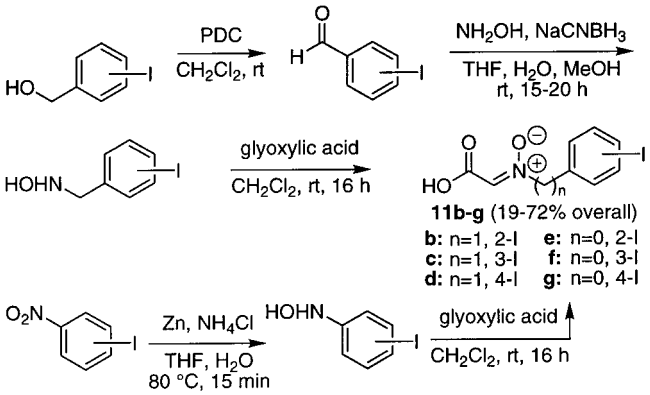
**Scheme 4**




cleavage can provide two additional orthogonal handles for functionalization.

Preliminary experiments showed that lactone **12a** reacted readily with primary amines in THF to form $\gamma$-hydroxyamides such as **13a** (Scheme 4). The liberated C-6 alcohol could then be acylated with phenyl isocyanate to form **15**. Alternatively, the epoxide underwent $Yb(OTf)_3$-mediated aminolysis[38] to form aminodiols such as **14**, although this product was recovered in low yield and purity. Nitriles were used to open the epoxide of **12a** in another $Yb(OTf)_3$-mediated reaction to form **16** and **17**,[39,40] however these compounds were also recovered in low yield. Further investigations indicated that the Geysen linker was prone to $\beta$-elimination under these conditions, highlighting the need for exceedingly mild reaction conditions when using this linker-support combination.[41]

**Scheme 5**

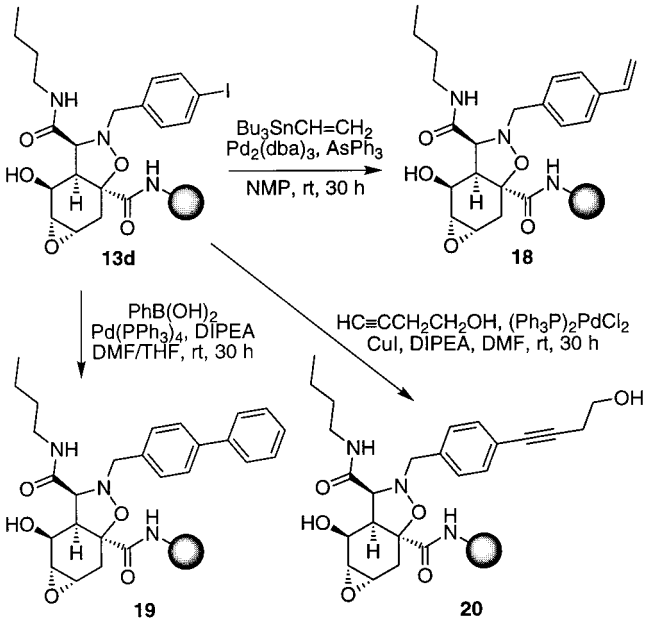



We next considered strategies for introducing substituents at the isoxazolidine nitrogen. The most obvious approach was to perform solution-phase syntheses of a large number of the nitrone carboxylic acids, **11**, then apply these compounds to the solid-phase tandem reaction. However, it is more efficient to synthesize a limited number of nitrone acids that contain a reactive functional group. After tetracycle formation, this group can undergo further solid-phase reactions with readily available building blocks, increasing library diversity without requiring multiple solution-phase syntheses. Aryl iodides were selected because they can be functionalized using palladium-catalyzed reactions. Three iodobenzyl nitrone acids, **11b**−**d**, were readily synthesized from the iodobenzyl alcohols with trituration at the final step sufficient to provide pure compounds (Scheme 5).[35,42,43] Treatment of **10** with **11b**−**d** and PyBroP/DIPEA/DMAP yielded the desired iodobenzyl tetracycles **12b**−**d** (Scheme 3).

The benzyl iodide substituents are amenable to a number of different palladium-catalyzed reactions (e.g., cross couplings, aminations, etherifications, carbonylations). This creates the potential for increased diversity through the application of both "building block-based" and "reaction-based" diversity. The former arises by introducing building blocks with a common functional group at a given site, the latter by using reactions that couple building blocks with different functional groups to the same site. With this approach in mind, Stille,[44,45] Suzuki,[46−49] and Sonogashira/Castro-Stephens[50−54] reactions were all successfully performed on **13d** at room temperature to form **18**−**20** (Scheme 6). Unfortunately, attempts at Heck[55,56] and catalytic

(38) Meguro, M.; Asao, N.; Yamamoto, Y. *J. Chem. Soc., Perkin Trans. 1* **1994**, 2597−2601.
(39) Ryan, R. J.; Julia, S. *Tetrahedron* **1973**, *29*, 3649−3654.
(40) Lindsay Smith, J. R.; Norman, R. O. C.; Stillings, M. R. *J. Chem. Soc., Perkin Trans. 1* **1975**, 1200−1202.
(41) Sternson, S. M.; Schreiber, S. L. *Tetrahedron Lett.* **1998**, *39*, 7451−7454.
(42) Borch, R. F.; Bernstein, M. D.; Durst, H. D. *J. Am. Chem. Soc.* **1971**, *93*, 2897−2904.
(43) Acheson, R. M.; Lee, G. C. M. *J. Chem. Soc., Perkin Trans. 1* **1987**, 2321−2332.
(44) Stille, J. K. *Angew. Chem., Int. Ed. Engl.* **1986**, *25*, 508−524.
(45) Despande, M. S. *Tetrahedron Lett.* **1994**, *35*, 5613−5614.
(46) Miyaura, N.; Suzuki, A. *Chem. Rev.* **1995**, *95*, 2457−2483.
(47) Backes, B. J.; Ellman, J. A. *J. Am. Chem. Soc.* **1994**, *116*, 11171−11172.
(48) Frenette, R.; Friesen, R. W. *Tetrahedron Lett.* **1994**, *35*, 9177−9180.
(49) Guiles, J. W.; Johnson, S. G.; Murray, W. V. *J. Org. Chem.* **1996**, *61*, 5169−5171.
(50) Sonogashira, K.; Tohda, Y.; Hagihara, N. *Tetrahedron Lett.* **1975**, 4467−4470.
(51) Stephens, R. D.; Castro, C. E. *J. Org. Chem.* **1963**, *28*, 3313−3315.
(52) Young, J. K.; Nelson, J. C.; Moore, J. S. *J. Am. Chem. Soc.* **1994**, *116*, 10841−10842.
(53) Collini, M. D.; Ellingboe, J. W. *Tetrahedron Lett.* **1997**, *38*, 7963−7966.
(54) Odingo, J.; Sharpe, B. A.; Oare, D. Presented at the 213th National Meeting of the American Chemical Society, San Francisco, CA, April 1997; ORGN 574.
(55) Heck, R. F. In *Comprehensive Organic Synthesis*; Trost, B. M., Fleming, I., Eds.; Permangon Press: Oxford, 1991; Vol. 4, pp 833−863.

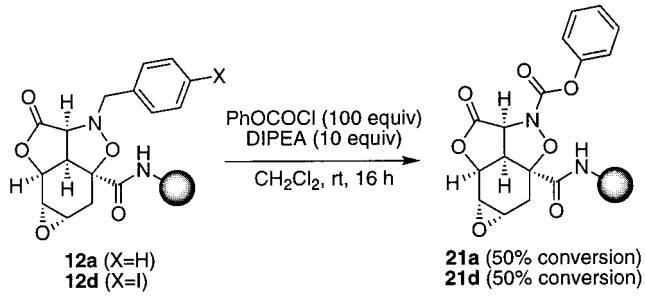
**Scheme 6**




amination[57−59] reactions appeared to result in decomposition of the starting material or $\beta$-elimination of the linker.

The palladium chemistry seemed well-suited for a late step of the synthesis because of the mild, site-selective nature of these reactions. However, when 4-iodobenzyltetracycle **12d** was aminolyzed to $\gamma$-hydroxyamide **13d** (Scheme 4), a significant amount (59%) of epoxycyclohexenol **10** was formed along with the desired product. Reexamination of the aminolysis of the parent desiodotetracycle **12a** to $\gamma$-hydroxyamide **13a** revealed 24% formation of **10**. Whether this side reaction occurred by a true cycloreversion or some other mechanism,[60] it presented a serious challenge because of the stringent requirements for purity and yield ($\geq$90%) in split-pool synthesis. After screening a wide range of acylation catalysts,[61] we found that addition of the tautomeric catalyst 2-hydroxypyridine[62,63] substantially reduced the amount of **10** formed during the reaction, however not to an acceptable level for split-pool synthesis (17% for **13a** and 21% for **13d**).

**Scheme 7**




We investigated other isoxazolidine nitrogen substituents that might reduce or eliminate the aminolysis side reaction problem. Chloroformates have been used to dealkylate tertiary amines, yielding secondary carbamates.[64] Application to the tertiary isoxazoline nitrogen would provide various carbamate products that might not undergo the aminolysis side reaction. Furthermore, use of the appropriate chloroformate would allow subsequent carbamate cleavage to form the free isoxazoline that could be functionalized by alkylation or acylation after lactone aminolysis.[65−68] Treatment of **12a** and **12d** with a large excess of phenylchloroformate resulted in formation of the desired carbamates **21a** and **21d** (Scheme 7). A smaller excess of DIPEA was added to quench adventitious HCl and prevent halohydrin formation at the epoxide. Unfortunately, we were unable to drive this reaction beyond 50% conversion and therefore did not pursue it further.

Three iodophenyl nitrones, **11e**−**g**[69,70] (Scheme 5), and the corresponding iodophenyl tetracycles, **12e**−**g** (Scheme 3), were also synthesized in an effort to overcome the aminolysis side reaction problem.[71] We were encouraged to find that tetracycle **12g** could be aminolyzed with 4-methoxybenzylamine to $\gamma$-hydroxyamide **22** (Scheme 8) with minimal side reaction (<10% yield of **10**). Several other pilot reactions with these iodophenyl tetracycles appeared promising, including the facile epoxide (thio)acidolyisis of **12e** to form **26** and **27** (Scheme 9). Alcohol **27** was readily acylated with propionyl chloride to yield **28**. However, presumably due to the electron-rich nature of the aryl iodides, these iodophenyl compounds proved poor substrates

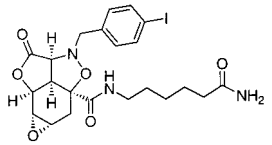
(56) Hiroshige, M.; Hauske, J. R.; Zhou, P. *Tetrahedron Lett.* **1995**, *36*, 4567−4570.

(57) Wolfe, J. P.; Wagaw, S.; Marcoux, J.-F.; Buchwald, S. L. *Acc. Chem. Res.* **1998**, *31*, 805−818.

(58) Hartwig, J. F. *Acc. Chem. Res.* **1998**, *31*, 852−860.

(59) Ward, Y. D.; Farina, V. *Tetrahedron Lett.* **1996**, *37*, 6993−6996.

(60) Analysis of the supernatant from the **12d** → **13d** aminolysis reaction seems to indicate that the side reaction is not a true cycloreversion. Additional experiments indicate that electron-withdrawing substituents on the aromatic ring potentiate this reaction. (For **12** → **13**: R = 4-$CF_3$-Bn, 69% side reaction; R = 4-I-Bn, 59%; R = 4-Cl-Bn, 32%; R = Bn, 24%; R = 4-tBu-Bn, 13%; R = 3-I-Bn, 46%; R = 2-I-Bn, 31%.) Replacement of the $\omega$-aminocaproic acid spacer with glycine or removal of the spacer altogether actually decreases the amount of **10** formed. (For **12**, R = 4-I-Bn: Aca spacer, 59% side reaction; Gly spacer, 36%; no spacer 29%.) Interestingly, substitution at the benzylic carbon eliminates the side reaction altogether (for **12**, R = 4-I-α-Me-Bn, 0% side reaction). Moreover, solution-phase aminolysis of the tetracycle below results in no side reaction. Control experiments verified that the side reaction is not caused by photocleavage.



(61) Aminolysis in the presence of sodium cyanide in MeOH, sodium hydride in DMA, sodium methoxide in benzene, ferric chloride in MeOH, lithium chloride in MeOH, and lithium hydroxide in MeOH all appeared to result in degradation of the linker. Aminolysis in the presence of the Otera organotin catalyst in toluene did not reduce the amount of epoxycyclohexenol recovered.

(62) Rony, P. R. *J. Am. Chem. Soc.* **1969**, *91*, 6090−6096.

(63) Openshaw, H. T.; Whittaker, N. *J. Chem. Soc. C* **1969**, 89−91.

(64) Hobson, J. D.; McCluskey, J. G. *J. Chem. Soc. C* **1967**, 2015−2017.

(65) Montzka, T. A.; Matiskella, J. D.; Partyka, R. A. *Tetrahedron Lett.* **1974**, 1325−1327.

(66) Olofson, R. A.; Schnur, R. C.; Bunes, L.; Pepe, J. P. *Tetrahedron Lett.* **1977**, 1567−1570.

(67) Olofson, R. A.; Martz, J. T.; Senet, J.-P.; Piteau, M.; Malfroot, T. *J. Org. Chem.* **1984**, *49*, 2081−2082.

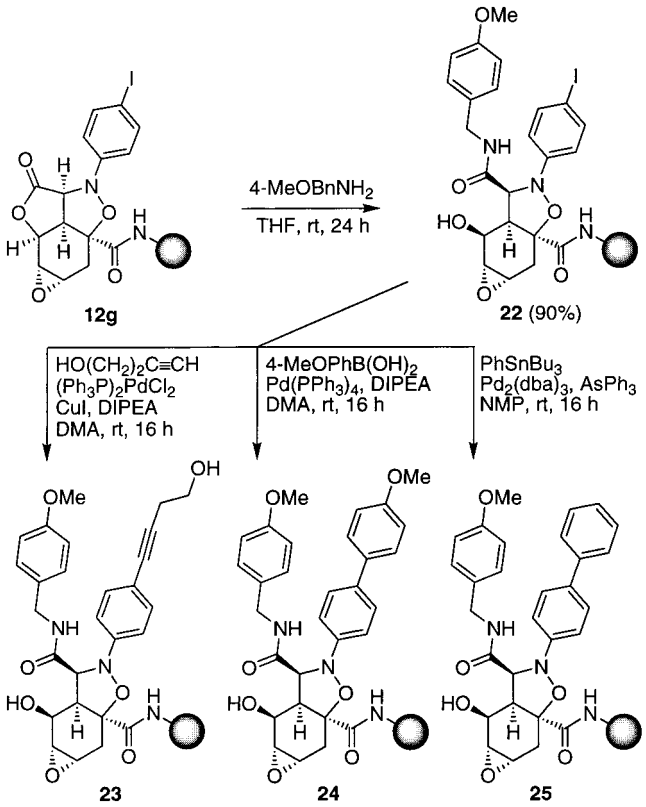
(68) Jung, M. E.; Lyster, M. A. *J. Chem. Soc., Chem. Commun.* **1978**, 315−316.

(69) McClure, K. F.; Danishefsky, S. J. *J. Am. Chem. Soc.* **1993**, *115*, 6094−6100.

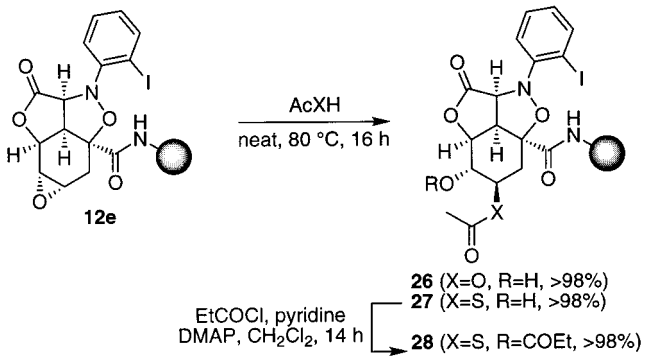
(70) Kamm, O. *Organic Syntheses*; Wiley & Sons: New York, 1941; Collect. Vol. I, pp 445−447.

(71) Use of the appropriate coupling reagent proved critical to the success of these tandem reactions. After extensive screening of coupling conditions, HATU/DMAP was found effective in forming **12e** and **12g**, while only PyBOP/DMAP successfully mediated conversion to **12f**. The following reaction conditions were unsuccessful: (1) in situ activation of acid with PyBroP, DIPEA, ±DMAP, ±4 Å MS, in NMP, DMF, dioxane, or $CH_2Cl_2$; (2) preactivation of acid with oxalyl chloride in $CH_2Cl_2$; (3) preactivation of acid with cyanuric fluoride in $CH_2Cl_2$; (4) in situ activation of acid with DIPC, DMAP, $CH_2Cl_2$; (5) in situ activation of acid with CDI, catalytic pyridine, $CH_2Cl_2$; (6) in situ activation of acid with 2,4,6-trichlorobenzoyl chloride, $CH_2Cl_2$; and (7) esterification with the Otera organotin catalyst.

**Scheme 8**




**Scheme 9**

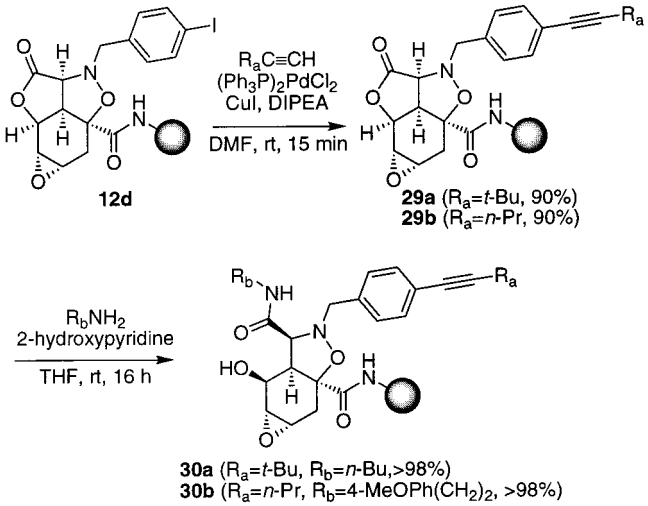



for the palladium-catalyzed reactions and yielded impure products (**23**−**25**).

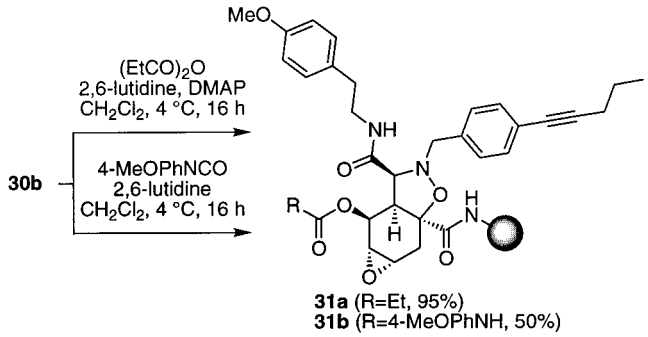
Ultimately, we overcame the aminolysis side reaction problem by changing our reaction sequence. Iodobenzyl tetracycle **12d** was first converted to alkynylbenzyl tetracycles **29a** and **29b** using the Sonogashira/Castro-Stephens reaction (Scheme 10). 2-Hydroxypyridine-mediated lactone aminolysis to $\gamma$-hydroxyamides **30a** and **30b** proceeded effectively without any detectable formation of epoxycyclohexenol **10**.[72]

Lactone aminolysis not only couples an amine diversity element but also simultaneously unmasks the C-6 hydroxyl group that can undergo further reactions. Acid chlorides, anhydrides, sulfonyl chlorides, chloroformates, carbamoyl chlorides, and aryl and alkyl isocyanates were combined with a

**Scheme 10**




**Scheme 11**




variety of acylation catalysts, bases, and solvents in efforts to cap this hydroxyl group.[73] Conditions for efficient reactions with anhydrides (**31a**) and aryl isocyanates (**31b**) were developed (Scheme 11). However, in the latter case, even after extensive resin washing, the photolysis product of carbamate **31b** contained substantial amounts of di(4-methoxyphenyl)urea. Although isocyanate reactions have been reported on polystyrene solid supports,[74−79] it is possible that the poly(ethyleneglycol) portion of the TentaGel polymer prevented effective removal of the urea byproduct.

The alkynes introduced by the Castro-Stephens reaction provide a potential handle for coupling of a fourth diversity element. Considering the mild reaction conditions required by our linker−support combination, we investigated a recently reported rhodium-catalyzed hydroacylation reaction.[80] Since an ample number of salicyl aldehydes are commercially available, this reaction could generate substantial additional diversity.[81] Exposure of propionate **31a** to salicyl aldehyde in the presence of $[Rh(cod)Cl]_2$ and DPPF yielded the expected regioisomers

(72) Despite considerable efforts to understand this vexing side reaction, both its exact mechanism and a rationalization of the ameliorating effects of alkynylbenzyl substituents remain mysterious to us.

(73) Where appropriate, DMAP, $PBu_3$ (Vedejs, E.; Diver, S. T. *J. Am. Chem. Soc.* **1993**, *115*, 3358−3359), and Cu(I)Cl (Duggan, M. E.; Imagire, J. S. *Synthesis* **1989**, 131−132) were assayed as acylation catalysts; DIPEA, pyridine, and 2,6-lutidine were assayed as bases; $CH_2Cl_2$, THF, and DMF were assayed as solvents.

(74) Hobbs DeWitt, S.; Kiely, J. S.; Stankovic, C. J.; Schroeder, M. C.; Reynolds Cody, D. M.; Pavia, M. R. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6909−6913.

(75) Kick, E. K.; Ellman, J. A. *J. Med. Chem.* **1995**, *38*, 1427−1430.

(76) Kurth, M. J.; Ahlberg Randall, L. A.; Takenouchi, K. *J. Org. Chem.* **1996**, *61*, 8755−8761.

(77) Beebe, X.; Schore, N. E.; Kurth, M. J. *J. Org. Chem.* **1995**, *60*, 4196−4203.

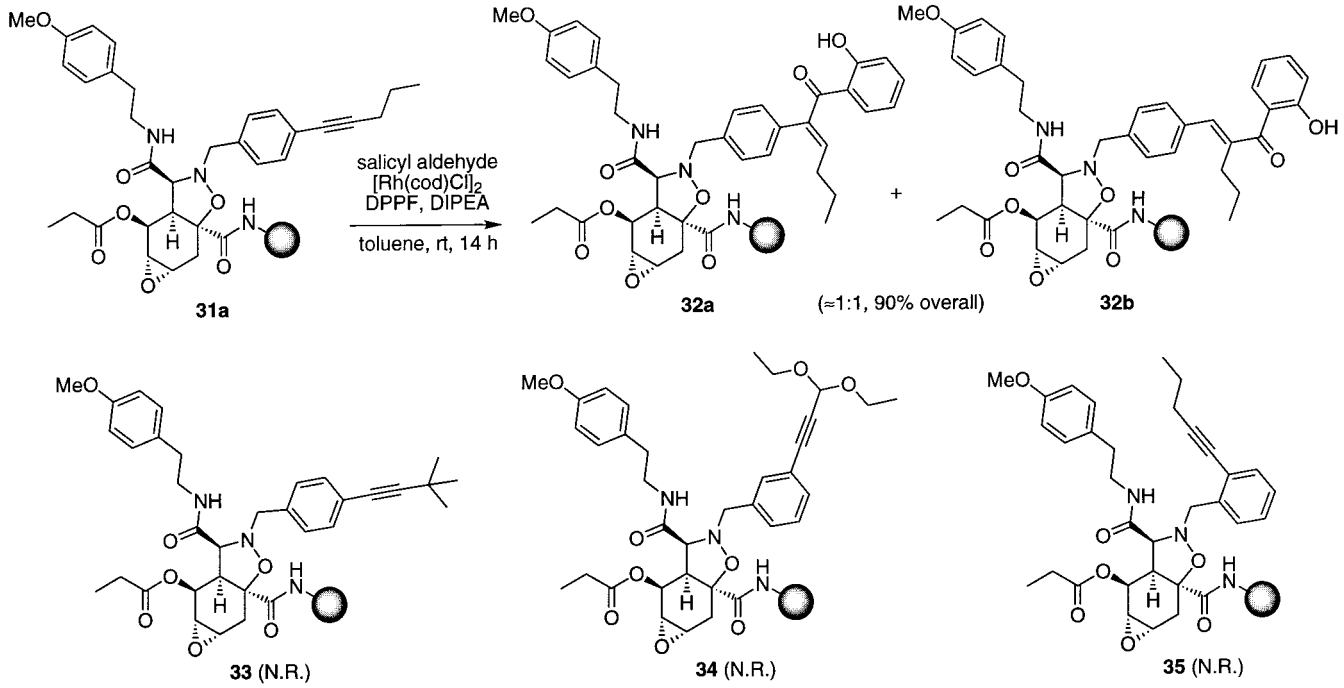
(78) Beebe, X.; Schore, N. E.; Kurth, M. J. *J. Am. Chem. Soc.* **1992**, *114*, 10061−10062.

(79) Meyers, H. V.; Dilley, G. J.; Durgin, T. L.; Powers, T. S.; Winssinger, N. A.; Zhu, H.; Pavia, M. R. *Mol. Div.* **1995**, *1*, 13−20.

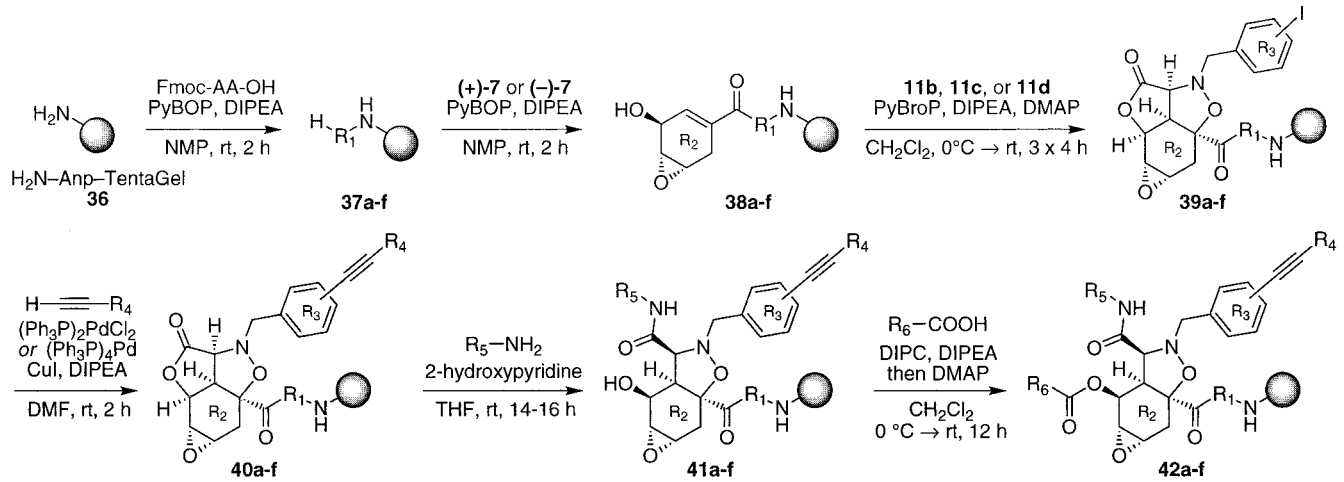
(80) Kokubo, K.; Matsumasa, K.; Miura, M.; Nomura, M. *J. Org. Chem.* **1997**, *62*, 4564−4565.

(81) The 1996−1997 Aldrich Structure Index lists 24 applicable salicyl aldehydes.

**Scheme 12**




**Scheme 13**




| | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ | $R_6$ |
|---|---|---|---|---|---|---|
| **a:** | Ø | **(+)-7** | *ortho* | benzyl | cyclobutyl | 4-methoxybenzyl |
| **b:** | Ø | **(+)-7** | *meta* | isopropenyl | allyl | phenyl |
| **c:** | Ø | **(+)-7** | *para* | 4-chlorophenyl | 2-methoxybenzyl | isopropyl |
| **d:** | Aca | **(+)-7** | *ortho* | *t*-butyl | 4-methoxybenzyl | isobutyl |
| **e:** | Aca | **(+)-7** | *meta* | diethoxymethyl | 2,2-dimethoxyethyl | 1-propenyl |
| **f:** | Aca | **(+)-7** | *para* | propyl | 4-methoxyphenethyl | ethyl |

**32a** and **32b** in approximately a 1:1 ratio (Scheme 12). However, this reaction proved quite sensitive to sterics as compounds **33**−**35** all failed to react under these conditions.

**Reaction Optimization.** The three highly efficient Sonogashira/Castro-Stephens, lactone aminolysis, and esterification reactions offered the opportunity to generate a library of substantial size and diversity. Additional diversity could be generated by the use of the three iodobenzyl nitrone isomers **11b**−**d**, both enantiomers of epoxycyclohexenol **7**, and several different spacer elements between the photolinker and the library (Scheme 13). These reactions were studied in detail to define the scope, limitations, and optimal conditions for each.

Initial efforts at the Sonogashira/Castro-Stephens reaction (**39** → **40**) yielded approximately 20% of an unidentified low molecular weight byproduct. However, shorter reaction times (15−120 min instead of 16 h) reduced the amount of this byproduct to an acceptable 5−10% while still resulting in complete conversion of **39** to **40**. Additionally, terminal diynes were coupled efficiently by using $(PPh_3)_4Pd$ as the palladium source and a larger excess of diyne (50 equiv instead of 20 equiv for monoynes) to avoid intrabead cross-linking. Attempts to perform this reaction with catalytic amounts of palladium or copper iodide resulted in slow reaction rates.

Preliminary studies had indicated that inclusion of 2-hydroxypyridine in the lactone aminolysis reaction (**40** → **41**) yielded the $\gamma$-hydroxyamide products in higher purity. However, this modification led to epoxide aminolysis as a competing side reaction. Application of 25 equiv of amine, 5 equiv of 2-hydroxypyridine, and shorter reaction time (14−16 h instead of 20−24 h) resulted in complete conversion with little or no evidence of epoxide aminolysis. α-Branched amines could also open the lactone efficiently when applied in larger excess (50

equiv of amine, 10 equiv of 2-hydroxypyridine). Unfortunately, we were unable to develop conditions for lactone opening reaction with secondary amines, anilines, *N*,*N*-dialkylhydrazines, *O*-alkylhydroxylamines, hydrazides (R-CO-$NHNH_2$), carbazates (RO-CO-$NHNH_2$), or (thio)carbazides (RHN-C(O/S)-$NHNH_2$). Exposure of the tetracycle to hydrazine in MeOH appeared to result in degradation of the photolinker.

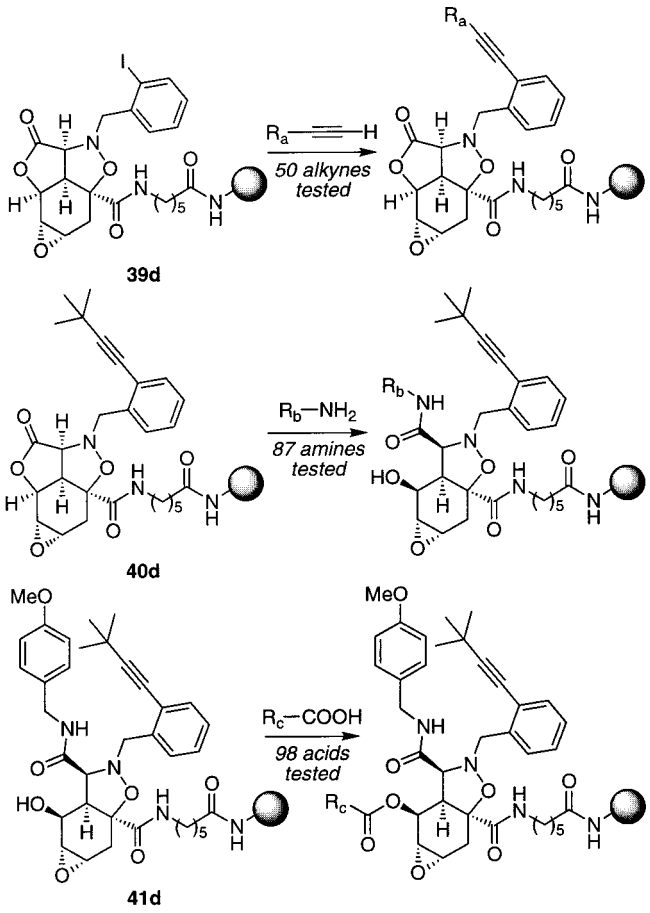
A limited number of symmetrical anhydrides are commercially available and useful for the esterification reaction (**41** → **42**).[82] Thus, a general method of coupling carboxylic acids to the C-6 hydroxyl group was desired. After surveying several conditions,[83] we found that the reaction proceeded efficiently when DIPC and DIPEA were used to preactivate the carboxylic acids as symmetrical anhydrides that were subsequently added to the resin with DMAP.[84] Straight-chain, α-branched, conjugated, and aromatic carboxylic acids were all coupled successfully under these conditions. However, we were unable to find conditions that would allow coupling of *N*-Boc- or *N*-Fmoc-protected amino acids. Completion of these reaction optimization experiments set the stage for the detailed library validation studies required prior to library synthesis.

**Library Validation Protocols.** Split-pool synthesis provides the theoretical means to synthesize the full matrix of every combination of building blocks in a multistep synthesis. Such large numbers of molecules will likely be required for successful outcomes in chemical genetic screens. However, these syntheses present enormous analytical challenges. We have developed a four-stage validation protocol to provide maximum confidence that a complex, split-pool synthesis of encoded molecules yields the anticipated products in high purity and efficiency. The first three protocols are concerned with the synthetic molecules and the fourth with the encoding step.

First, the suitability of the reaction sequence for library synthesis was demonstrated by execution of the entire reaction sequence six times, each time using different building blocks. The fully elaborated final products, **42a**−**f**, were recovered in 80−90% purity following photolysis. These products, as well as the 20 intermediates preceding them (**38a**, **38d**, **39a**−**f**, **40a**−**f**, **41a**−**f**), were fully characterized by multidimensional $^1$H NMR, HR-FAB-MS, TLC, and HPLC. This experiment showed that the reaction sequence could be used to synthesize library members in satisfactory purity.

Second, potential building blocks were tested by reaction with a selected substrate at each step (Scheme 14). While it is impossible to test the complete matrix of building block combinations, this experiment indicated which building blocks are compatible with the coupling reactions. Thus, 50 alkynes (Figure 5) were tested in reactions with iodobenzyltetracycle **39d**, 87 amines (Figure 6) in reactions with alkynylbenzyltetracycle **40d**, and 98 acids (Figure 7) in reactions with γ-hydroxyamide **41d**. Nearly every commercially available terminal alkyne was tested, along with a variety of amines and acids representing different steric and electronic functional groups. Photocleavage products were analyzed by HPLC and LC-MS[85] and their purities and percent conversions were estimated from these data (Figure 8). There were no obvious trends among the

**Scheme 14**




alkynes that were unsuitable for the Sonogashira/Castro-Stephens reaction; however, in the aminolysis and esterification reactions, electron-poor amines and electron-rich or enolizable acids generally did not react with suitable efficiency. In addition, several of the acids were insoluble under the reaction conditions. Of the building blocks tested, 23 alkynes, 54 amines, and 44 acids reacted with ≥90% conversion and purity. These building blocks, along with a limited number of less optimal candidates (generally reacting with ≥70% conversion and purity), were selected for inclusion in library synthesis.

Third, a small test library was generated from iodobenzyltetracycle **39b** (Figure 9) to investigate whether any unforeseen complications, such as interactions between building blocks coupled at different sites, might arise during synthesis in a split-pool format. The building blocks were carefully selected such that every product within each final acylated pool would have a unique mass (Figure 10), allowing analysis by LC-MS. Thus, the tetracycle-containing resin was divided into eight portions and the seven alkynes were coupled to the first seven portions. The eighth portion was left as the parent aryl iodide, representing a "skip codon".[86] After pooling and splitting, the seven amines were coupled and the eighth portion of resin was left as the lactone-closed skip codon. Finally, after a third round of pooling and splitting, the seven acids were coupled and the eighth portion was left as the free C-6 hydroxyl skip codon. Because all eight final pools, designated **43**{**X,X,1**} through **43**{**X,X,8**},[87] con-

(82) The 1996−1997 Aldrich Structure Index lists 54 applicable symmetrical anhydrides compared to over 1000 carboxylic acids.

(83) DIPC (0.5 and 1.0 equiv relative to acid), HATU, and TFFH (Carpino, L. A.; El-Faham, A. *J. Am. Chem. Soc.* **1995**, *117*, 5401−5402) were assayed as coupling reagents with DIPEA in $CH_2Cl_2$, both with and without DMAP.

(84) Atuegbu, A.; Maclean, D.; Nguyen, C.; Gordon, E. M.; Jacobs, J. W. *Bioorg. Med. Chem.* **1996**, *4*, 1097−1106.

(85) Certain acid-sensitive products were also analyzed by TLC and FAB-MS.

(86) 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.

(87) Test library compounds are designated as **43**{$\mathbf{R_a,R_b,R_c}$} where $R_a$ signifies the alkyne building block, $R_b$ signifies the amine building block, and $R_c$ signifies the acid building block. Pools of compounds are signified by $R_n$ = X, where X represents all eight building blocks at a given position.

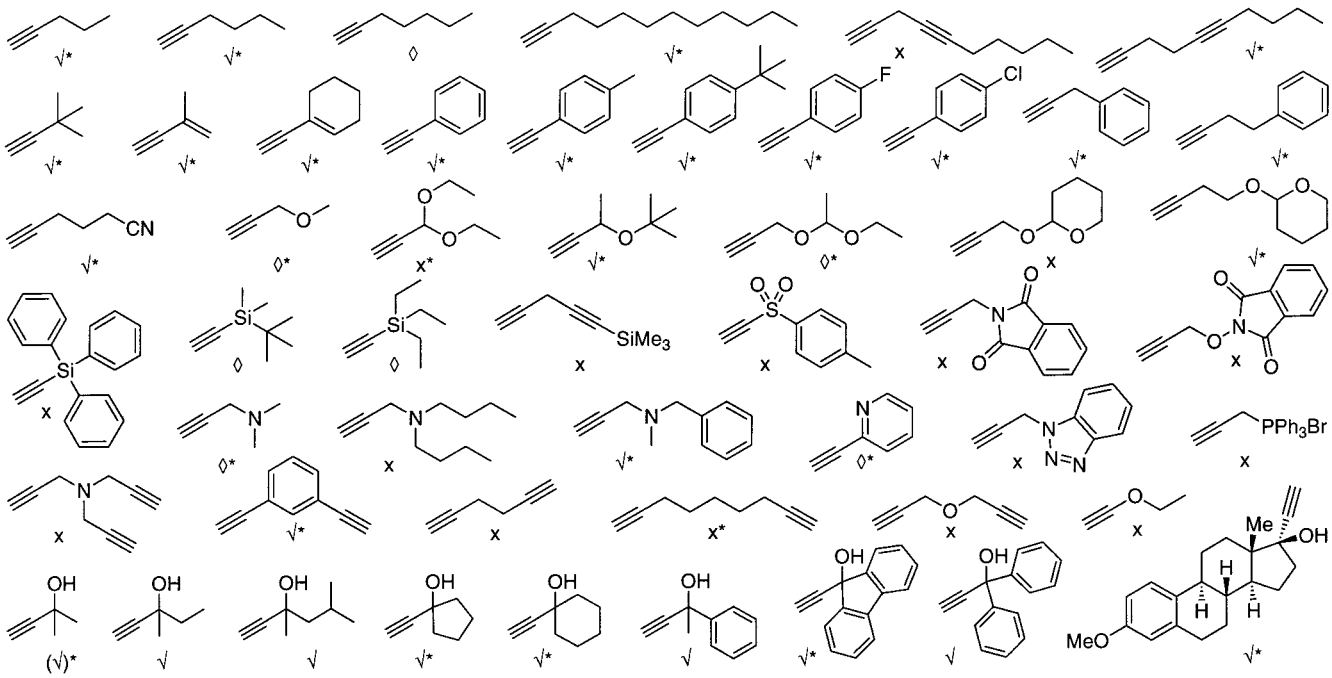


**Figure 5.** Alkyne building blocks tested in Sonogashira/Castro-Stephens reaction. Building blocks reacting with ≥80% conversion and purity are denoted by √, 50−80% by ◊, <50% by ×. Building blocks included in the full-scale library are denoted by *.

tained the same eight γ-butyrolactone compounds corresponding to the aminolysis skip codon, a total of 456 compounds was generated.

Each pool was photocleaved to yield a mixture of 64 compounds that were analyzed by LC-MS (Figure 11). Of the 456 expected masses, all 456 (100%) were detected at some level, 418 (92%) were detected at ≥10% of the average intensity for the given pool, and 400 (88%) were detected at ≥20% of the average intensity for the given pool. All of the weak signals resulted from compounds having one of two building blocks at the amine position. 1-(3-Aminopropyl)-2-pyrrolidinone (**Amine 6**) is known to cyclize to DBN with loss of $H_2O$. Since strong bases had been found to be incompatible with our linker−support combination, this building block was excluded from full-scale library synthesis. The skip codon (**Amine 1**) left lactone-closed tetracycles that were partially hydrolyzed during the final acylation step. As a result, during full-scale library synthesis, the aminolysis skip codon pool was set aside before the final pooling, splitting, and acylation steps.

**Binary Encoding.** Assuming an ideal, efficient split-pool synthesis, each support carries a single compound. Several solutions to the problem of compound identification have been developed, falling into two general categories: recursive deconvolution and encoding.[88] Since recursive deconvolution requires several rounds of resynthesis, we chose the particularly powerful *binary* encoding strategy. This method has been used to discover small molecules that bind to carbonic anhydrase,[89,90] SH3 domains, [86,91,92] angiotensin-converting enzyme,[93] *Bauhinia purpurea* lectin,[94] matrilysin,[95] and aspartyl proteases.[96] Still's polyhaloaromatic EC-GC tags, **44**, were selected since they are relatively unreactive and can be coupled directly to the polystyrene backbone of beads using mild carbene insertion chemistry (Scheme 15).[19,97]

Unfortunately, the published procedures gave inconsistent and unsatisfactory results in our hands. Substitution of the reported rhodium bis(trifluoroacetate) catalyst with a bulkier rhodium bis(triphenylacetate) catalyst[98] suppressed solution-phase diazoketone dimerization and substantially improved the efficiency of tag-bead coupling to form cycloheptatrienes **45**. We also found that, after reaction with an initial set of tags, attachment of subsequent tags to the same beads required multiple couplings. It is possible that the initial reactions occurred at the most accessible sites in the polymer, making subsequent reactions more difficult. Finally, the reaction conditions for oxidative cleavage of the tags from **45** with ceric ammonium nitrate (CAN) were optimized, reducing the required cleavage time to 10 min from the reported 4 h. This improved the yields of the polyhaloaromatic alcohol products, **46**, and allowed rapid and consistent analysis by EC-GC.

**Full-Scale Library Planning and Synthesis.** Completion of the validation protocols above set the stage for full-scale encoded library synthesis. First, building blocks were selected for each step of the synthesis (Scheme 13). ω-Aminocaproic acid and glycine were selected as spacer elements with the "no spacer" skip codon providing a third structure for **37**. Use of both enantiomers of epoxycyclohexenol **7** resulted in six structures

(88) Czarnik, A. W. *Curr. Opin. Chem. Biol.* **1997**, *1*, 60−66.
(89) Burbaum, J. J.; Ohlemeyer, M. H. J.; Reader, J. C.; Henderson, I.; Dillard, L. W.; Li, G.; Randle, T. L.; Sigal, N. H.; Chelsky, D.; Baldwin, J. J. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6027−6031.
(90) Baldwin, J. J.; Burbaum, J. J.; Henderson, I.; Ohlemeyer, M. H. J. *J. Am. Chem. Soc.* **1995**, *117*, 5588−5589.
(91) Kapoor, T. M.; Hamilton Andreotti, A.; Schreiber, S. L. *J. Am. Chem. Soc.* **1998**, *120*, 23−29.
(92) Morken, J. P.; Kapoor, T. M.; Feng, S.; Shirai, F.; Schreiber, S. L. *J. Am. Chem. Soc.* **1998**, *120*, 30−36.
(93) Maclean, D.; Schullek, J. R.; Murphy, M. M.; Ni, Z.-J.; Gordon, E. M.; Gallop, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 2805−2810.
(94) Liang, R.; Yan, L.; Loebach, J.; Ge, M.; Uozumi, Y.; Sekanina, K.; Horan, N.; Gildersleeve, J.; Thompson, C.; Smith, A.; Biswas, K.; Still, W. C.; Kahne, D. *Science* **1996**, *274*, 1520−1522.
(95) Schullek, J. R.; Butler, J. H.; Ni, Z.-J.; Chen, D.; Yuan, Z. *Anal. Biochem.* **1997**, *246*, 20−29.
(96) DiIanni Carroll, C.; Johnson, T. O.; Tao, S.; Lauri, G.; Orlowski, M.; Gluzman, I. Y.; Goldberg, D. E.; Dolle, R. E. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 3203−3206.
(97) Nestler, H. P.; Bartlett, P. A.; Still, W. C. *J. Org. Chem.* **1994**, *59*, 4723−4724.
(98) Callot, H. J.; Metz, F. *Tetrahedron* **1985**, *41*, 4495−4501.

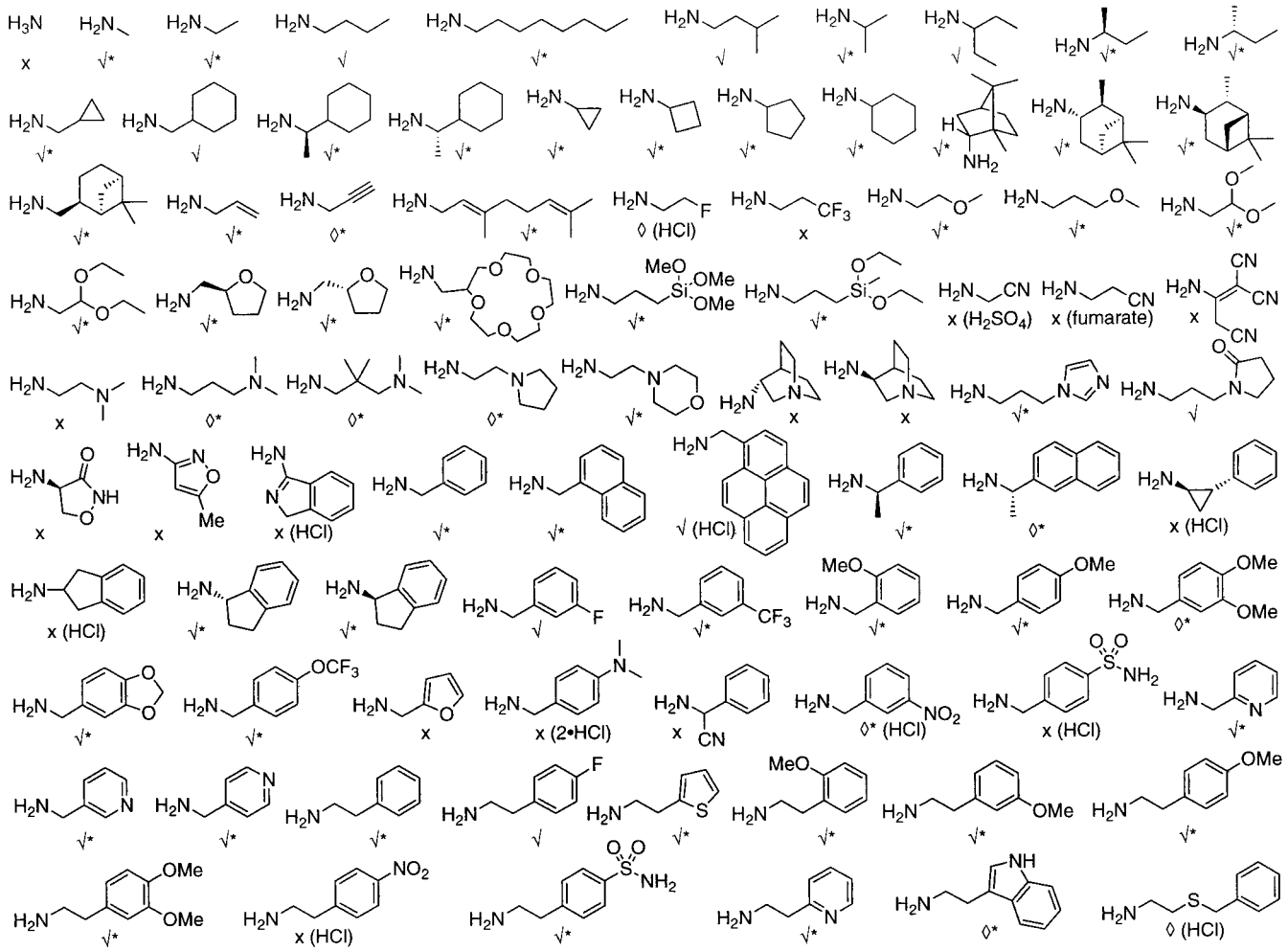


**Figure 6.** Amine building blocks tested in lactone aminolysis reaction. Salts indicated in parentheses were neutralized in situ with DIPEA. Building blocks reacting with ≥80% conversion and purity are denoted by √, 50−80% by ◊, <50% by ×. Building blocks included in the full-scale library are denoted by *.

for **38**. Inclusion of all three iodobenzyl nitrone carboxylic acids **11b**−**d** led to 18 iodobenzyltetracycle structures for **39**.

The three remaining diversity positions, corresponding to the Sonogashira/Castro-Stephens, lactone aminolysis, and C-6 acylation reactions, allowed the use of substantially larger numbers of building blocks. Optimal use of the binary encoding tags dictates that $2^n - 2$ building blocks should be used at a given position. This accounts for one skip codon (the "all one" code) and allows for exclusion of the undesirable "all null" code that cannot be differentiated from a failed tagging reaction. As a practical matter, coupling of up to $2^6 - 2 = 62$ building blocks at a given step was deemed feasible.

Only $2^5 - 2 = 30$ building blocks were selected for the Sonogashira/Castro-Stephens reaction because of the relatively small number of available terminal alkynes. Most of the alkynes that reacted efficiently during building block screening (Figure 5) were selected. Several racemic alkynes were also included, although diastereomeric products would likely result. Furthermore, several alkynols were included, despite their potential reactivity in the final acylation reaction. Control experiments indicated that these hydroxyl groups were efficiently acylated by a variety of alkyl and aromatic acids under the DIPC-mediated coupling conditions. Coupling of 30 different alkynes with exclusion of a 31st skip codon portion would result in 558 structures for **40**.

Wider selections of building blocks were available at the amine and acid positions. For each reaction, 62 building blocks were selected, representing a range of sizes and functional groups (Figures 6 and 7). Coupling of 62 different amines with exclusion of a 63rd aminolysis skip codon portion would result in 35154 structures for **41**. As discussed above, the 558 lactone-closed compounds corresponding to the aminolysis skip codon would not react at the final acylation step. Therefore, the total number of final library compounds, **42**, resulting from acylation with 62 acids and exclusion of a 63rd skip codon portion is calculated as follows: $[(62 \times 558 = 34596) \times 63 = 2179548] + 558 = 2180106$ compounds.

Synthesis of three copies of the library was planned, based upon a calculation that indicates that three copies should be screened to ensure 95% confidence that every compound has been sampled.[99] Although this calculation does not address the number of copies required to ensure that every possible compound has been synthesized, we recognized that, if necessary, the library could be resynthesized on larger scale in the future.

Library synthesis began with coupling of Fmoc-protected Geysen linker to 90 $\mu$m TentaGel S $NH_2$. After deprotection, the resin, **36**, was split into three portions by weight and labeled with the tags corresponding to the spacer position. For the fourth validation protocol, after each tagging step in the synthesis, several beads were removed from every portion of the resin and the tags were cleaved and analyzed to ensure adequate

(99) Nolan, G. P., FACS Screening Web Page. http://www.stanford.edu/group/nolan/FACSscrn.html (accessed Jun 1999).

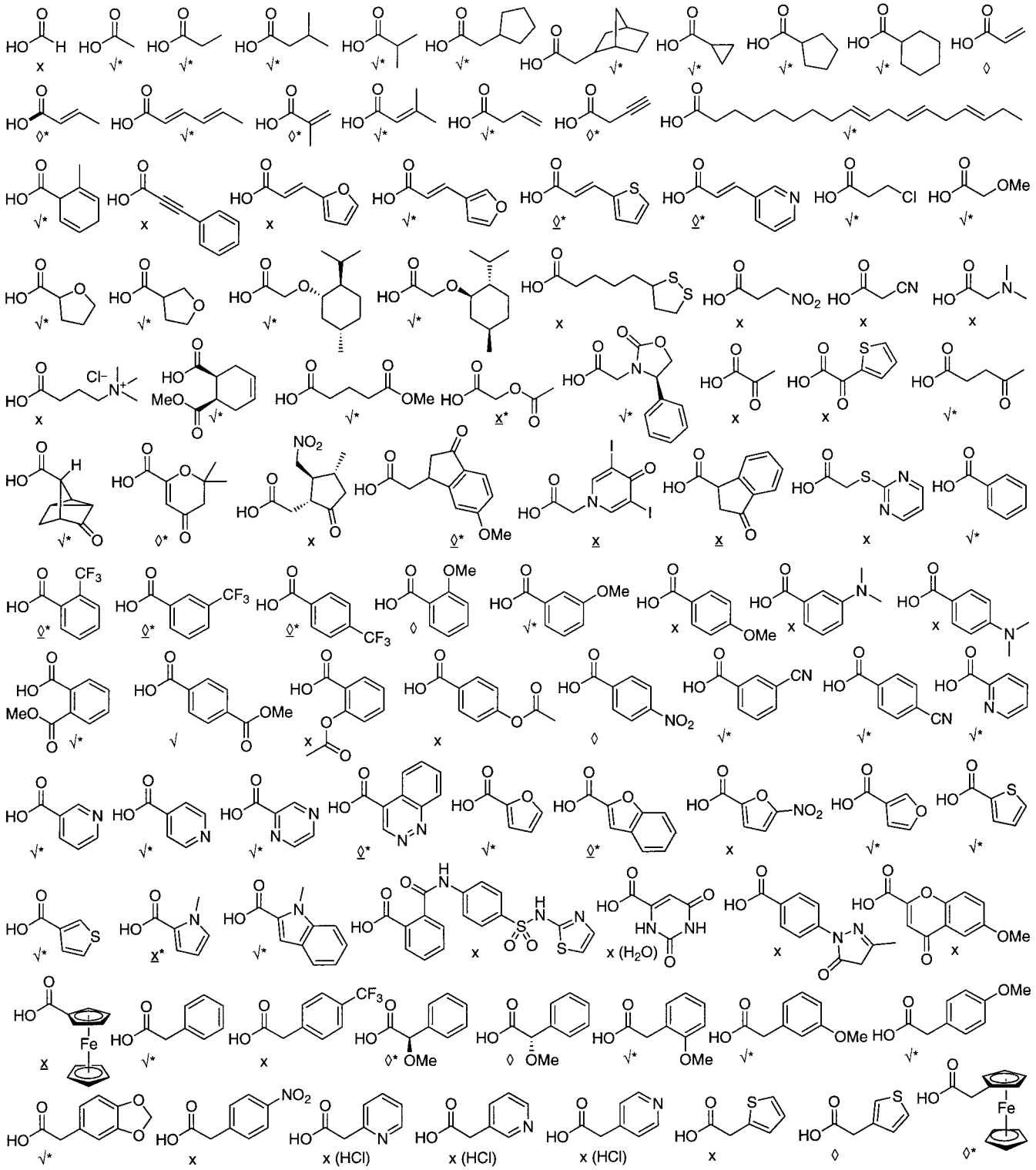


**Figure 7.** Carboxylic acid building blocks tested in acylation reaction. Building blocks reacting with ≥80% conversion and purity are denoted by √, 50−80% by ◊, <50% by ×. Underlining indicates starting material **41d** was of questionable purity. Building blocks included in the full-scale library are denoted by *.

incorporation levels. Fmoc-Aca-OH and Fmoc-Gly-OH were then coupled to two of the portions and deprotected under standard conditions. The resin, **37**, was pooled, mixed, and split into two equal portions. After tagging, one enantiomer of epoxycyclohexenol **7** was coupled to each portion. Resin **38** was then pooled, mixed, and split into three equal portions for tagging and reactions with iodobenzyl nitrones **11b**−**d** to yield resin **39**. These tetracycle-containing resins were pooled, mixed, and split into 31 equal portions. Each was tagged and the appropriate 30 terminal alkynes were coupled using the Sonogashira/Castro-Stephens reaction. The 31st portion of resin was set aside as the skip codon. Resin **40** was pooled, mixed, and split into 63 portions. In this case, the 63rd portion of resin, corresponding to the aminolysis skip codon, was 1/63rd the size of the other 62 portions. This modification was required to avoid overrepresentation of lactone-closed compounds in the completed library. The 62 large portions of resin were tagged and reacted with the appropriate amines to yield resin **41**. The 63rd

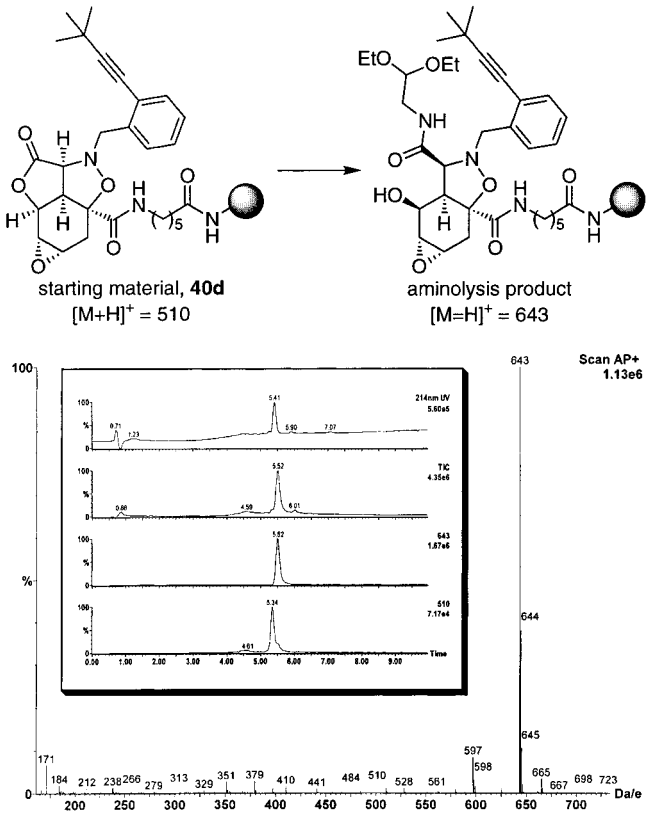



**Figure 8.** Representative LC-MS data for testing of building blocks. Top: Starting material and product structures. Bottom: Mass spectrum of product peak. Inset: 214 nm UV trace, Total Ion Count (TIC) trace, product mass trace, and starting material mass trace. While the starting material is difficult to detect in the UV and TIC traces, a small amount is clearly seen in the single mass trace at 4% relative intensity compared to the product. Note the slight (0.09 min) delay between the UV detector and mass detector retention times.

skip codon portion was set aside for the remainder of the synthesis to avoid hydrolysis of the lactone-closed compounds during the final acylation step. The remaining 62 portions of resin were pooled, mixed, and split into 63 equal portions. After tagging, the appropriate 62 acids were coupled and the 63rd portion of resin was left as the unreacted C-6 hydroxy compounds. Finally, the 63 acylation portions and the lactone aminolysis skip codon portion were pooled and mixed to yield the final library **42**, calculated to contain three copies of 2180106 compounds. The entire process was completed by two of us (D.S.T. and M.A.F.) working over a period of 3 weeks. The bulk of this time was spent verifying that the encoding tags had successfully coupled to every portion of the resin during each step of the synthesis.

**Cell Permeation and Pathway Modulation.** It seemed worthwhile to begin an analysis of these compounds by screening the 456-compound test library (Figure 9) in cellular assays even before the full-scale library was completed. Although our compounds were designed to contain structural features common to natural products, we had no general sense of their ability to either permeate cells or alter cellular pathways.

These initial studies relied upon traditional nonminiaturized assays, requiring more material than was contained on a single synthesis bead. Thus, the test library was screened as mixtures of compounds cleaved in bulk. The eight final pools, **43**{**X,X,1**} through **43**{**X,X,8**}, each contained 64 compounds. The compounds in each pool were photolyzed from the resin, recovered, and dissolved in DMSO at an estimated concentration of 1 mM per compound (64 mM overall).[100] The eight pools, assayed at

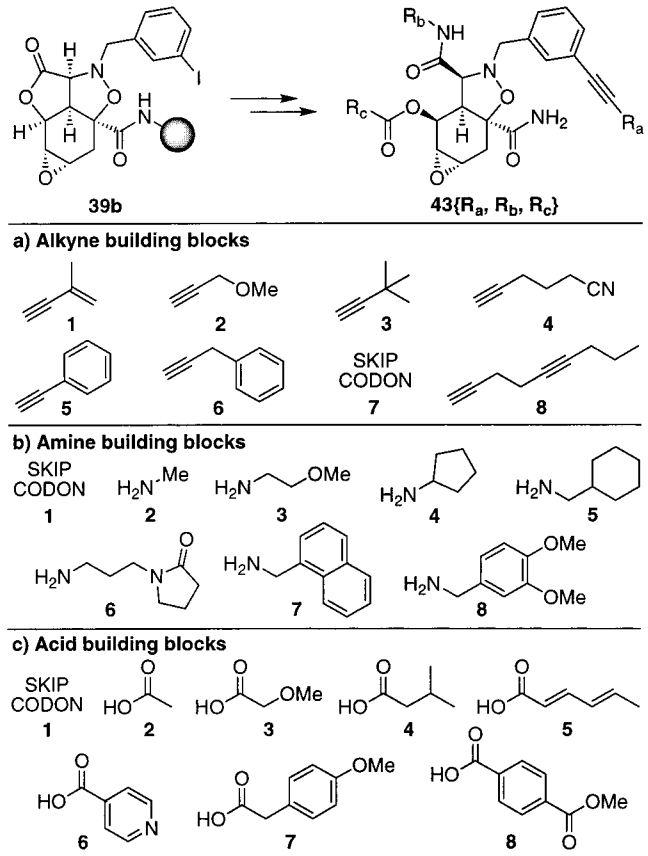



**Figure 9.** Tetracycle and building blocks used in the test library.

**Amines**

| Alkynes | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| **MW** | 0 | 31 | 75 | 85 | 113 | 142 | 157 | 167 |
| **1** | 66 | 381 | 412 | 456 | 466 | 494 | 523 | 538 | 548 |
| **2** | 70 | 385 | 416 | 460 | 470 | 498 | 527 | 542 | 552 |
| **3** | 82 | 397 | 428 | 472 | 482 | 510 | 539 | 554 | 564 |
| **4** | 93 | 408 | 439 | 483 | 493 | 521 | 550 | 565 | 575 |
| **5** | 102 | 417 | 448 | 492 | 502 | 530 | 559 | 574 | 584 |
| **6** | 116 | 431 | 462 | 506 | 516 | 544 | 573 | 588 | 598 |
| **7** | 128 | 443 | 474 | 518 | 528 | 556 | 585 | 600 | 610 |
| **8** | 134 | 449 | 480 | 524 | 534 | 562 | 591 | 606 | 616 |

**Figure 10.** Alkyne and amine building block masses and the resulting 64 unique $\gamma$-hydroxyamide product masses. Acylation of the C-6 alcohol with a carboxylic acid shifts all of the product masses for that pool by the same value (mass of the acid minus water).

concentrations up to 10 $\mu$M per compound, showed no suppression of rapamycin-based growth inhibition in *S. cerevisiae*.[101] In addition, none of the pools, assayed at concentrations up to 12.5 $\mu$M per compound, showed inhibitory activity in a *Xenopus laevis* oocyte extract assay that indicates modulation of the cyclin B degradation pathway.[102]

However, all eight pools showed a significant inhibitory effect on mink lung cell proliferation when assayed at a concentration of 1 $\mu$M per compound (Figure 12). Moreover, when the library was assayed at a concentration of 250 nM per compound, pool **43**{**X,X,8**} (Figure 13), was found to activate a TGF-$\beta$-

(100) Based upon previous results, a 50% yield was assumed.
(101) The rapamycin concentration was 100 nM. J. Huang and S. L. S., unpublished results.
(102) R. W. King, unpublished results.
(103) Carcamo, J.; Zentella, A.; Massague, J. *Mol. Cell. Biol.* **1995**, *15*, 1573−1581.

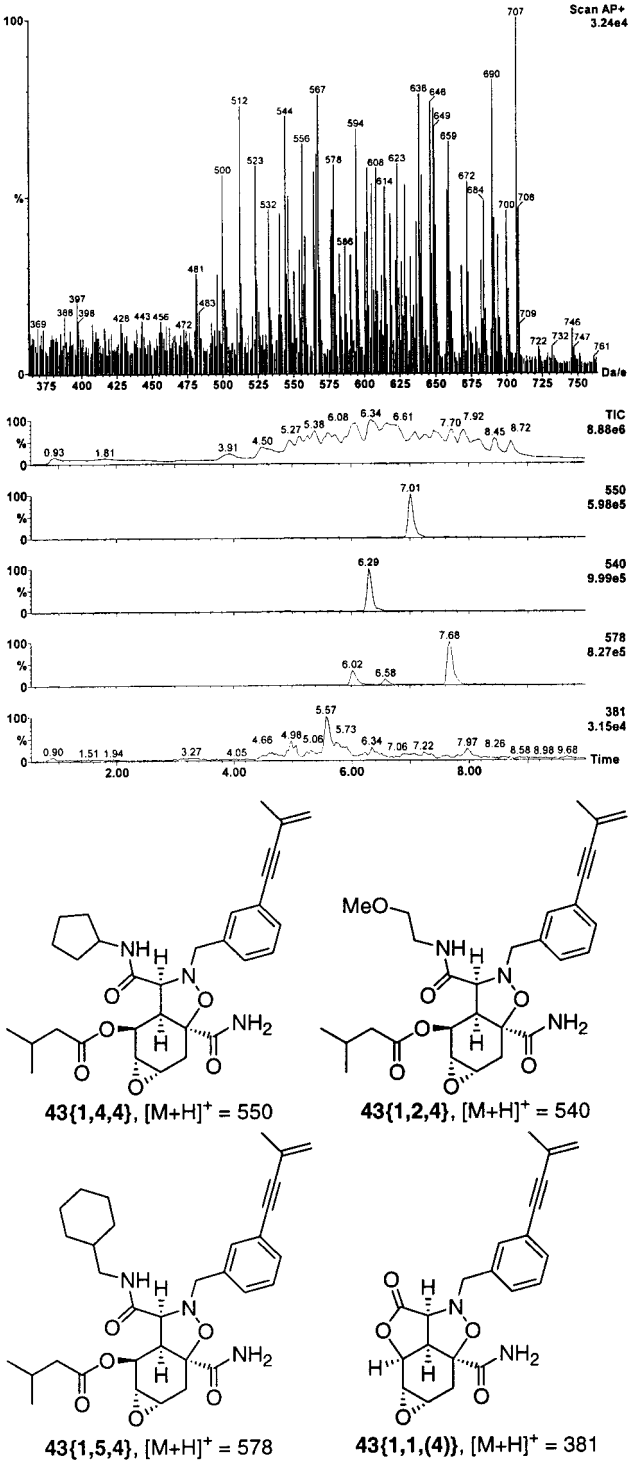



**Figure 11.** Representative LC-MS data for test library pool **43**{**X,X,4**} acylated with **Acid 4**. Top: Mass spectrum averaged over the entire chromatogram. Middle: TIC trace and single mass traces. Bottom: Corresponding structures. The 550 and 540 single mass traces represent typical "clean" product signals. Smaller peaks (6.02 min, 6.58 min) in the 578 single mass trace arise from isotopic compositions of lower mass products (FW = 575, 576). The 381 single mass trace is representative of a "weak" product signal.

responsive reporter gene[103] in a stably transfected mink lung cell line (Figure 14).[104] Since this library is not encoded with chemical tags, a recursive deconvolution strategy was used to investigate this activity further.

(104) Stockwell, B. R.; Schreiber, S. L. *Curr. Biol.* **1998**, *8*, 761−770.

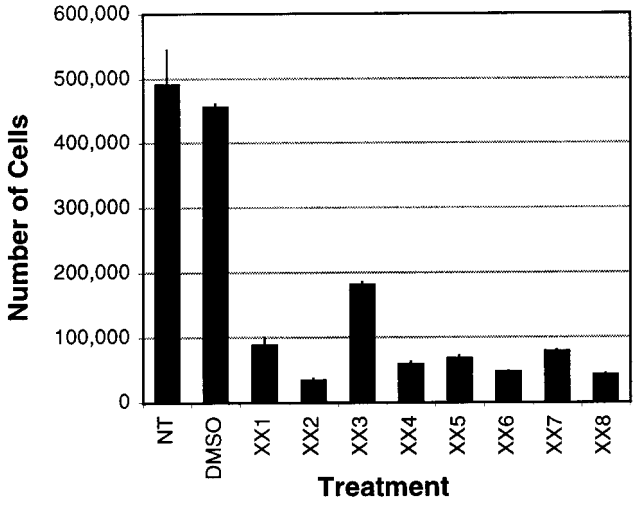
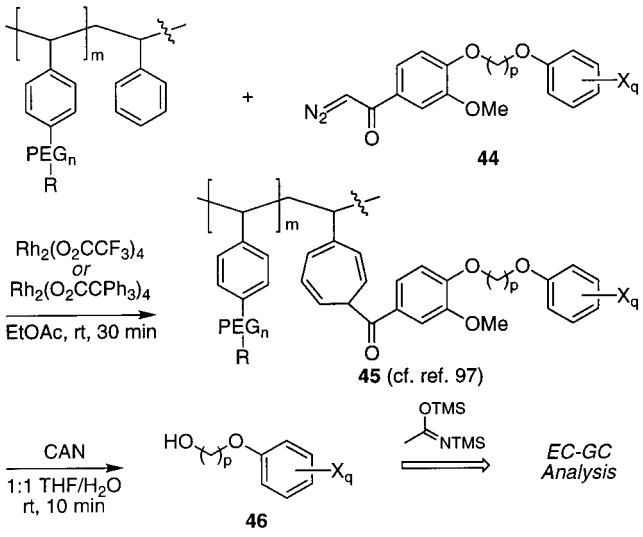



**Figure 12.** Mink lung cell proliferation assay. Lane 1: No treatment. Lane 2: 0.1% DMSO control. Lanes 3−10: Pools **43**{**X,X,1**} through **43**{**X,X,8**} at 1 $\mu$M concentration per compound. Data represent the average of three experiments with error bars indicating one SD.

**Scheme 15**




The 64 compounds in **43**{**X,X,8**} were resynthesized as eight pools, designated **43**{**X,1,8**} through **43**{**X,8,8**}, each containing eight different compounds. Each pool was aminolyzed with a different amine and all pools were acylated with **Acid 8**. As a negative control, pool **43**{**X,X,3**} was also resynthesized as eight pools, **43**{**X,1,3**} through **43**{**X,8,3**}. Surprisingly, pool **43**-{**X,8,3**}, assayed at a concentration of 1 $\mu$M per compound, was a stronger activator of the TGF-$\beta$-responsive reporter gene than any of the other 15 eight-compound pools or either of the parent 64-compound pools (data not shown).

A final round of resynthesis deconvoluted pool **43**{**X,8,3**}, yielding single compounds **43**{**1,8,3**} through **43**{**8,8,3**}. These eight compounds and all of the 16 intermediates preceding them were recovered in high purity as determined by $^1$H NMR and HR-TOF-ESI-MS analysis. All 24 compounds were screened and compounds **43**{**5,8,3**} and **43**{**6,8,3**} were found to activate the TGF-$\beta$-responsive reporter gene (Figure 15). However, the alkynylbenzyltetracycle (**43**{**5,1,1**} and **43**{**6,1,1**}) and $\gamma$-hydroxyamide (**43**{**5,8,1**} and **43**{**6,8,1**}) precursors to these compounds were even more active. The six active compounds were purified by silica gel chromatography and reassayed, verifying that the desired major product is responsible for the

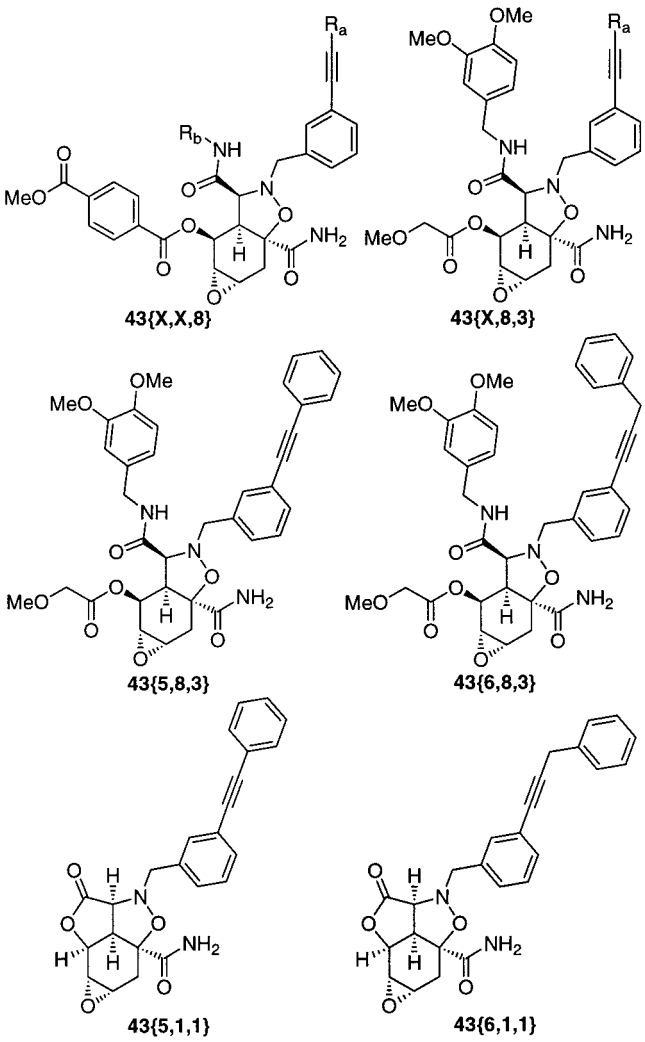



**Figure 13.** Activators of the TGF-$\beta$-responsive reporter gene.

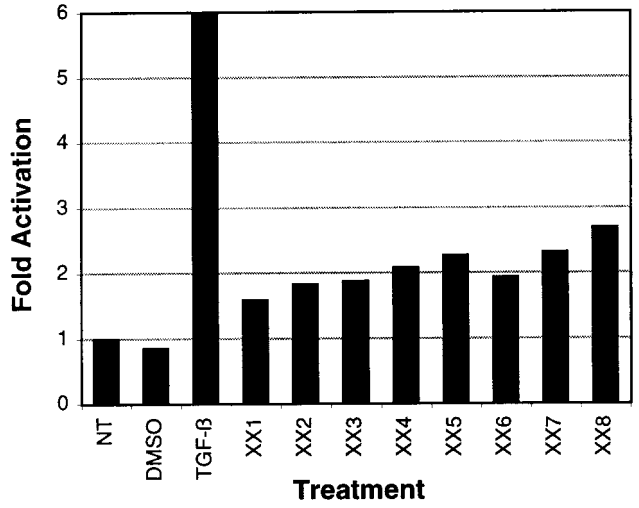



**Figure 14.** TGF-$\beta$-responsive reporter gene assay. Lane 1: No treatment. Lane 2: 0.1% DMSO control. Lane 3: 200 pM TGF-$\beta$1. Lanes 4−11: Pools **43**{**X,X,1**} through **43**{**X,X,8**} assayed at a concentration of 250 nM per compound. Data represent a single experiment with fold activation calculated relative to untreated cells.

activity in each case. The $EC_{50}$ for the strongest activator, **43**-{**6,1,1**}, is approximately 50 $\mu$M.

Discovery of these compounds, while somewhat fortuitous, gave us a measure of confidence in the cell permeating and pathway modulating properties of members of the library. These results also highlight the shortcoming of screening mixtures of compounds. Active compounds may be masked by competing cytotoxic, cytostatic, or antagonistic compounds in the same pool. Alternatively, activity may arise from synergistic effects between two or more compounds, making deconvolution to a single structure impossible. For these reasons, we generally avoid screening mixtures in high-throughput chemical genetic screens.

**Chemical Genetic Screens Using Compounds Released from Single Beads.** Solid-phase synthesis was originally developed to facilitate the purification of peptides from reaction mixtures. Much of the more recent solid-phase organic synthesis has been used for a similar purpose, but involves non-oligomeric small molecules. Split-pool synthesis provides a new and arguably more powerful incentive for performing solid-phase organic synthesis: It yields large numbers of spatially segregated small molecules. To take full advantage of this feature, assay formats must be developed that use compounds derived from single beads. In our early studies of assay formats using single beads, two such systems were used. Binding-based assays were performed by detecting soluble proteins that had been recruited to individual beads via their interactions with a tethered small molecule.[86,91,92] Phenotype-based assays were performed using cells and synthesis beads contained in small volumes of cell culture. These "nanodroplets" were generated either stochastically[26] or on a molded, poly(dimethylsiloxane) (PDMS) grid.[25] Although these assays have yielded useful information, they suffer from several shortcomings, including the fact that the synthesis beads are not easily used more than once. Moreover, the synthesis described above was performed on 90 $\mu$m TentaGel beads, having a loading capacity of only approximately 80−100 picomolar equivalents per bead. In mink lung cell cytoblot assays searching for small molecule suppressors of trapoxin's and rapamycin's actions and using a subset of approximately 77000 beads,[105] these shortcomings manifested themselves in several ways, including the need for resynthesis of compounds corresponding to apparent positives. On the other hand, our early experience in screening these synthesis beads has revealed the ability to recover positive beads, cleave their EC-GC tags, and decode them successfully.

To address the problem noted above, colleagues at the Harvard Institute of Chemistry and Cell Biology have developed instrumentation and robotics that provide efficient arraying of synthesis beads into high-density wells, releasing and transferring of compounds into high-density stock solutions, and transferring of these solutions into high-density PDMS assay plates.[106] These assay plates are the preferred format for cytoblot assays,[16] and the high-density stock solutions are also well-suited for small molecule printing.[18] To allow efficient release of compounds into storage wells without the need for removal of toxic byproducts, we have explored various linkers. To provide sufficient amounts of released compounds to allow their use in large numbers of assays, we have explored beads with higher loading capacities.

## Conclusion

We have completed the design, development, validation, and split-pool synthesis of a binary encoded library of complex small molecules having structural features reminiscent of natural products. Highly efficient coupling reactions were used to attach

(105) B. R. S. S. J. Haggarty, A. J. You, B. K. Wagner, and S. L. S., unpublished results.

(106) L. Walling and R. W. King, unpublished results.

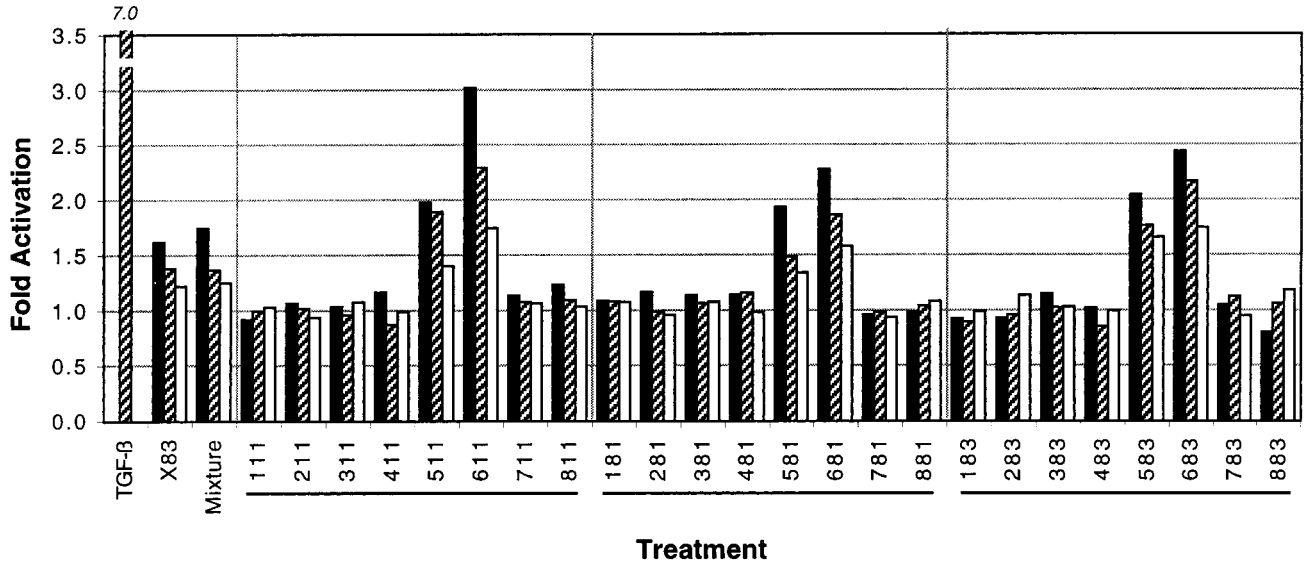



**Figure 15.** TGF-$\beta$-responsive reporter gene assay. Lane 1: 1 nM TGF-$\beta$1. Lanes 2 and 3: Pool **43**{**X,8,3**} and a mixture of individually synthesized compounds **43**{**1,8,3**} through **43**{**8,8,3**}, assayed at 2.5 (black bars), 1.25 (striped bars), and 0.625 $\mu$M (white bars) per compound. Lanes 4−27: Alkynylbenzyltetracycle precursors **43**{**1,1,1**} through **43**{**8,1,1**}, $\gamma$-hydroxyamide precursors **43**{**1,8,1**} through **43**{**8,8,1**}, and acylated final compounds **43**{**1,8,3**} through **43**{**8,8,3**}, assayed at 25, 12.5, and 6.25 $\mu$M per compound. Data represent the average of two experiments with fold activation calculated relative to untreated cells (data not shown). Note that the background signal from the instrument has not been subtracted.

a variety of building blocks to a rigid, densely functionalized tetracyclic template. We have described a four-stage library validation protocol that provides maximum confidence that library members are synthesized in high purity and efficiency. Preliminary biological assays have identified molecules that activate a TGF-$\beta$-responsive reporter gene in mammalian cells.

This work also highlights the need for further development of mild, chemoselective coupling reactions that are compatible with both the wide range of functional groups found in split-pool libraries and the solid support−linker combinations useful in chemical genetic assays. Furthermore, our preliminary biological results indicate the dangers involved in screening mixtures of compounds. The development of assay formats that instead use individual compounds should eliminate these concerns.

Our current research efforts are aimed at adapting the synthetic chemistry described in detail herein to new linkers and beads. The results are highly encouraging, and we can now anticipate adapting the underlying chemistry, as well as other synthetic pathways in development, in numerous ways. Such syntheses should allow many thousands of experiments where synthetic small molecules are used to modulate the circuitry of cellular systems.

## Experimental Section

See Supporting Information.

**Acknowledgment.** D.S.T. was supported by NDSEG and Roche Fellowships; M.A.F. was supported by Glaxo-Wellcome; B.R.S. is a Howard Hughes Medical Institute Predoctoral Fellow; M.D.S. was an NIH postdoctoral fellow. We are grateful to Dr. Andrew Tyler, Ms. Jennifer Lynch, and Mr. John Athanasopoulos for expert mass spectrometry support. We thank Mr. Kuenley Chiu and Ms. Lygia Snow for assistance with EC-GC tag analysis. The Harvard University Mass Spectrometry and NMR Facilities were funded by grants from the NSF (CHE-9020043, CHE88-14019) and NIH (S10-RR06716, 1-S10-RR04870-01). Financial support for this work was provided by the National Institute of General Medical Sciences. The ICCB is supported by the National Cancer Institute, and Merck & Co. S.L.S. is an Investigator at the Howard Hughes Medical Institute.

**Supporting Information Available:** Complete experimental procedures and analytical data for demonstration compounds; building block testing; test library synthesis, screening, and deconvolution; binary encoding; and full-scale library synthesis (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA992144N