Diversity-Based Organic Synthesis in the Era of Genomics and Proteomics**

Prabhat Arya,* Doug T. H. Chou, and Myung-Gi Baek

Dedicated to Dr. David Thomas

Synthetic organic chemistry has always been a frontier area of research due to its impact on the materials and biological sciences. The scientific community involved in this area is constantly being challenged to develop efficient methodologies, novel reactions, and processes that will lead to the synthesis of desired target molecules and their derivatives.[1] Not only is synthetic organic chemistry a tool for obtaining compounds that can be utilized for understanding biological functions or the behavior of materials, but it also leads to the creation of novel drug/drug-like candidates and of novel materials with interesting properties.

During the last few decades, there has been a great push for the development of novel synthetic methodologies that lead to complex natural products, natural product like molecules, and simpler analogues of natural products in an efficient manner. Tremendous progress has been made in asymmetric synthesis for the development of stereo- and enantioselective reactions. Combinatorial chemistry dealing with organic synthesis emerged during the last decade.[2] This was largely due to the extensive time period required for the identification of drug-like candidates and novel materials. In a classical manner, lead compounds are derived from the extraction of natural products from plants, animals, insects, or microorganisms. Following positive responses from several fractions, the next challenge is identification of the chemical entities responsible for the biological activity. Synthesis is then undertaken to obtain specific compound(s) in large quantities or to obtain simpler analogues that may exhibit similar biological response. Overall, traditional medicinal chemistry approaches toward these goals have been very time consuming and very expensive. By applying combinatorial chemistry, it is possible to obtain large sets of organic/bioorganic compounds over short periods of time, either by applying parallel high-throughput synthesis or by a "mix and split" technology. In the coming years, it is expected that combinatorial chemistry will play a crucial role in understanding the biological functions of new targets or in the identification of novel targets in the area of genomics and proteomics research. Over the past few years, we have observed the impact of combinatorial chemistry in the development of high-throughput synthesis of focused libraries. It is anticipated that the next era will involve diversity-based synthesis of natural product like compounds to explore their applications in emerging genomics- and proteomics-related research activities. Although very recent, these efforts are already in action within the combinatorial chemistry community and are highlighted in this article.

Recently, Schreiber has given an elegant description of the differences between target-oriented synthesis and diversity-based synthesis.[3] As Schreiber explains, in a drug discovery process, target-oriented synthesis of small molecules is undertaken on the basis of retrosynthetic analysis. In general, this approach requires preselection of the protein target(s) which are then utilized for developing small molecules as selective binding agents, either by traditional methods or by solid-phase combinatorial chemistry based on organic synthesis. This can further be applied in a parallel manner or by using a mix and split technology for developing high-throughput organic synthesis. Using a mix and split method, it is possible to synthesize a large set of compounds which are then screened as mixtures. The identification coding technique helps to identify the positive lead compounds in mixtures. Parallel synthesis generally does not result in higher numbers as it provides only one compound per well.

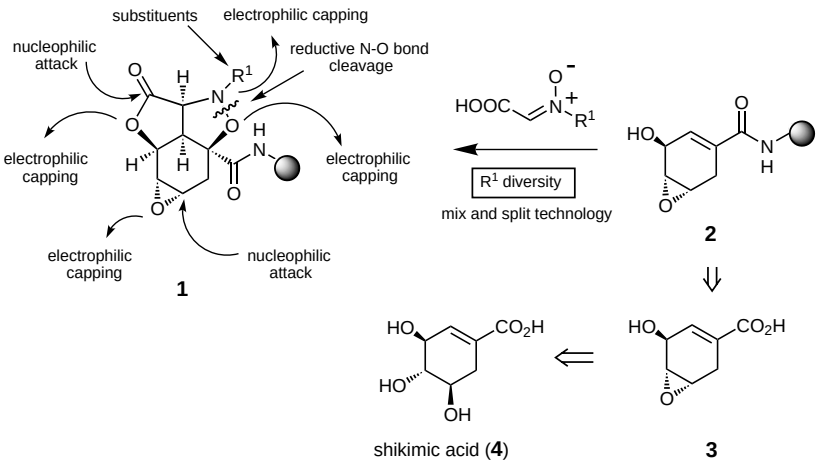
In contrast, and as described by Schreiber, diversity-based organic synthesis is not focused toward a given target. Rather, it is used for identifying new targets or for understanding biological functions of new targets. Moreover, it does not rely upon retrosynthetic analysis and is aimed at developing novel methodologies for obtaining small molecules with structural complexity and diversity. Schreiber compares this to the random mutation approach, which is a well-established technique for modulating specific, cellular, biological responses. Schreiber further defines this technique as a "chemical genetics approach" in which small-molecule libraries are used for probing cellular responses in search of "therapeutic target validation" and "chemical target validation".[4] Given

[*] Dr. P. Arya, Dr. D. T. H. Chou, Dr. M.-G. Baek
Chemical Biology Program
Steacie Institute for Molecular Sciences
National Research Council of Canada
100 Sussex Drive, Ottawa, ON, K1A 0R6 (Canada)
Fax: (+1) 613-952-0068
E-mail: prabhat.arya@nrc.ca

[**] NRC publication no. 43843. David Thomas is a senior colleague and mentor.

  1433-7851/01/4002-0339 $ 17.50+.50/0

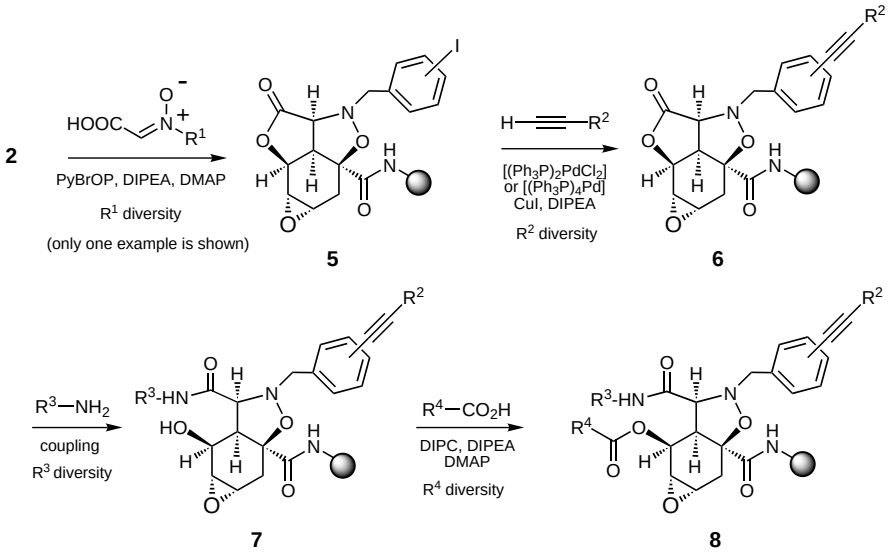
that genomics and proteomics research will provide rapid information on the discovery of novel genes and proteins, we are faced with the challenge of understanding their functions in a timely fashion. Diversity-based synthesis leading to small-molecule libraries will play a crucial role in this area, and it will likely rival the mutational or gene knock-out experiments commonly carried out in biological laboratories. Instead of carrying out labor-intensive gene knock-out experiments, the use of chemical-based small-molecule libraries would provide similar answers with relative ease. In particular, natural product like libraries will be in great demand, complementing the available natural products that are utilized for modulating (that is, activating or inactivating) protein function(s).



Scheme 1. Shikimic acid derived tetracyclic template for the synthesis of small-molecule libraries.

It is well known that several, small-molecule, complex natural products interfere with protein–protein, carbohydrate–protein, and DNA–protein interactions. For example, in the area of protein–protein interactions, it is generally believed that the size and the rigidity of the structural elements within a small molecule tend to provide better binding/interfering agents. This is largely due to flat surfaces that are involved in most protein–protein interactions, although this may not be valid for the active sites for individual enzymes. The two approaches of Schreiber and co-workers for developing diversity-based high-throughput syntheses are summarized as follows.

Using a mix and split solid-phase approach, combined with the encoding technology, Schreiber et al.[5] synthesized a diversity-based small-molecule library on a tetracyclic template **1** (see Scheme 1 for various diversity-based reactions on the template) which is attached to a TentaGel-S NH_2 [poly(ethyleneglycol)–polystyrene copolymer resin]. This template, anchored onto the solid support, was prepared from the epoxide derivative **2** of shikimic acid on reaction with various nitrone carboxylic acid derivatives. The tandem acylation, followed by intramolecular 1,3-dipolar cycloaddition, gave the product with complete stereo- and regioselectivity. 4,5-Epoxy-3-hydroxy-1-cyclohexene-carboxylic acid (**3**) was synthesized from shikimic acid (**4**). Template **1** possesses a high degree of rigidity and is highly functionalized (with the isoxazolidine, lactone, and epoxide). This could further be subjected to a variety of organic and organometallic reactions for the generation of small-molecule libraries. It is interesting to note that the synthetic plan used by Schreiber does not require use of protective groups. Validation of several key reactions was thoroughly carried out, before the sequence of steps was applied with the mix and split encoded technology for a full-scale library synthesis. The validation steps involved the repetition of the reaction sequence for suitability, with testing of potential building blocks at every step, the generation of a small library, and the investigation of unforeseen complications. For example, the lactone and the epoxide moiety from template **1**, reacted readily with nucleophiles such as $BuNH_2$, PhNCO, Et_2NH, and PhCN in a typical organic solvent to obtain the hydroxyl derivatives. Subsequently, the alcohols generated were coupled with carboxylic acids or acid halides to form the ester derivatives. The thioacidolysis of the epoxide was successfully performed using thioacetic acid to yield another hydroxyl adduct. Since the isoxazolidine nitrogen could have various substituents, a number of nitrone acids containing reactive functional groups were synthesized. For example, a number of iodoaryl nitrone acids were selected for synthesis of the iodoaryl tetracyclic templates (**5**; Scheme 2). Iodo substituents could further be subjected to palladium-catalyzed, cross-coupling reactions to obtain compound **6**, or could be subjected to amination, etherification, or carbonylation. Moreover, the iodo func-



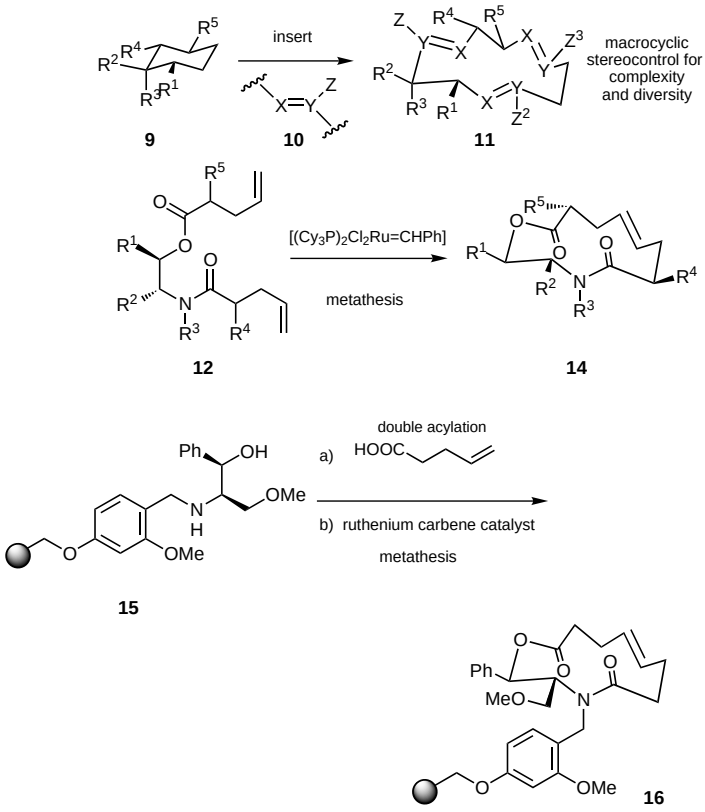
Scheme 2. The synthesis of small-molecule libraries from the tetracyclic template. DIPEA = diisopropylethylamine, DMAP = 4-dimethylaminopyridine, DIPC = diisopropylcarbodiimide.

  1433-7851/01/4002-0340 $ 17.50+.50/0 *Angew. Chem. Int. Ed.* **2001**, *40*, No. 2

tional group could be further substituted to alkenes, alkynes, and chloroformates for the generation of highly diverse templates. After optimization of a series of reactions, lactone aminolysis (to form **7**) and esterification (to form **8**), led to the synthesis of a full-size library (approximately 2.18 million compounds) with a high degree of complexity and diversity. The library was analyzed for its inhibitory activity toward a *Xenopus lavis* oocyte extract assay, mink lung cell proliferation, and rapamycin-based growth inhibition in *Saccharomyces cerevisiae*.[6] Interestingly, several library members were identified as activators of the reporter gene in mink lung cells. It is remarkable that the library design strategy relied upon the use of highly mild reagents and efficient stereoselective synthetic methods. This indeed is a very good lesson for the combinatorial chemistry community. Examples of library synthesis that utilize a simple, chiral starting material for obtaining structurally complex and diverse building block templates are likely to appear in the literature in coming years.

Another interesting example of a similar approach appeared in 1999. In this communication,[7] Schreiber and coworkers reported an efficient approach to the synthesis of macrocyclic libraries. A key step in their strategy involved the application of a well-known ring-closing metathesis reaction,[8] which is then followed by exploring macrocyclic-based stereocontrol reactions. It is well-known that several biologically active natural products bear medium to large rings. A number of antibiotics (for example, erythromycin) belong to this category of macrocyclic-base, highly functionalized compounds. Complex small-molecule libraries having medium to large size ring derivatives can provide biologically enriched compounds. Successful strategies will largely depend upon the structure of the acyclic precursors having optimized conformational features suited for the ring closure. In general, six-membered ring cyclization reactions are highly efficient and high yielding. Macrocyclic ring closure could also exhibit similar properties by introducing the structural and the functional elements that would favor the ring cyclization. Schreiber et al. envisioned that inserting *trans*-olefinic moieties **10** into the alternative sp^3 bonds of a six-membered ring **9** would favor the macrocyclic ring cyclization that yields **11** due to the release of the *trans* annular effects and the torsional strains (Scheme 3). For the successful cyclization metathesis, it is important that the reactive termini in both acyclic precursors are oriented in a close proximity suitable for the ring closure. Olefinic and carbonyl groups in a macrocycle could then undergo various macrocyclic-based stereocontrolled reactions (such as epoxidation or enol ether chemistry).

Before library synthesis, the ring-closing metathesis reaction was tested in the solution phase with the model compound. Compound **12** was obtained from the sequential acylation of a 1,2-amino alcohol derivative. Ring-closing metathesis was carried out in the presence of Grubbs' catalyst (**13**; 5–15 mol %) to give the desired product **14** in average to excellent yields. Further, it was shown that epoxidation and enol ether alkylation reaction with the macrocycle gave the epoxide product and the alkylated derivative with complete stereocontrol. At this stage, the ring closing metathesis based macrocyclization was tried on the solid phase. A chiral amino

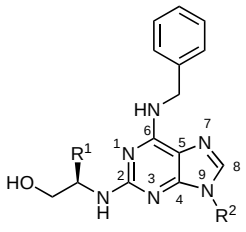


Scheme 3. Schreiber et al.'s macrocyclic ring formation approach for the potential synthesis of small-molecule libraries. Top: theoretical reaction; middle: model solution-phase system; and bottom: solid-phase reaction.

propanol derivative was anchored onto the solid support using a traceless linker yielding compound **15**. This was first subjected to double acylation followed by ring-closing metathesis that gave the chiral macrocyclic derivative **16**. This sets the stage to carry out a mix and split encoded process toward the synthesis of stereochemically complex, chiral macrocyclic compound libraries. The biological outcome from this class of libraries will be seen in coming years.

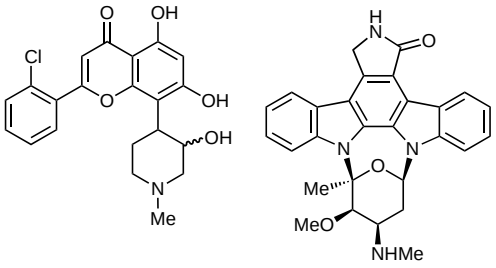
With a similar goal of using small-molecule libraries in genomics and proteomics research, the Schultz group is developing purine-based small-molecule libraries. In combination with the structural features of several kinases and genomics research, these libraries are tested for the identification of kinase inhibitors. The purine ring constitutes a common structural motif in molecules that play vital roles in many biological processes. For example, it is an important component in proteins such as cellular kinases, specifically cyclin-dependent kinases (CDKs).[9, 10] These enzymes regulate many biological processes such as cell growth, DNA replication, and cell division.[9] As such, the synthesis of kinase inhibitors has grown rapidly over the last few years. Some of the earlier CDK inhibitors are shown in Scheme 4 (**17**–**19**).[11] These compounds displayed reasonable inhibition, but still lacked the potency and selectivity to be ideal drug candidates. In order to facilitate the synthesis of purine derivatives with greater affinity for CDK enzymes, it was postulated, based on several X-ray crystallographic studies, that the synthesis of

  1433-7851/01/4002-0341 $ 17.50+.50/0



17

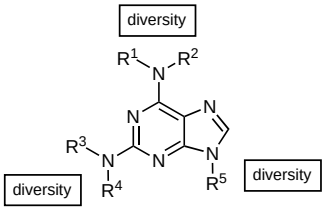
olomoucine: R^1= H, R^2 = Me
(*R*)- roscovitine: R^1= Et, R^2 = *i*Pr



flavopiridol (**18**) staurosporine (**19**)

Scheme 4. Natural product based compounds which are CDK inhibitors.

structurally diverse purine analogues (with diversity at the C2, C6, and N9 positions) might lead to the identification of more potent CDK inhibitors (Scheme 5; **20**).[12]

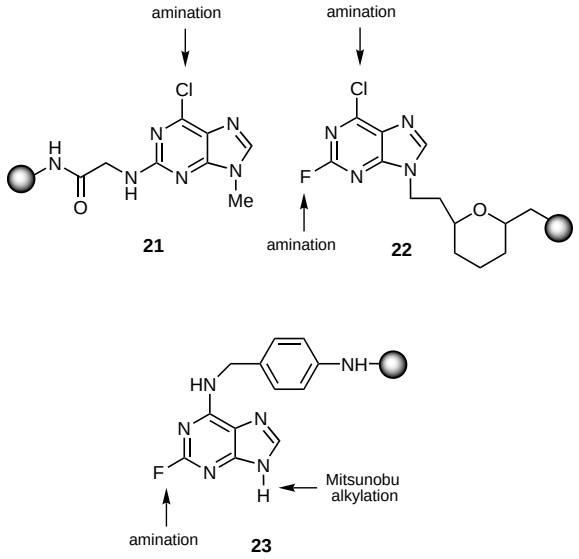


20

Scheme 5. General structure of the purine-based compounds to be synthesized in libraries for testing as CDK inhibitors.

However, traditional synthetic methodologies involving solution-phase chemistry often prove to be difficult and time-consuming processes if a large number of diverse molecules are desired. On the other hand, combinatorial chemistry offers a practical solution to produce compound libraries for their biological significance. Libraries of natural products or collections of small organic molecules are being explored for their ability to affect particular pathways in cells or organisms. These libraries will enable the identification of new targets for their effectiveness against particular pathways, in the hope of discovering new drugs. The rapid screening of these libraries is now possible with modern technologies. Several strategies have been employed by Schultz and co-workers for solution/solid-phase synthesis of purine derivatives.[11] These initial targets will allow a primary determination of structure–activity relationships (SARs) against CDK enzymes.

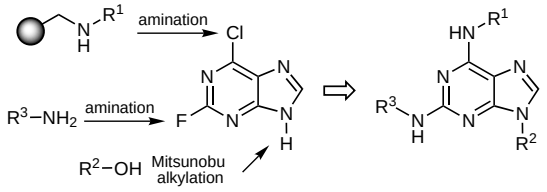
The solid-phase synthesis developed by Schultz and co-workers involves the attachment of a solid support to one of the purine positions (C2, N9, or C6) with glycinamido, ethyl tetrahydropyran ether, or 4-aminobenzyl amino groups (Scheme 6; **21**–**23**, respectively).[11b,c,e] The main drawback of this strategy is the loss of a potentially useful site for



21 **22** **23**

Scheme 6. Solid-phase approaches by Schultz et al. for the synthesis of purine-based libraries.

structural variation. A combination of solid- and solution-phase chemistry was utilized to overcome this problem. This strategy involves immobilization of the purine ring at the C6 position allowing a variety of substitution to occur at the C2, C6, and N9 positions (Scheme 7; **24**, **25**). Initially, the solid-

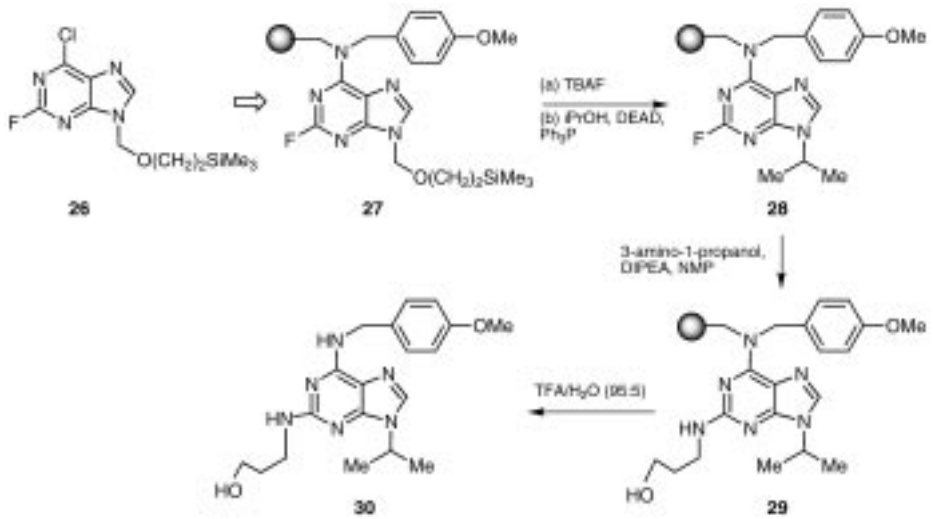


24 **25**

Scheme 7. A combination of solid- and solution-phase chemistry utilized by Schultz et al.

phase synthesis involves the reductive amination of 2,3-dithiol-1-propanol derivatized 4-methylbenzhydrylamine (MBHA) with various primary amines to give resin-bound amines.[11c] The amines are then reacted with compound **26**, to achieve combinatorial diversity at the C6 position. Mitsunobu and amination chemistry are then applied for further functionalization at the N9 and C2 positions. Scheme 8 illustrates an example of this type of approach. Like the previous strategy, this approach also suffers from two key set backs: 1) direct introduction of a secondary amine at the C6 position is not possible, and 2) amination with bulky groups at the C2 position occurs slowly and in low yields. Schultz and co-workers tackled these difficulties with two new solution-phase strategies, allowing diversification of the trisubstituted purine libraries shown in Scheme 9.[13]

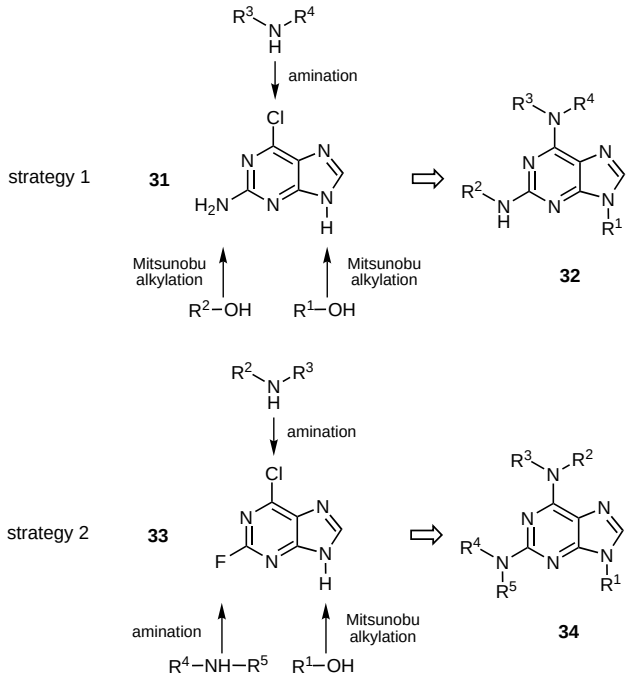
  1433-7851/01/4002-0342 $ 17.50+.50/0 *Angew. Chem. Int. Ed.* **2001**, *40*, No. 2



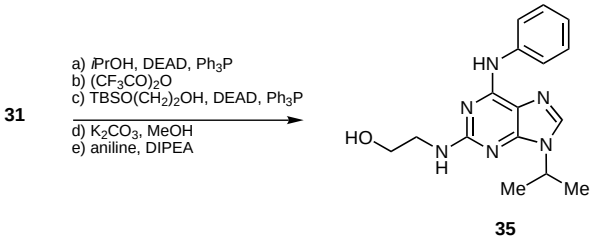
Scheme 8. Multistep solid-phase synthesis of purine-based small-molecule libraries, developed by Schultz et al. TBAF = tetrabutylammonium flouride, DEAD = diethylazodicarboxylate, NMP = *N*-methylpyrrolidine, TFA = trifluoroacetic acid.

duction of various amine derivatives at the C6 position was accomplished in the presence of *n*-butanol at elevated temperatures (see, for example, compound **35**). Strategy 2 is different from strategy 1 in the order of substitution at various positions, but the main chemistry for both strategies is virtually identical. An example of solution chemistry using this strategy is shown in Scheme 11.

Using a combination of solution/solid-phase approaches, several purine libraries containing hundreds to thousands of purine analogues with structural diversity at the C2, C6, and N9 positions were synthesized. Iterative synthesis can later

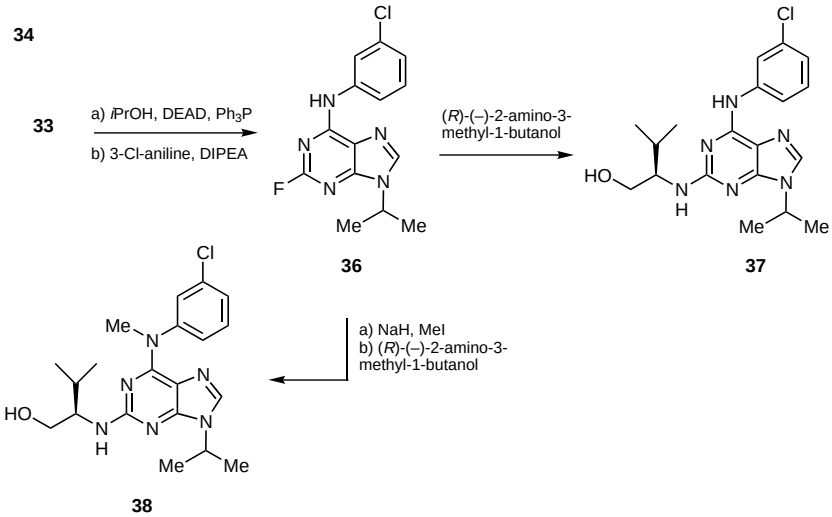


Scheme 9. Solution-phase strategies for the synthesis of purine-based libraries.



Scheme 10. An example of solution-phase synthesis for purine-based libraries (strategy 1 from Scheme 9). TBS = *tert*-butyldimethylsilyl.

The strategies involved the use of 2-amino-6-chloropurine (**31**) and 2-fluoro-6-chloropurine (**33**). Compound **33** can be easily synthesized from commercially available **31**.[11b] In strategy 1 (Scheme 10), **31** was functionalized in a stepwise fashion at the N9, C2, and C6 positions. Mitsunobu chemistry using primary and secondary alcohols allows the introduction of various substitutents at the N9 position. The amino group at the C2 position was acylated with trifluoroacetic anhydride in order for Mitsunobu alkylation to occur efficiently. Lastly, introbe applied to enhance the affinity of the lead compounds for CDK activity. To date, biological assays have been able to identify a number of potent cyclin-dependent kinase inhibitors from the libraries. In addition, other purine derivatives have been shown to inhibit cellular proliferation at specific phases of the cell cycle, as well as the growth of cells linked to



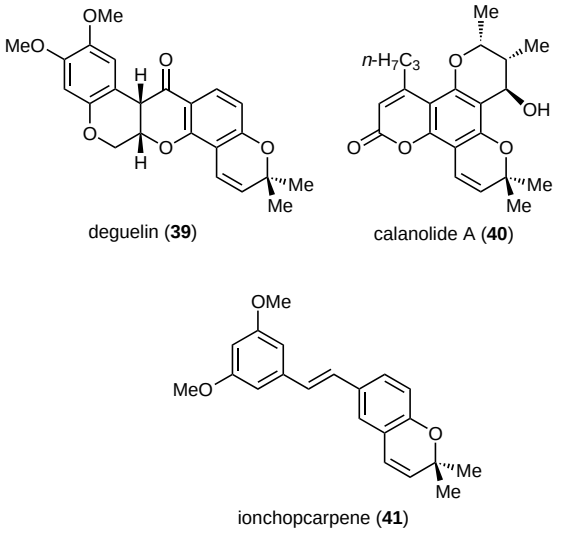
Scheme 11. An example of solution-phase synthesis for purine-based libraries (strategy 2 from Scheme 9).

  1433-7851/01/4002-0343 $ 17.50+.50/0

specific tumors.[13] Schultz and co-workers have recently isolated, from the purine-based libraries, a compound named myoseverin which has the capability to induce the reversible fission of myotubes (muscle cells) into mononucleated fragments.[14] This finding has implications in the process of tissue regeneration and wound healing.

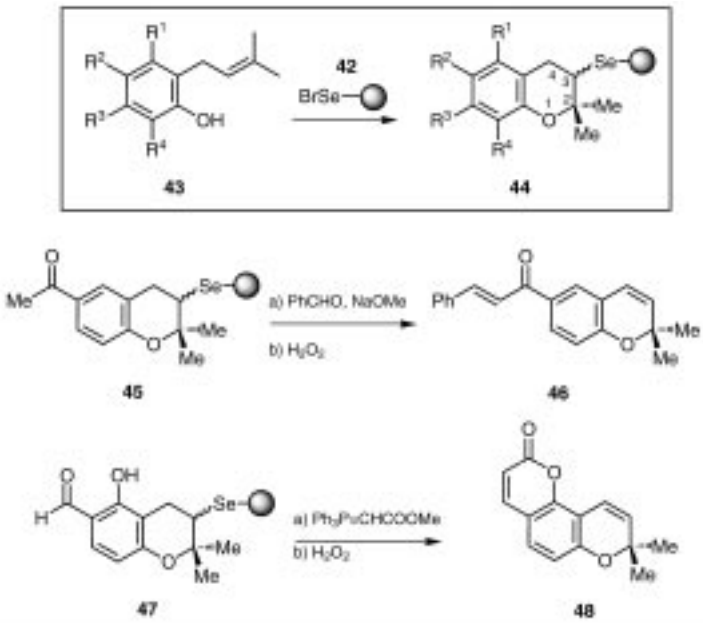
More recently, Bertozzi and co-workers[15] utilized kinase-based libraries to select for carbohydrate sulfotransferases inhibitors given that both enzymes use a similar substrate. Surprisingly, several purine derivatives were identified as lead compounds with an IC_{50} value of 20–40 μM. These initial compounds will serve as templates for the design of future libraries with greater affinity for carbohydrate sulfotransferase enzymes.

Nicolaou and co-workers are actively involved in the area of solid-phase synthesis leading to natural product like, small-molecule libraries. Recently, they reported a solid-phase selenium-based linker and its application for the synthesis of benzopyran derivatives.[16] Several natural products, such as flavonoids, coumarins, rotenoids, stilbenoids, and chromene glycosides contain the dimethylbenzopyran moiety as a part of their core structure. Some representative examples are shown in Scheme 12. A high-throughput synthesis of their



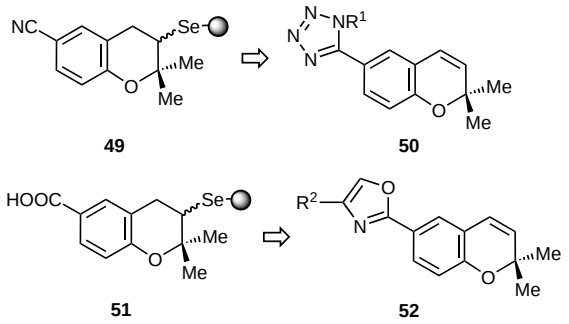
Scheme 12. Examples of dimethylbenzopyran-based natural products.

derivatives could provide interesting lead compounds with a broad spectrum of biological properties. A key step in the approach of Nicolaou et al. is the synthesis of the selenenyl bromide resin, **42** (Scheme 13) prepared from commerically available polystyrene resin. Lithiation, followed by treatment with dimethyl selenide gave the methyl selenide resin. This was then converted into bromide derivative **42** on treatment with bromine. Resin **42** reacted with compound **43** to give the selenyl benzopyran derivative **44**. Under oxidative work-up (33% H_2O_2), it was possible to cleave the benzopyran moiety from the resin with the introduction of the double bond between the C3 and C4 positions. The methodology is highly versatile and could be applied to a variety of substrates bearing different functional groups on the aromatic ring of the



Scheme 13. Nicolaou et al.'s selenide resin based approach for the synthesis of dimethylbenzopyrans.

benzopyran moiety. For example, compound **45** could undergo an enol ether based aldol reaction followed by oxidative work-up to give compound **46** in high yield. Similarly, the selenide resin bound benzopyran derivative **47** could react under Wittig conditions to yield the coumarin derivative **48** after the oxidative work-up. Nicolaou and co-workers have further shown that functional groups from the benzopyran selenide moiety anchored onto a solid support could be converted into useful derivatives. For example, cyano derivative **49** (Scheme 14) was converted into the corresponding



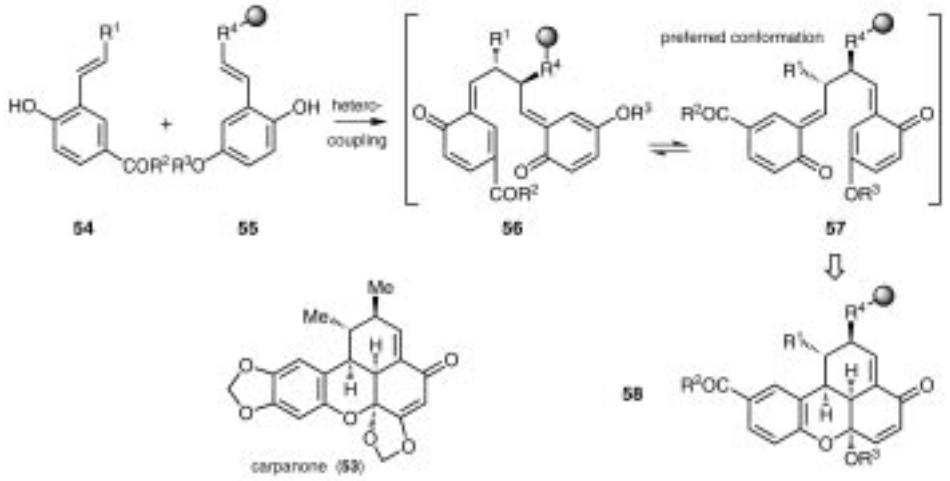
Scheme 14. Examples of functional-group transformation in selenide resin based dimethylbenzopyrans.

tetrazole derivative in a few steps that involve reaction with azidotrimethyl tin followed by N–$SnMe_3$ cleavage under trifluoroacetic acid (TFA) conditions. In a similar manner, oxazole derivative **52** was obtained from compound **50** in a number of steps. In general, these transformations are high yielding and provide a useful entry to the synthesis of biologically relevant, natural product like compounds. Following success with the use of selenium-based resin for the synthesis of dimethylbenzopyrans, a solid-phase synthesis of indoline-like compounds was also accomplished using sele-

  1433-7851/01/4002-0344 $ 17.50+.50/0 *Angew. Chem. Int. Ed.* **2001**, *40*, No. 2

nium chemistry.[16c] The approaches developed by Nicolaou and co-workers are simple, highly efficient, and could be utilized for obtaining a variety of dimethylbenzopyran- and indole-based small-molecule libraries.

Shair and co-workers are developing solid-phase approaches for the synthesis of natural product like compounds that mimic biosynthetic pathways.[17] Biomimetic synthesis often requires the use of simple organic precursors and mild reaction conditions. Shair et al. planned to utilize biomimetic-type reactions on the solid phase that could further be combined with the mix and split technology for library synthesis. Carpanone (**53**, Scheme 15), a natural product from the benzoxanthenone family, was selected for study. The biomimetic synthesis of carpanone was achieved by Chapman in 1971 and involves a stereoselective, oxidative coupling followed by an intramolecular Diels – Alder reaction.[18]



Scheme 15. Shair et al.'s biomimetic approach for the synthesis of benzoxanthenone-based natural product like compounds.

Following the biomimetic synthesis of carpanone, Shair and co-workers utilized a hetero-coupling reaction followed by an intramolecular Diels – Alder reaction on the solid phase.[17] The plan was to use two different hydroxystyrenes **54** and **55**, which differ in their electronic configuration. Electron-rich sytrene derivative **55** was anchored onto a solid support to avoid any homodimerization. Several oxidizing agents were then tried, to study the cross-coupling reaction, and the use of $PhI(OAc)_2$ afforded the desired heterocoupled product **58** on the solid phase. It was interesting to note that only a single isomer was detected after cleavage of the product from the support, and the cycloaddition reaction proceeded with a remarkable stereoselectivity. Several examples were tried using this method, and tetracyclic benzoxanthenone derivatives were isolated in respectable yields after cleavage from the support. This is the first example in which a biomimetic synthesis has been successfully transferred onto the solid phase. The approach presented by Shair and co-workers is highly unique and it could lead to the library synthesis of structurally complex and diverse natural product like compounds, which would allow exploration of their applications in genomics and proteomics research.

To summarize, in recent years, several interesting approaches leading to the synthesis of structurally complex and diverse compounds have appeared in the literature. Collectively, these methods allow the synthesis of a variety of natural product like compounds in a high-throughput manner. If the last century was the era of focused stereoselective organic synthesis, in the coming years diversity-based organic synthesis will play an important role in understanding the function of biological targets which are emerging from the growing efforts in genomics and proteomics research.

[1] For an excellent, comprehensive recent review on "the art and science of total synthesis at the dawn of the twenty-first century", see: K. C. Nicolaou, D. Vourloumis, N. Winssinger, P. S. Baran, *Angew. Chem.* **2000**, *112*, 46; *Angew. Chem. Int. Ed.* **2000**, *39*, 44.

[2] Several text books covering different aspects of combinatorial chemistry have been published in last few years, see: a) "Molecular Diversity and Combinatorial Chemistry Libraries and Drug Discovery": *ACS Symp. Ser.* **1996**; b) *Combinatorial Chemistry—Synthesis and Applications* (Eds.: S. R. Wilson, A. W. Czarnik), Wiley, Toronto, **1997**; c) *Combinatorial Peptides and Nonpeptide Libraries—A Handbook* (Ed.: G. Jung), VCH, New York, **1996**.

[3] S. L. Schreiber, *Science* **2000**, *287*, 1964.

[4] S. L. Schreiber, *Bioorg. Med. Chem.* **1998**, *6*, 1127.

[5] a) D. S. Tan, M. A. Foley, M. D. Shair, S. L. Schreiber, *J. Am. Chem. Soc.* **1998**, *120*, 8565; b) D. S. Tan, M. A. Foley, B. R. Stockwell, M. D. Shair, S. L. Schreiber, *J. Am. Chem. Soc.* **1999**, *121*, 9073, and references therein.

[6] a) Unpublished results from Schreiber and co-workers; b) J. Carcamo, A. Zentella, J. Massague, *Mol. Cell. Biol.* **1995**, *15*, 1573; c) B. R. Stockwell, S. L. Schreiber, *Curr. Biol.* **1998**, *8*, 761.

[7] D. Lee, J. Sello, S. L. Schreiber, *J. Am. Chem. Soc.* **1999**, *121*, 10648.

[8] R. H. Grubbs, S. Chang, *Tetrahedron* **1998**, *54*, 4413.

[9] P. Cohen, *Curr. Opin. Chem. Biol.* **1999**, *3*, 459.

[10] T. M. Sielecki, J. F. Boylan, P. A. Benfield, G. L. Tranior, *J. Med. Chem.* **2000**, *43*, 1.

[11] a) M. T. Fiorini, C. Abell, *Tetrahedron Lett.* **1998**, *39*, 1827; b) N. S. Gray, S. Kwon, P. G. Schultz, *Tetrahedron Lett.* **1997**, *38*, 1161; c) N. S. Gray, L. Wodicka, A.-M. W. H. Thunnissen, T. C. Norman, S. Kwon, H. Espinoza, D. O. Morgan, G. Barnes, S. LeClerc, L. Meijer, S.-H. Kim, D. J. Lockhart, P. G. Schultz, *Science* **1998**, *281*, 533; d) M. Legraverend, O. Ludwig, E. Bisagni, S. LeClerc, L. Meijer, *Bioorg. Med. Chem. Lett.* **1998**, *8*, 793; e) T. C. Norman, N. S. Gray, J. T. Koh, P. G. Schultz, *J. Am. Chem. Soc.* **1996**, *118*, 7430; f) S. R. Schow, R. L. Mackman, C. L. Blum, E. Brooks, A. G. Horsma, A. Joly, S. S. Kerwar, G. Lee, D. Shiffman, M. G. Nelson, X. Wang, M. M. Wick, X. Zhang, R. T. Lum, *Bioorg. Med. Chem. Lett.* **1997**, *7*, 2697; g) D. A. Nugiel, L. A. M. Cornelius, J. W. Corbett, *J. Org. Chem.* **1997**, *62*, 201.

[12] a) U. Schulze-Gahmen, J. Brandsen, H. D. Jones, D. O. Morgan, L. Meijer, J. Vesely, S.-H. Kim, *Proteins Struct. Funct. Genet.* **1995**, *22*, 378; b) W. F. De Azevedo, Jr., H.-J. Mueller-Dieckmann, U. Schulze-Gahmen, P. J. Worland, E. Sausville, S.-H. Kim, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 2735.

[13] Y.-T. Chang, N. S. Gray, G. R. Rosania, D. P. Sutherlin, S. Kwon, T. C. Norman, R. Sarohia, M. Leost, L. Meijer, P. G. Schultz, *Chem. Biol.* **1999**, *6*, 361.

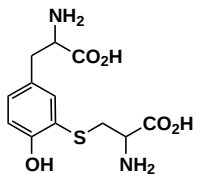
[14] G. R. Rosania, Y.-T. Chang, O. Perez, D. Sutherlin, H. Dong, D. J. Lockhart, P. G. Schultz, *Nat. Biotechnol.* **2000**, *18*, 304.

[15] J. I. Armstrong, A. R. Portley, Y.-T. Chang, D. M. Nierengarten, B. N. Cook, K. G. Bowman, A. Bishop, N. S. Gray, K. M. Shokat, P. G. Schultz, C. R. Bertozzi, *Angew. Chem.* **2000**, *112*, 1359; *Angew. Chem. Int. Ed.* **2000**, *39*, 1303.
[16] a) K. C. Nicolaou, J. A. Pfefferkorn, G.-Q. Cao, *Angew. Chem.* **2000**, *112*, 750; *Angew. Chem. Int. Ed.* **2000**, *39*, 734; b) K. C. Nicolaou, G.-Q. Cao, J. A. Pfefferkorn, *Angew. Chem.* **2000**, *112*, 755; *Angew. Chem. Int. Ed.* **2000**, *39*, 739; c) K. C. Nicolaou, A. J. Roecker, J. A. Pfefferkorn, G.-Q. Cao, *J. Am. Chem. Soc.* **2000**, *122*, 2966.
[17] C. W. Lindsley, L. K. Chan, B. C. Goess, R. Joseph, M. D. Shair, *J. Am. Chem. Soc.* **2000**, *122*, 422.
[18] O. L. Chapman, M. R. Engel, J. P. Springer, J. C. Clardy, *J. Am. Chem. Soc.* **1971**, *93*, 6696.

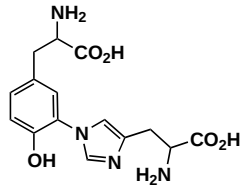
Chemically Modified Amino Acids in Copper Proteins That Bind or Activate Dioxygen**

Malcolm A. Halcrow*

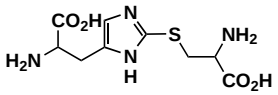
It is an interesting observation that four of the six classes of copper oxidase or monooxygenase enzymes whose structures are known[1–3] contain a chemically modified amino acid either ligated to, or in close proximity to, the active-site copper center. Some of these unusual residues, namely the TyrCys, HisCys, and TyrHis moieties, are derived from a cross-linking reaction between two amino acid side chains. Alternatively, the 2,4,5-trihydroxyphenylalaninequinone (topaquinone, TPQ) and lysyl tyrosyl quinone (LTQ) cofactors are generated by monooxygenation of a tyrosine side chain, followed by nucleophilic attack at the resultant *ortho*-quinone. The HisCys,[4] TyrCys,[5] and TPQ[6] residues were first detected in their respective proteins between nine and eighteen years ago. However, our understanding of the functional roles of these and the other, more recently discovered,[7, 8] residues is often still rather limited. Attention has also now begun to focus on the biosynthesis of these cofactors, which in many cases must involve unusual organic transformations that have not been previously characterized in copper biochemistry.



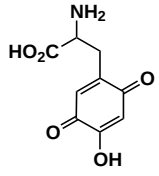
TyrCys



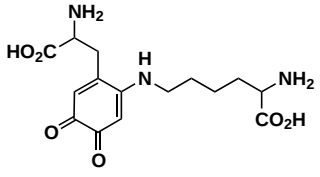
TyrHis



HisCys



TPQ



LTQ

The cross-linked amino acid whose function is best understood is the TyrCys residue found in radical copper oxidases, of which galactose oxidase (GO) is the prototype.[1] This side chain acts as a ligand to a monocopper center in the enzymatic active site (Scheme 1 A),[5] and is oxidized to a [TyrCys]$^{\bullet}$ radical in the active resting enzyme; this phenoxyl radical abstracts a hydrogen atom from the alcohol substrate during turnover.[1] The oxidation potential of the TyrCys moiety is 0.5 V less positive than would be expected for an unmodified tyrosine residue.[9] Consistent with this, several model studies have shown that *ortho*-sulfide substitution of a free or coordinated phenoxide will lower its oxidation potential by 250 – 500 mV.[10] However, recent calculations have shown that the S atom of an *ortho* sulfide substituted phenoxyl group only accepts up to 15 % of the radical unpaired spin, and does not significantly perturb the spin distribution over the rest of the molecule.[11] In addition, it has also been calculated that the O–H homolytic bond dissociation energy (BDE) of an *ortho* sulfide substituted phenol is identical, to within 2 kcal mol^{-1}, to that of an otherwise identical phenol lacking this substituent.[12] Hence, the electronic structure of the [TyrCys]$^{\bullet}$ radical of GO and its reactivity towards H-atom donors are not significantly perturbed by the sulfide substituent, and it appears that the function of this cross-link is purely to lower the oxidation potential of the oxidizable tyrosine ring.

Less is known about the other cross-linked amino acids. The TyrHis cross-link in cytochrome *c* oxidase (CcO) forms part of the heme/copper site of this enzyme (Scheme 1 B),[7] which catalyzes the reduction of dioxygen to water.[3] The currently favored mechanism for this reaction implies that 4-electron reduction of O_2 by CcO occurs in one mechanistic step,[13] and it is thought that the TyrHis residue acts as both as a Brønsted

[*] Dr. M. A. Halcrow
School of Chemistry
University of Leeds
Woodhouse Lane, Leeds LS2 9JT (UK)
Fax: (+44) 113-233-6565
E-mail: M.A.Halcrow@chem.leeds.ac.uk

[**] The author acknowledges the Royal Society (London) for a University Research Fellowship.

  1433-7851/01/4002-0346 $ 17.50+.50/0