Anal. Chem. **2002,** *74,* 3963−3971

Articles

High-Throughput Method for the Production and Analysis of Large Natural Product Libraries for Drug Discovery

Gary R. Eldridge, Hélène C. Vervoort, Chris M. Lee, Peadar A. Cremin, Caroline T. Williams, Shane M. Hart, Matt G. Goering, Mark O'Neil-Johnson, and Lu Zeng*

Sequoia Sciences, Inc., San Diego, California 92121

High-throughput methods were applied to the production, analysis, and characterization of libraries of natural products in order to accelerate the drug discovery process for high-throughput screening in the pharmaceutical and biotechnology industries. Library production integrates automated flash chromatography, solid-phase extraction, filtration, and high-throughput parallel four-channel preparative high-performance liquid chromatography to obtain the libraries in 96- or 384-well plates. Libraries consist of purified fractions with approximately one to five compounds per well. Libraries are analyzed prior to biological screening by a high-throughput parallel eight-channel liquid chromatography−evaporative light scattering detection-mass spectrometry system to determine the molecular weight, number, and quantity of compounds in a fraction. After biological screening, active fractions are rapidly purified at the microgram level and individual compounds are rescreened for confirmation of activity. Structures of active compounds are elucidated by NMR spectroscopy and mass spectrometry. Utilization of a novel microcoil probe allows NMR data to be gathered on 50 μg. As a demonstration, a library was made from the stem bark of *Taxus brevifolia*. Biological screening in the National Cancer Institute's in vitro panel of three cancer cell lines demonstrates that the process enables the discovery of active anticancer compounds not detected in the flash fractions from which the library originates.

Natural products have been a major resource for chemical diversity in the pharmaceutical industry. Currently, ∼60% of the antitumor and anti-infective agents commercially available or in the later stages of clinical trails are of natural product origin.[1] Since 1990, pharmaceutical and biotechnology companies have gradually moved toward high-throughput screening (HTS) of small-molecule compounds for their discovery programs. On average, 300 000 compounds were screened in HTS programs in the year 2000 and this number is expected to grow.[2] The majority of compounds for HTS are provided by large combinatorial chemistry libraries that are synthesized either by solid-phase or by solution-phase chemistry varying from peptide libraries to small-molecule libraries.[3−7] However, 40% of the structurally diverse compounds found in the *Dictionary of Natural Products* are not represented in these combinatorial libraries.[8] To date, only a few libraries based on natural product scaffolds have been screened and these libraries, being synthetic, are usually much smaller than combinatorial libraries.[9−12]

Historically, natural product drug discovery has been a time- and resource-intensive process. Primary screening of crude extracts of plants or microbial fermentations followed by bioassay-guided fractionation, dereplication of active components, and isolation and structure elucidation of novel bioactive compounds can take months to a year. However, with the advent of technologies in HTS and combinatorial chemistry, drug discovery cycles have become much shorter, resulting in many industrial programs discontinuing the screening of biomass extracts. Clearly, an integration of methods and technologies is needed to speed up the isolation, purification, and characterization of bioactive natural products.

Over the past few years, several techniques have been developed for combinatorial chemistry that can be used to reintroduce natural products as an important source for novel leads in the drug discovery process. The production of high-quality

* Corresponding author: (e-mail) lzeng@sequoiasciences.com; (tel) 858-623-0800; (fax) 858-623-0805.

(1) Shu, Y.-Z. *J. Nat. Prod.* **1998**, *61*, 1053−1071.
(2) Fox, S.; Wang, H.; Sopchak, L.; Khoury, R. *J. Biomol. Screening* **2001**, *6*, 137−140.
(3) Dolle, R. E. *Mol. Diversity* **1998**, *4*, 199−223.
(4) Dolle, R. E. *Mol. Diversity* **1998**, *4*, 233−256.
(5) Dolle, R. E.; Nelson, K. H., Jr. *J. Comb. Chem.* **1999**, *1*, 235−282.
(6) Dolle, R. E. *J. Comb. Chem.* **2000**, *2*, 383−433.
(7) Dolle, R. E. *J. Comb. Chem.* **2001**, *2*, 477−517.
(8) Henkel, T.; Brunne, R. M.; Muller, H.; Reichel, F. *Angew. Chem., Int. Ed.* **1999**, *38*, 643−647.
(9) Xiao, X.-Y.; Parandoosh, Z.; Nova, M. P. *J. Org. Chem.* **1997**, *62*, 6029−6033.
(10) Nicolaou, K. C.; Winssinger, N.; Vourloumis, D.; Ohshima, T.; Kim, S.; Pfefferkorn, J.; Xu, J.-Y.; Li, T. *J. Am. Chem. Soc.* **1998**, *120*, 10814−10826.
(11) Nicolaou, K. C.; Pfefferkorn, J. A.; Roecker, A. J.; Cao, G.-Q.; Barluenga, S.; Mitchell, H. J. *J. Am. Chem. Soc.* **2000**, *122*, 9939−9953.
(12) Nicolaou, K. C.; Pfefferkorn, J. A.; Mitchell, H. J.; Roecker A. J.; Barluenga, S.; Cao, G.-Q.; Affleck, R. L.; Lillig, J. E. *J. Am. Chem. Soc.* **2000**, *122*, 9954−9967.

10.1021/ac025534s CCC: $22.00 © 2002 American Chemical Society
Published on Web 07/18/2002

combinatorial libraries has depended on automated HPLC separation using fast gradient elution protocols to purify the compounds.[13−15] Subsequent purity assessments and confirmation of synthesized compounds were achieved by liquid chromatography−mass spectrometry (LC−MS) coupled with ultraviolet-evaporative light scattering detection (UV-ELSD) or other detection methods.[16,17] To increase the throughput of the purification process, parallel preparative high-performance liquid chromatography (HPLC) systems were developed, which allow one HPLC system to separate 2−4 times as many samples as a conventional (single-channel) HPLC system.[18,19] In addition, the development of a four- or eight-channel mass spectrometry interface has allowed parallel LC−MS systems to analyze combinatorial libraries at ultrahigh-throughput level.[20−22] We have recently reported the application of parallel eight-channel liquid chromatography−evaporative light scattering detection-mass spectrometry (LC−ELSD-MS) in the detection of natural products, including flavonoids, alkaloids, terpenoids, and steroids.[23] Also, the recent progress in the construction of high-field magnets, pulsed-field gradients, solvent suppression, and flow probe technology has given new stimulus to LC−NMR, which has emerged since the mid-1990s as a powerful method for the on-line identification of organic molecules.[24]

Our objective has been to combine these advances in high-throughput technologies with more conventional natural products chemistry techniques to rapidly generate large libraries of purified fractions of small-molecule natural products for HTS. Our production process begins with the fractionation of polar and nonpolar plant extracts using automated flash chromatography.[30] The resulting flash fractions are subjected to batch-throughput solid-phase extraction to remove tannins and molecular weight cutoff filters to remove high molecular weight components in flash fractions from polar extracts. Flash fractions are further separated using a high-throughput parallel four-channel reversed-phase preparative HPLC system, resulting in fractions containing a mixture of ∼5 compounds/well. The fractions are analyzed by a high-throughput parallel eight-channel analytical LC−ELSD-MS system. The fractions containing detectable compounds are collectively called "the library" from which more focused libraries are drawn for biological screening.

Currently, the library consists of 36 000 fractions of approximately one to five small-molecule natural products per fraction. They originate from a diverse collection of 176 plant families and 561 genera from the United States and Gabon, Africa. After biological screening, the individual compounds of the active fractions are rapidly purified to confirm biological activity. Using a Bruker Avance 600-MHz NMR spectrometer equipped with a new capillary (5 μL) microcoil flow probe with an active volume of only 1.5 μL, 5−10 μg of a pure compound is sufficient for ^{1}H and correlation spectroscopy (COSY) experiments to dereplicate known structures. To characterize novel structures, ∼50 μg of pure compound is needed to acquire additional experiments such as a gradient heteronuclear multiple-quantum coherence (g-HMQC) and a gradient heteronuclear multiple-bond correlation (g-HMBC).

We describe here the sample preparation and high-throughput methods used in our library production process and illustrate it using an extract of *Taxus brevifolia*, the pacific yew tree containing paclitaxel (Taxol) and its derivatives. The resulting *Taxus* library was analyzed by parallel eight-channel LC−ELSD-MS. Library fractions were submitted to the National Cancer Institute (NCI) for screening in three cancer cell lines. Screening results indicate that natural product libraries, produced as described here, enable the discovery of active, otherwise undetected minor metabolites. Several flash chromatography fractions were obtained during the production of the *Taxus* library that showed no biological activity. However, library fractions that resulted from these flash chromatography fractions did show activity in the same assays. The active compounds, 7-(β-xylosyl)taxol, 7-(β-xylosyl)taxol C, and 7-(β-xylosyl)-10-deacetyltaxol C, were quickly dereplicated using their molecular weights determined during LC−ELSD-MS analysis of the library, and the structures were characterized using ^{1}H and COSY experiments. These compounds were previously reported from the stem bark of *Taxus baccata* and possess potent activity against B16 melanoma.[25,26]

EXPERIMENTAL SECTION

Instrumentation. Automated flash chromatography separations were performed on 50-g Si and C18 flash columns (International Sorbent Technology Ltd., Mid Glamorgan, U.K.) using a Flash Master II automated chromatographic system (Jones Chromatography Inc., Lakewood, CO). The removal of tannins was performed using a 500-mg or a 2.5-g polyamide-filled cartridge (Jones Chromatography Inc.).

Preparative HPLC separations were performed on Betasil C18 columns (20 × 100 mm, 5 μm, Keystone Scientific Inc., Bellefonte, PA). A parallel four-channel preparative HPLC system was assembled and consisted of four Beckman System Gold 126 gradient HPLC pumps (Beckman Coulter Inc., Fullerton, CA) with system controllers and four-way solvent delivery modules, four Beckman System Gold 166 single-wavelength UV detectors with preparative flow cells, a Gilson 215/849 multiple probe autosampler (Gilson Inc., Middleton, WI), and four Gilson 204 fraction

(13) Weller, H. N.; Young, M. G.; Michalczyk, S. J.; Reitnauer, G. H.; Cooley, R. S.; Rahn, P. C.; Loyd D. J.; Fiore, D.; Fischman, S. J. *Mol. Diversity* **1997**, *3*, 61−70.
(14) Zeng, L.; Burton, L.; Yung, K.; Shushan, B. Kassel, D. B. *J. Chromatogr., A* **1998**, *794*, 3−13.
(15) Weller, H. N. *Mol. Diversity* **1999**, *4*, 47−52.
(16) Fang, L.; Wan, M.; Pennacchio, M.; Pan, J. *J. Comb. Chem.* **2000**, *2*, 254−257.
(17) Taylor, E. W.; Qian, M. G.; Dollinger, G. D. *Anal. Chem.* **1998**, *70*, 3339−3347.
(18) Zeng, L.; Kassel, D. B. *Anal. Chem.* **1998**, *70*, 4380−4388.
(19) Schultz, L.; Garr, C. D.; Cameron, L. M.; Bukowski, J. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2409−2414.
(20) Kassel, D. B.; Wang, T.; Zeng, L. US Patent, 6,066,848, May 23, 2000.
(21) Wang, T.; Zeng, L.; Cohen, J.; Kassel, D. B. *Comb. Chem. High Throughput Screening* **1999**, *2*, 327−334.
(22) De Biasi, V.; Haskins, N.; Organ, A.; Bateman, R.; Giles, K.; Jarvis, S. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 1165−1168.
(23) Cremin, P.; Zeng, L., submitted to *Anal. Chem.*
(24) Wolfender, J. L.; Ndjoko, K.; Hostettmann, K. *Phytochem. Anal.* **2001**, *12*, 2−22.
(25) Sénilh, V.; Blechert, S.; Colin, M.; Guénard, D.; Picot, F.; Potier, P.; Varenne, P. *J. Nat. Prod.* **1984**, *47*, 131−137.
(26) Murray, C. K.; Bailey, D. T.; Peterson, S. K.; Beckvermit, J. T. U.S. Patent 5,356,928, October 18, 1994.
(27) Olson, D. L.; Peck, T. L.; Webb, A. G.; Magin, R. L.; Sweedler, J. V. *Science* **1995**, *270*, 1967−1970.
(28) Olson, D. L.; Lacey, M. E.; Sweedler, J. V. *Anal. Chem.* **1998**, *70*, 645−650.
(29) Anderson, K. J.; Teuber, S. S.; Gobeille, A.; Cremin, P. A.; Waterhouse, A. L.; Steinberg, F. M. *J. Nutr.* **2001**, *131*, 2837−2842.
(30) Still, W. C.; Kahn, M.; Mitra, A. *J. Org. Chem.* **1978**, *43*, 2923−2925.

collectors. The system was controlled by Beckman 32 Karat chromatography software.

A Mega 1200 evaporator (Genevac Technologies, Suffolk, U.K.) was used to remove solvents from the preparative HPLC fractions. The preparative HPLC fractions were transferred from tubes to 96-deep-well plates by a Packard MultiProbe II liquid handling system (Packard BioScience Co., Meriden, CT). Focused libraries for screening were prepared in either 96- or 384-well plates using the same liquid handling system. A Genevac HT-12 evaporator was used to remove solvents from the 96- and 384-well plates.

A parallel eight-channel LC−ELSD-MS system was assembled and consisted of a LCT time-of-flight mass spectrometer with an eight-way MUX electrospray interface (Micromass Ltd, Manchester, U.K.), a Waters 600E Multisolvent Delivery System (Wasters Corp., Milford, MA) to pump solvents through an eight-way manifold that splits the flow to eight HPLC columns (4.6 × 50 mm, 3 μm, Keystone Betasil C-18), a Gilson 215/889 multiple probe autosampler, and eight Alltech 500 ELSD detectors (Alltech Associates Inc., Deerfield, IL). The system was controlled by MicroMass MassLynx software. Data analysis was performed using the OpenLynx Software followed by Extractor, a customized software package developed for Sequoia Sciences by Koch Associates (La Jolla, CA).

The isolation of individual compounds was performed using semipreparative Keystone Betasil C18 or C8 columns (8 × 250 mm i.d., 5 μm) on a single-channel Beckman HPLC system consisting of a Beckman 168 diode array UV detector, Sedex ELSD detector (Richard Scientific Inc., Novato, CA), and Gilson 204 fraction collector (a splitter is used to split the flow in 10:90 to ELSD and fraction collector). Size exclusion chromatographic analyses were conducted on a single-channel analytical Beckman HPLC system using Macrosphere GPC column (4.6 × 250 mm, 7 μm, Alltech Associates, Inc.) and Sedex ELSD detector.

NMR data for the structure elucidation of compounds were acquired by utilizing a Bruker Avance 600-MHz NMR system (Bruker, Rheinstetten, Germany) and a 5-μL capillary microcoil NMR flow probe with 1.5-μL active volume (Magnetic Resonance Microsensors, Savoy, IL), a Harvard 22 syringe pump (Harvard Apparatus Inc., Holliston, MA), and a Valco six-port injection valve (Valco Instruments Co. Inc., Houston, TX) with a 3-μL loop as the sample loading device.

Plant Collection. Plants were collected in Gabon, Africa, or in the United States by the Missouri Botanical Garden (St. Louis, MO). Plant samples consisted of the whole plant or separated plant parts such as roots, stems, leafs, flowers, and fruits or various combinations of parts. Voucher specimens were deposited at the Missouri Botanical Garden and the Institute of Pharmacopeia and Traditional Medicine under the National Center of Scientific Research and Technology (Libreville, Gabon). The plant samples from Gabon were dried immediately after collection above a gas-powered plant drier. Plant samples from the United States were shipped frozen to Sequoia Sciences. Frozen plants samples were lyophilized upon arrival. Low-purity Taxol (extract of the stem bark of *T. brevifolia* adsorbed onto silica) was purchased from Hauser Chemical Research Inc. (Boulder, CO).

Extraction Process. Dried plant material (150 g) was ground to a homogeneous powder. The powder was sonicated for 30 min in an organic solvent mixture of EtOH/EtOAc (50:50) followed by vigorous shaking for exhaustive extractions (two times, 4 and 8 h each). After filtration and removal of the organic solvents by rotary evaporation, the organic extract was obtained. The remaining residue was exhaustively extracted using an aqueous solvent mixture of H_2O/MeOH (30:70) (two times, 4 and 8 h each). The aqueous extract was obtained after removing the solvents by rotary evaporation. The low-purity taxol powder was exhaustively extracted with EtOH/EtOAc (50:50). After filtration, the taxol preparation was dried by rotary evaporation. This taxol extract (TX001) was treated as an organic extract.

Automated Flash Chromatographic Separation. Organic extract material (1 g) was dissolved in 5 mL of MeOH/EtOAc (50:50) and adsorbed onto 5 g of silica powder. The dried powder was brought onto a 50-g silica column and eluted on the flash chromatography system using a step gradient of (1) 75% hexanes, 25% EtOAc, (2) 50% hexanes, 50% EtOAc, (3) 100% EtOAc, (4) 75% EtOAc, 25% MeOH, and (5) 50% EtOAc, 50% MeOH. The Flash Master II was modified to collect large fractions of 250 mL of solvent per gradient step. The system was set up to perform automated separations of 10 samples per loading. Flash fraction 1 was discarded, whereas fractions 2−5 were dried by rotary evaporation. Flash fractions 4 and 5 were screened for the presence of tannins by LC−MS and passed over a 2.5-g polyamide column if results were positive.[29] Flash fractions produced from the *Taxus* organic extract were named TX002−TX005.

Aqueous extract (2 g) was dissolved into 10 mL of water, and the resulting suspension was centrifuged. The aqueous layer was brought onto a 50-g C18 column (prerinsed with one column volume of methanol and five column volumes of water). Any insoluble material was again dissolved into 10 mL of water using sonication. The suspension was centrifuged again. The aqueous layer was also brought onto the column. The column was then rinsed with five column volumes of water, and the effluent was discarded. The remaining insoluble material was subsequently taken into 10 mL of methanol. The methanol layer was brought onto the column. The column was eluted with one column volume of methanol to remove water from the column, and a 500-mg polyamide cartridge in methanol was attached to the bottom of the column. The column was eluted with five column volumes of methanol. The resulting fraction (flash fraction 6, 100 mg) was dissolved in MeOH/H_2O (60:40, 15 mL) and filtered at 3000*g* for 8 h using Centricon filter units with a molecular weight cutoff of 3000. The retentate, typically 1−2 mL, was discarded. Analytical size exclusion chromatography showed that the content of high molecular mass constituents (>3000 Da) in the filtrate was reduced significantly from up to 75% to less than 10% of the total amount of material using ELSD detection.

Parallel Four-Channel Preparative HPLC Separation. Flash fraction material (50 mg) was dissolved into either 1000 μL of MeOH/EtOAc (70:30) (for flash fractions 2 and 3 of organic extracts) or 100% MeOH (for flash fractions 4 and 5 of the organic extracts and flash fraction 6 of the aqueous fraction) and filtered where necessary. The fractions were separated into 40 fractions (20 mL/min, 1 min per collection per tube) using the parallel four-channel preparative HPLC system. A different 35-min gradient was applied to each flash fraction for adequate separation: flash fraction 2, 40−80% acetonitrile in water; flash fraction 3, 30−70% acetonitrile in water; flash fraction 4, 20−60% acetonitrile in water;

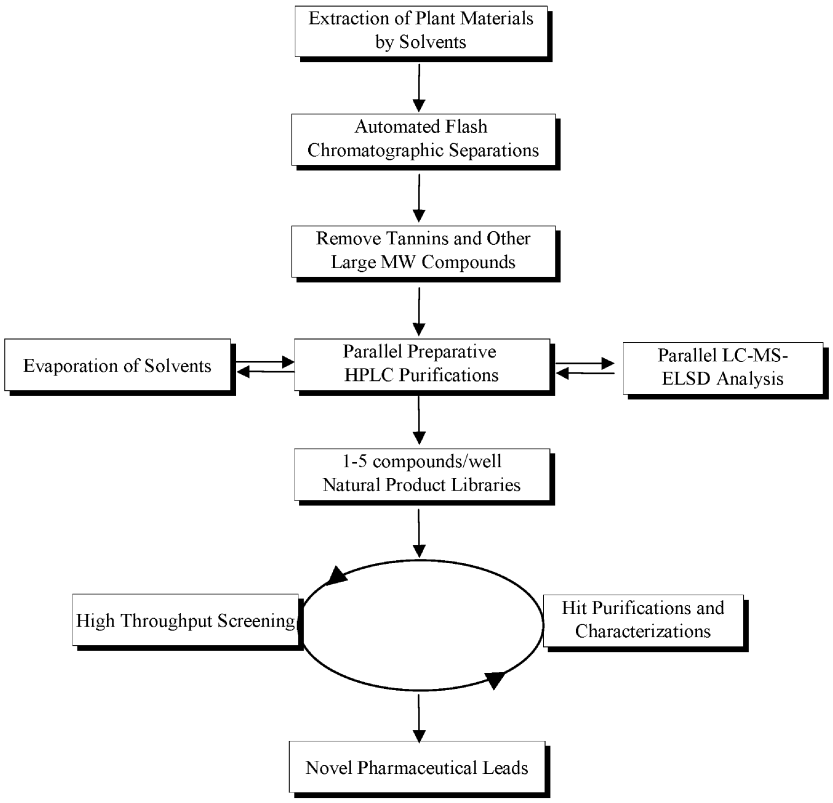

Figure 1. Schematic representation of high-throughput methods applied to the process of drug discovery from natural resources.

flash fractions 5 and 6, 10−50% acetonitrile in water. The 40 tubes containing HPLC fractions were dried in the Mega 1200 evaporator. The HPLC fractions were transferred to 96-deep-well plates using the liquid handling system (Packard MultiProbe II). A *Taxus* library was made from *Taxus* flash fractions (TX002−TX005) consisting of a total of 160 samples named TX002-1−40 to TX005-1−40, respectively.

Parallel Eight-Channel LC−ELSD-MS Analysis. All samples were analyzed by a parallel eight-channel LC−ELSD-MS system with chromatographic conditions of 5% acetonitrile in water for the first 1.0 min, a linear gradient of acetonitrile from 5 to 95% in 8.0 min, followed by 95% acetonitrile in water for 1.0 min. After each analysis, the column was equilibrated at 5% acetonitrile in water for 2.5 min.

The MS data were acquired on an LCT orthogonal TOF mass spectrometer (Micromass Ltd.) with an eight-channel multiplexed electrospray interface (MUX). MUX was operated under MassLynx V3.4 software. The instrument was operated in positive electrospray mode with capillary voltages of 3.5 kV being used. A desolvation temperature of 380 °C and a source temperature of 120 °C were used for all experiments. The nitrogen desolvation and nebulizer gas flow rates were set to 1100 and 300 L/h, respectively. Sample cone voltage was set to 50 V with extraction cone voltage set at 2.0 V. The rf lens voltage was set to 350V.

Data processing was performed automatically starting with OpenLynx, followed by a customized software package, Extractor, to automatically extract all graphic information, such as retention times, mass spectra, and peak integrations, and convert it to text to allow it to be transferred to a database for storage and analysis.

Biological Activity of the *Taxus* Library. In vitro cytotoxicity tests were conducted at the NCI using an in vitro three-cell-line panel consisting of MCF7 breast cancer, NCI-H460 lung cancer, and SF-268 CNS cancer. Each cell line was inoculated and incubated in microtiter plates. After 24 h, test samples were added to a final assay concentration of 2 μg/mL and the culture was incubated for 48 h. Results for each test sample were reported as the percentage of growth of the treated cells when compared to untreated controls. Compounds that reduced the growth of any one of the cell lines by 32% or less against standard were considered to be active.

Structure Elucidation of Bioactive Taxane Compounds. The constituents of active library fractions were purified using a single-channel HPLC system. The gradient applied to the separations was based on the elution profile observed during the preparative HPLC separation that created the fraction and was optimized for baseline separation of the compounds. The purification required ~100 μg/separation, and the yield per compound was typically in the range of 5−50 μg/compound. Pure compounds were dissolved into 3 μL of CD_3OD and loaded onto the microcoil NMR flow probe using a syringe pump equipped with a sample injection valve and capillary tubing. A sample of 5−50 μg was used to run a 1D ^{1}H spectrum (64 scans, 8 or 16 increments) and a gCOSY spectrum (256 scans, 16 increments). The probe was operated at a temperature of 293 K. Pulse widths were 5.5 μs at a power of 23 dB for the ^{1}H spectra. This information together with the molecular ions from the LC−ELSD-MS analysis was used to verify the structure of previously reported molecules. Novel structures for which 50 μg could be obtained were identified using additional experiments such as gHMQC, and gHMBC and high-resolution mass spectra were generated by TOF mass spectrometry for the determination of molecular formula.

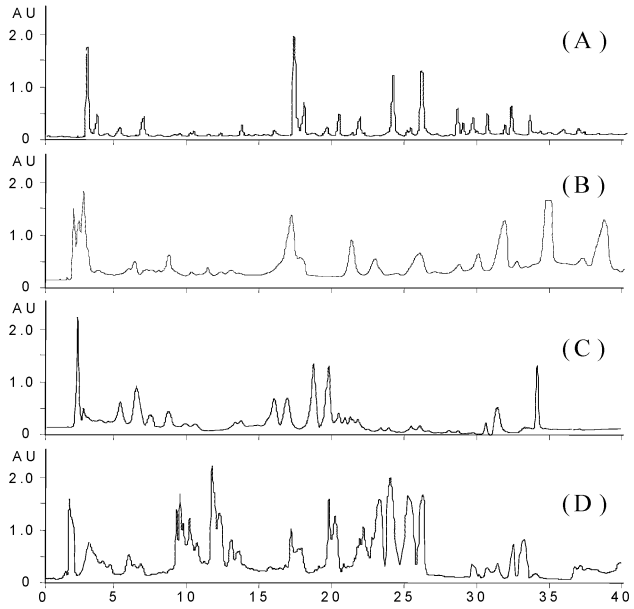

Figure 2. LC−UV chromatograms of four preparative HPLC purifications, each yielding 40 natural product library fractions. Chromatograms A−D show channels 1−4 in one parallel preparative HPLC run.

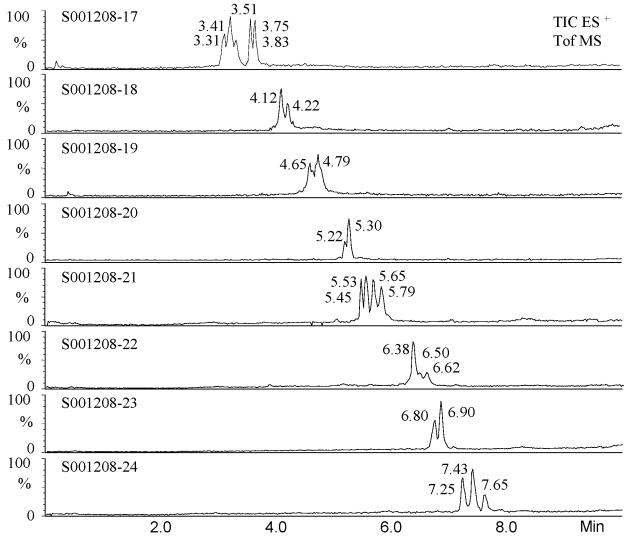

Figure 3. LC−MS (TIC) chromatograms of eight natural product library fractions (S001208-17−24) obtained during a single parallel LC−MS run.

RESULTS AND DISCUSSION

Preparation of Natural Product Libraries for Biological Screening. Most drug discovery programs today are capable of screening large numbers of compounds against multiple targets using nanogram to low-microgram quantities of material. To meet the demand for large numbers of structurally diverse compound libraries, we developed a high-throughput method to accelerate the natural product drug discovery process. Figure 1 is a schematic presentation that depicts the strategy by which we produced natural product libraries, screened them, purified active library fractions, and characterized individual, biologically active compounds for pharmaceutical discovery programs.

The process employed a parallel four-channel, reversed-phase preparative HPLC system for the production of library fractions. Reversed-phase, preparative HPLC is routine for the purification of pharmaceutical compounds, but the design and application of the automated, parallel four-channel preparative system has increased the efficiency of this technique 4-fold.[19] Our four-channel preparative HPLC system was customized in our laboratory. It operated four gradient pumping systems independently but simultaneously, permitting parallel separations of four samples per run. Since the systems were delivering the effluents independently to each preparative column, the separations at each channel had the same efficiency. Because of the complexity of natural product extracts, the separation of naturally occurring druglike compounds with acceptable resolution and yield requires several steps of cleanup and preseparation before samples are loaded onto a preparative column. Plant extracts were prepared for preparative HPLC using normal-phase or reversed-phase flash chromatography followed by solid-phase extraction over polyamide and filtration through molecular weight cutoff filters. These procedures remove highly lipophilic and hydrophilic compounds, pigments, large molecular weight tannins, polysaccharides, and other nondruglike molecules. Our optimized gradient for preparative HPLC chromatography separated 50 mg of a flash fraction per channel and collected 1 min/tube yielding 40 tubes/collection. Based on one sample per hour per channel, the parallel four-channel system was able to purify 32 samples in a working day, generating $32 \times 40 = 1280$ fractions per day. Those preparative HPLC fractions containing quantifiable compounds, ∼60% of the total, constituted the library. These fractions consisted of approximately 1−5 compounds/well and primarily from 0.1 to 1 mg of material. Using these methods, a library consisting of 36 000 fractions containing 1−5 compounds/well was produced. Figure 2 is a typical example of chromatograms obtained from one preparative HPLC run for the parallel purification of four library fractions.

Library production was validated by means of processing a single sample repeatedly and comparing the resulting LC−ELSD-MS data of the preparative HPLC fractions. The LC−ELSD-MS information of these samples proved to be very reproducible and accurate.

Analysis of the Library by Parallel Eight-Channel LC−ELSD-MS. To enable the analysis of large numbers of natural products contained in the library, we developed a method using a parallel eight-channel LC−ELSD-MS system.[23] Parallel LC−MS technology has recently been introduced to combinatorial chemistry.[20−22] The method employs a multiple sampling mass spectrometry interface now referred to as "MUX technology". In addition to the interface, the system used during analysis of the library incorporated ELSD as a quantitative tool to determine the quantity of compounds in each fraction.[23] ELSD is considered a universal detector with relatively good sensitivity and accuracy.[16] Fractions produced by parallel preparative HPLC were transferred to 96-deep-well plates, rows 2−11. Prior to LC−ELSD-MS analysis of a plate, row 1 containing a mixture of three standards was injected to ensure acceptable system performance. The system operated at 10 min/run with 8 samples/run, one plate every 2.5 h, 8 plates/day, and 40 plates/week. After large numbers of samples were analyzed, the retention time of the standards shifted slightly. Therefore, standards served as reference points and retention times were normalized.

Table 1. LC−MS Data of Sample S001208 Fractions 17−24[a]

samples	components per sample	retention time (min)	M + H
S001208-17	5	3.31	491.4
S001208-17	5	3.41	404.5
S001208-17	5	3.51	238.5
S001208-17	5	3.75	597.2
S001208-17	5	3.83	475.5
S001208-18	2	4.12	371.2
S001208-18	2	4.22	497.5
S001208-19	2	4.65	509.4
S001208-19	2	4.79	509.4
S001208-20	2	5.22	509.4
S001208-20	2	5.30	799.5
S001208-21	4	5.45	922.2
S001208-21	4	5.53	1207.3
S001208-21	4	5.65	936.3
S001208-21	4	5.79	813.2
S001208-22	3	6.38	904.5
S001208-22	3	6.50	537.4
S001208-22	3	6.62	537.2
S001208-23	2	6.80	1052.2
S001208-23	2	6.90	918.4
S001208-24	3	7.25	1066.3
S001208-24	3	7.43	1066.2
S001208-24	3	7.65	1062.4

[a] Mass spectra were acquired by electrospray positive ion mode.

Data processing by the workstation of the parallel eight-channel LC−ELSD-MS was quite a challenge because each file contained large amounts of information. A software package, Extractor, was developed to process the data and obtain information such as molecular weights, number of compounds per well, retention times, and peak integrations. Subsequently, the software compared samples with standards to correct for retention time shifts. Ultimately, data were exported in text format to a database and used for characterization and dereplication. Figure 3 shows a typical series of parallel eight-channel LC−MS data from the library. Mass spectra of relevant peaks for sample S001208 fraction 17−24 are presented in Table 1.

Sample Tracking and Data Archiving. Sample tracking and archiving of data are important when large natural product libraries are produced. The taxonomic identity of each plant sample was determined after collection by plant taxonomists at the Missouri Botanical Garden. If the various plant parts were separated, a single species of plant yielded multiple samples. Plant samples were numbered and assigned a unique bar code after collection. Their weights were recorded electronically together with their taxonomic information, collection location (GPS coordinates), and any other ecological information. As a plant or a sample was further fractionated, the resulting extracts or fractions were assigned new bar codes at every step of the process. The use of bar codes and computerized balances ensured electronic data entry of weights and sample identifications. The data were entered and stored in a combination of Microsoft Excel and Access tables with customized queries for sample selection and forms for data entry.

Purification of Hits and Confirmation of Biological Activities. Thus far, 36 000 natural product library fractions have been produced that have been screened in various drug discovery programs against different biological targets. The hit rates have been 0.5% or lower depending upon the biological assay. Active compounds were isolated in quantities of 5−50 μg from ~100 μg of a preparative HPLC fraction. Since elution conditions, quantities, and molecular ions of the compounds in the library were known, the isolation of individual compounds with purities greater than 85% became a process amenable to standardization. After pure compounds were obtained, their biological activities were confirmed. The standardization of the high-throughput library production and compound isolation process as described here significantly reduced the time needed to purify active materials.

Figure 4 is an example of the purification of a bioactive library fraction containing four peaks into a total of four individual components using a shallow-gradient, semipreparative HPLC separation. The sample loading onto the column was ~100 μg, and the recovery of each peak was ~5−50 μg.

Characterization of Hit Compounds. High-resolution NMR is a routine tool used by chemists to elucidate the structures of compounds. Since conventional NMR employs either 5- or 3-mm tubes, most laboratories need low-milligram quantities of sample to acquire all homo- and heteronuclear correlations for structure determination. If compounds are mass limited, as is the case with natural product drug discovery, obtaining low-milligram quantities

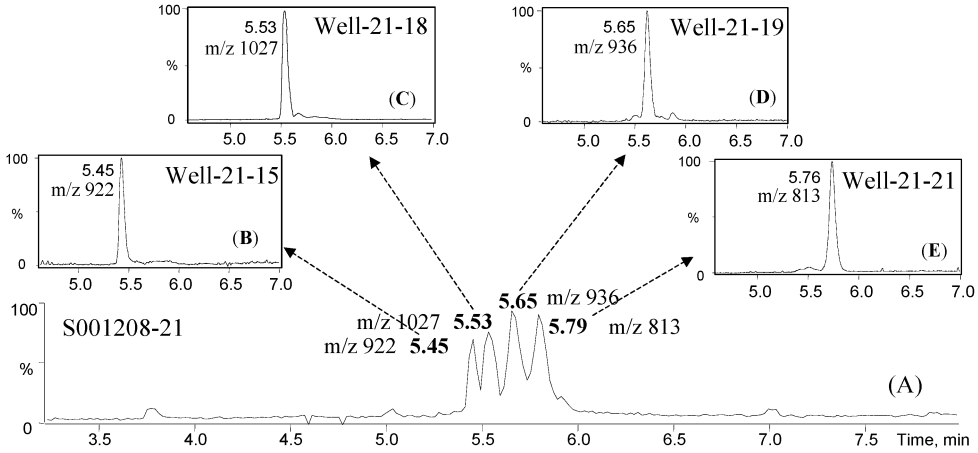

Figure 4. An example showing that a bioactive library fraction containing four components was separated into pure compounds by adopting the preparative HPLC condition as described in the text: (A) partial LC−MS (TIC) chromatogram of four components; (B−E) partial LC−ELDS chromatograms of each single compound after purification.

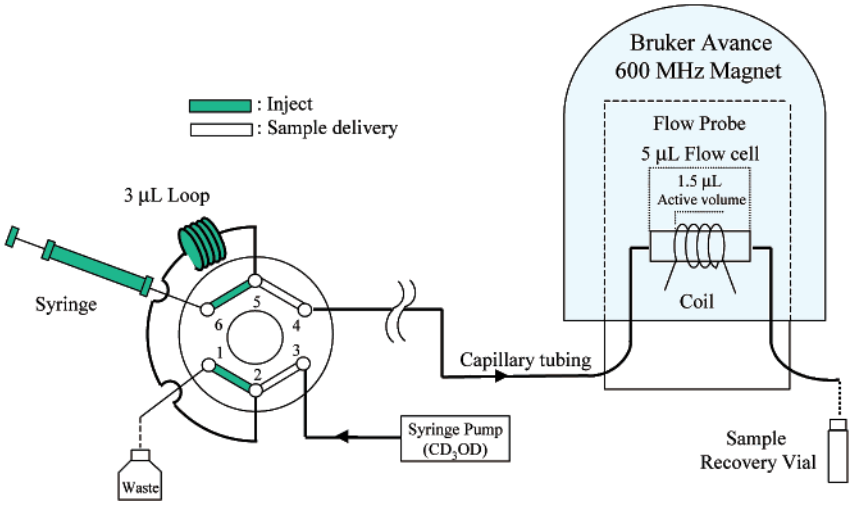

Figure 5. Schematic representation of the 600-MHz NMR spectrometer equipped with the 5-μL microcoil flow probe, syringe pump, and injection valve for low-microgram sample handling.

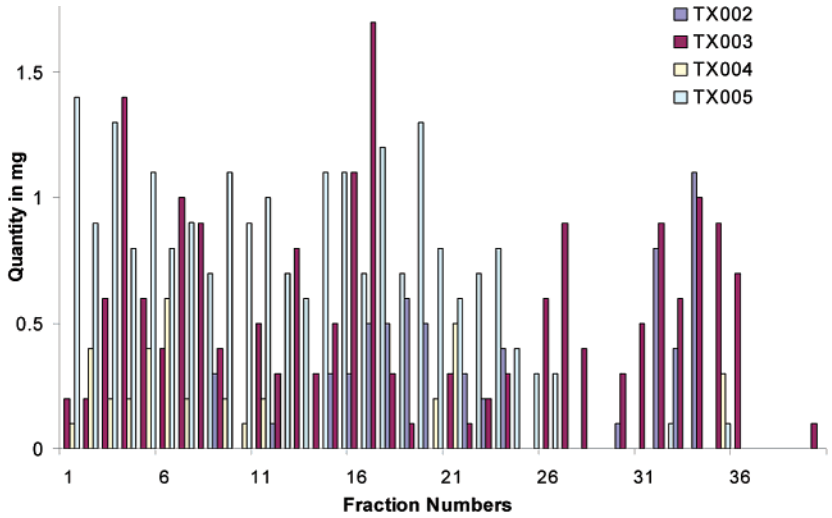

Figure 6. *Taxus* library made using our high-throughput method and the weights of the library fractions (in milligrams).

of sample requires multiple steps of separations and weeks or even months of time. With the advancement of NMR technologies, in particular the microcoil flow probe, with a capillary-based microliter-volume flow cell, the acquisition of NMR spectra on samples in trace quantities dramatically improves.[27,28] The new 5-μL microcoil flow probe reduces the active volume inside the coil to 1.5 μL, when compared to the conventional microliter-volume flow probes (40−200 μL) currently on the market, and the placement of the coil directly around the capillary results in further enhancement of mass sensitivity (signal-to-noise ratio per mass unit). Our characterization laboratory has implemented a system that permits us to routinely work with mass-limited samples at the low-microgram levels. As shown in Figure 5, a syringe pump operating at 5 μL/min pushed the sample to the microcoil probe in 2.5 min and parked the sample inside the probe. After the acquisition was completed, the sample was collected in recovery vials. Sample loading was typically done in 3 μL with 5−50-μg quantities. 1D ^{1}H NMR spectra and high-resolution mass spectra were generated on isolated compounds of 5−50 μg, providing the necessary data for characterization and dereplication.

The combination of our library production process and the high sensitivities of the latest NMR and LC−MS technologies greatly reduced the amount of material needed for structure elucidation and dereplication, consistent with the small amounts of material currently needed for HTS. In addition, structure elucidation was achieved within a fraction of the time needed when compared to conventional natural products chemistry. The possibility of working with such small quantities is an important step forward in natural product drug discovery, and these techniques will open doors enabling chemists to readily discover bioactive components among the minor constituents of natural resources.

Purification and Identification of Bioactive Taxane Compounds from *T. brevifolia*. Flash fractions (TX002−TX005) were produced by processing 1 g of organic extract (TX001) from *T. brevifolia*. As previously described here, a *Taxus* library was produced from these flash fractions by utilizing a parallel four-channel preparative HPLC system. A total of 160 preparative HPLC fractions were collected. Analysis of the fractions by parallel LC−ELSD-MS showed that the fractions primarily contained one to five components. The quantities of the fractions in the library were from 100 μg to 1.7 mg as determined by ELSD (see Figure 6). A total of 147 compounds were detected in this library by positive ion ESI mass spectrometry. Among the library fractions, paclitaxel was identified in sample TX003-17 (flash fraction TX003, fraction

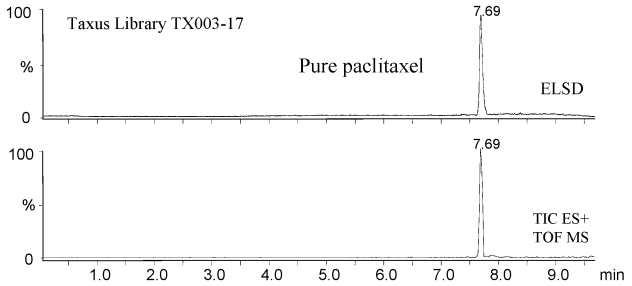

Figure 7. LC−ELSD-MS (TIC) chromatograms of a *Taxus* library fraction (TX003-17) obtained from our library production process, showing paclitaxel at a purity of over 95%.

Table 2. Results of One Dose Primary Anticancer Assays[a]

samples	MCF-7 breast cancer	H-460 lung cancer	SF-268 CNS cancer
TX001-1 (2 μg/mL)	31	91	75
TX001-2 (5 μg/mL)	23	27	55
TX001-3 (10 μg/mL)	17	25	43
TX002 (2 μg/mL)	94	117	110
TX003 (2 μg/mL)	19	23	29
TX004 (2 μg/mL)	23	61	60
TX005 (2 μg/mL)	85	107	105
TX003-17 (2 μg/mL)	21	30	29
TX005-28 (2 μg/mL)	21	71	49
TX005-30 (2 μg/mL)	30	93	55
TX005-31 (2 μg/mL)	14	43	38
TX005-32 (2 μg/mL)	15	45	52
paclitaxel (2 μg/mL)[b]	23	13	30
cephalomanine (2 μg/mL)[c]	15	19	30

[a] The numbers presented in the table showed the percentage of the reduced growth of the cell lines. Samples have less than 32% are active. [b,c] These are the standard compounds for positive control.

17) at the purity greater than 95% (Figure 7). All samples, including the original organic extract (TX001), flash fractions (TX002-005), and the *Taxus* library consisting of 160 preparative HPLC fractions were delivered to the NCI to be screened in vitro in the three-cell-line anticancer panel. The results are presented in Table 2 and show that the organic extract and flash fractions TX003 and TX004 exhibited anticancer activities. Sample TX0003-17, which proved to be pure paclitaxel, exhibited activity in all three cell lines. Samples TX005-28 and -30−32 also exhibited activity in the MCF7 breast cancer cell line assay, in which the original TX005 did not show activity (see Table 2).

The NMR data needed to identify the known taxanes (1D ^{1}H and COSY) were acquired using samples containing 5−50 μg of pure compound in 3 μL of CD_3OD in the microcoil flow probe. The samples were parked in the coil after two calibration runs using strychnine as a standard. A ^{1}H NMR spectrum of paclitaxel is shown in Figure 8 (50 μg of sample in 3 μL of CD_3OD). 1D ^{1}H and 2D COSY NMR spectra in combination with mass spectra allowed us to identify the major peaks from TX005-28, -30, and -31 as 7-(β-xylosyl)taxol, 7-(β-xylosyl)taxol C, and 7-(β-xylosyl)-10-deacetyltaxol C, respectively.[25,26]

CONCLUSION

High-throughput methods combined with conventional techniques for natural products chemistry can accelerate drug discovery from natural resources. We set out to automate and standardize the process by producing large libraries of small molecules from plants for high-throughput screening of pharmaceutical leads. In the high-throughput part of the process, each plant sample is separated by a parallel four-channel preparative HPLC into 200 fractions that are analyzed by a parallel eight-channel LC−ELSD-MS. On average, 60% of the analyzed fractions contain detectable compounds with one to five compounds per fraction. A total of 36 000 fractions containing detectable compounds were made, and these fractions are collectively called "the library" from which smaller, more focused libraries are drawn for screening.

Screening results indicate that hit rates are 0.5% or less and that the library facilitates the discovery of minor metabolites whose activity may go undetected upon the screening of crude extracts or even flash fractions. Since focused libraries are drawn using equal amounts of each library fraction, the concentration of minor metabolites in a screening assay is comparable to that of major metabolites.

Active compounds are rapidly purified from the library fractions for confirmation of biological activity. Characterization and struc-

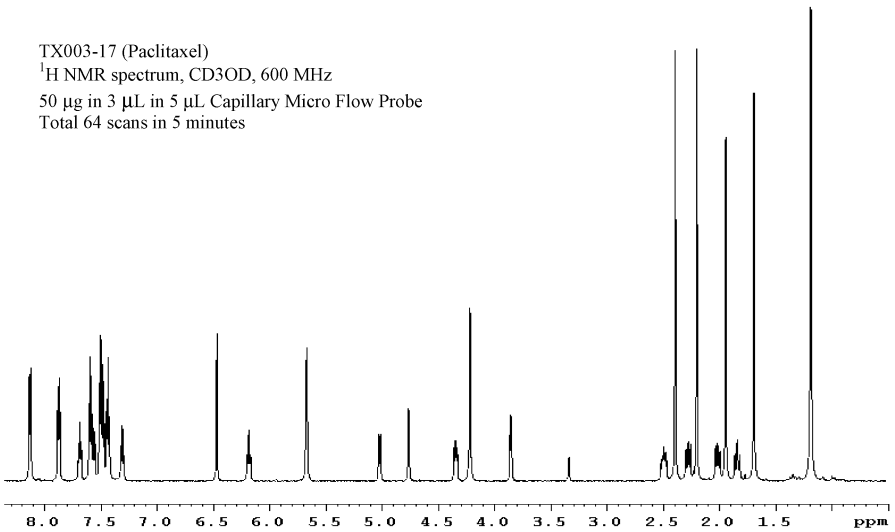

Figure 8. ^{1}H NMR spectrum of 50 μg of paclitaxel in 3 μL pf CD_3OD acquired using the 5-μL microcoil flow probe on the 600-MHz NMR spectrometer.

ture determination of individual compounds was done using the LC−ELSD-MS data and ^{1}H and COSY NMR experiments on as little as 5 μg using a Bruker Avance 600-MHz NMR spectrometer equipped with a new 5-μL microcoil flow probe. Using this new probe, the structures of novel compounds can be elucidated using as little as 50 μg with reasonable experiment times, increasing the throughput of the process. In addition, naturally occurring derivatives of active compounds can be quickly identified using the LC−ELSD-MS data and purified for structure−activity relationship studies during the screening process.

We believe that the production of natural product libraries for HTS and the ability to rapidly isolate and characterize bioactive library constituents reduces the cycle times for the discovery of natural product drug leads to levels comparable with combinatorial chemistry.

ACKNOWLEDGMENT

This project was partially supported by NIH SBIR Grant 1R43CA093098-01. The authors acknowledge Dr. Jim Miller, John Stone, Adam Bradley, and Gretchen Walters from Missouri Botanical Garden, for the plant collections and identifications, Matt Lesh, Peter Frank, Michael Sharpiro, Blanca Ruvalcaba, and Dr. Hye-Dong Yoo from Sequoia Sciences, Inc., for their technical assistance, and Dr. Charlie Koch from Koch Associate, La Jolla, CA, for programming the software that allowed the automated data processing of LC−ELSD-MS. We acknowledge Tim Peck, Dean Olson, and Jim Norcross from Magnetic Resonance Microsensors, Savoy IL, for making the first 5-μL proton indirect carbon gradient CapNMR flow probe available to Sequoia Sciences. We thank them for their valuable technical assistance.

Received for review January 22, 2002. Revised manuscript received April 11, 2002. Accepted June 18, 2002.

AC025534S