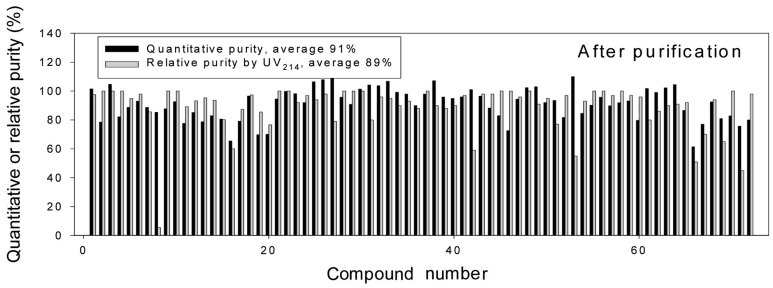
JOURNAL OF combinatorial CHEMISTRY

Article

Quality Control in Combinatorial Chemistry: Determination of the Quantity, Purity, and Quantitative Purity of Compounds in Combinatorial Libraries

Bing Yan, Liling Fang, Mark Irving, Sue Zhang, Armen M. Boldi, Frank Woolard, Charles R. Johnson, Tushar Kshirsagar, Gianine M. Figliozzi, Clinton A. Krueger, and Nathan Collins

J. Comb. Chem., **2003**, 5 (5), 547-559• DOI: 10.1021/cc030008f • Publication Date (Web): 31 May 2003



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 7 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

Journal of Combinatorial Chemistry is published by the American Chemical Society.
1155 Sixteenth Street N.W., Washington, DC 20036

Articles

Quality Control in Combinatorial Chemistry: Determination of the Quantity, Purity, and Quantitative Purity of Compounds in Combinatorial Libraries

Bing Yan,* Liling Fang, Mark Irving, Sue Zhang, Armen M. Boldi, Frank Woolard, Charles R. Johnson, Tushar Kshirsagar, Gianine M. Figliozzi, Clinton A. Krueger, and Nathan Collins

ChemRx Division, Discovery Partners International, Inc., 385 Oyster Point Boulevard, South San Francisco, California 94080

Received January 27, 2003

The quality of combinatorial libraries determines the success of biological screening in drug discovery programs. In this paper, we evaluate and compare various methods for measuring identity, purity, and quantity (yield) of combinatorial libraries. Determination of quantitative purity reveals the true library quality and often indicates potential quality problems before full-scale library production. The relative purity can be determined for every member in a large library in a high-throughput mode, but must be cautiously interpreted. In particular, many impurities are not observable by relative purity measurements using detectors such as UV_{214}, UV_{254}, and evaporative light-scattering detection. These "invisible" impurities may constitute a significant portion of the sample weight. We found that TFA, plastic extracts, inorganic compounds, and resin washout are among these impurities. With compelling evidence, we reach a conclusion that purification is the only way to remove "invisible" impurities and improve the quantitative purity of any compound even though some compounds may have a high relative purity before purification.

Introduction

As combinatorial chemistry[1] is becoming a more mature science, the urge for compound number is gradually converted to the desire for compound quality (purity and yield). High-throughput organic synthesis[2] is highly effective in producing a large number of compounds; however, the quality of compounds made in the past decade is only average or, in some cases, below average in organic chemistry standards. Three key parameters characterizing a synthetic product are its identity, purity, and yield (quantity). In traditional organic synthesis, identity is obtained primarily by NMR, purity by elemental analysis, and quantity by weighing the purified compound. The application of these compounds afterward, such as in biological screenings, is therefore unambiguous and quantifiable. However, the compound characterization has undergone a dramatic shift from traditional organic synthesis to combinatorial synthesis. Because of the limited throughput of traditional analytical methods, the compounds in a very large library, each present at a very low quantity, cannot be thoroughly characterized. NMR cannot be used for analyzing every member in large combinatorial libraries because of the slow spectral interpretation process, even though the high-throughput spectral acquisition has been achieved.[3] Elemental analysis is not feasible because of its low throughput and the inadequate sample quantity from combinatorial synthesis. Weighing is cumbersome for compounds synthesized in individual vials and is not feasible for synthesis performed on a microtiter plate. Because of these deficiencies, most combinatorial compounds that are used in high-throughput screening so far lack some of the information on identity, purity, and quantity.

Using compounds of questionable quality in high-throughput biological screening always generates ambiguous results. The frequent occurrence of false positives and the lack of reproducibility have hindered the lead discovery process, because it takes much more time and resources to reanalyze all the false positive hits. On the other hand, the frequent occurrence of false negatives defeats the purpose of lead discovery. In the past decade, we have developed technologies to generate hundreds-fold more compounds. However, we have not been able to obtain hundreds-fold more knowledge, primarily because of the questionable quality of the compounds. Therefore, an in-depth understanding of the library quality and ways to determine and improve the purity and yield of desired compounds are of utmost importance in combinatorial chemistry. High-

* To whom correspondence should be addressed. Phone: 650-228-1293. Fax: 650-228-0138. E-mail: byan@chemrx.com.

10.1021/cc030008f CCC: $25.00
Published on Web 05/31/2003

throughput purification[4] has been adopted as a way to improve the quality of combinatorial libraries. However, these three difficulties need to be resolved: selection of proper methods for purity determination before and after purification, assessment of the reaction yield and the quantity of library compounds before purification, and evaluation of the purification efficiency.

The meaning of purity often varies with the ways it is measured. In this paper, we evaluate various methods for measuring identity, purity, and quantity in combinatorial library analysis. We also assess advantages and limitations of current purity and quantity analysis methods and identify key technical challenges in characterization of combinatorial libraries. We classify experimentally measured purities into relative and quantitative purities. We further analyze discrepancies among them. We found that the existence of impurities invisible in UV_{214} and evaporative light-scattering detection (ELSD) measurement can only be detected by quantitative purity measurement. The sources of "invisible" impurities are also investigated and discussed. We conclude that high-throughput purification is the only way to remove all impurities and produce a high quality combinatorial library.

Materials and Methods

Calibration Curves for Quantitative Purity Measurement. HPLC separation was performed on a HP1100 system (Agilent, Palo Alto, CA), which consisted of a vacuum degasser, binary pump, autosampler, column compartment, and diode array detector. Data were processed by HP Chemstation software. Reversed-phase HPLC was carried out on a C18 column (3.0 × 100 mm, 5 μm, 100 A) from Phenomenex (Torrance, CA) at 40 °C with a flow rate of 1.0 mL/min. Two mobile phases (mobile phase A: 99% water, 1% acetonitrile, 0.05% TFA; mobile phase B: 1% water, 99% acetonitrile, 0.05% TFA) were employed to run a gradient condition from 0% B to 100% B in 6.0 min, 100% B for 2.0 min and reequilibrate at 0% B for 2.0 min. An injection volume of 10 μL was used.

All standards were weighed to the nearest 0.02 mg on an AT261 DeltaRange analytical balance by Mettler (Toledo, Columbus, OH). The stock solutions of 1.00 mg/mL (for making calibration curves) for each standard were prepared using either methanol or acetonitrile as solvent. The stock solution was diluted to make a series of calibration solutions with concentrations of 0, 25, 50, 75, 100, 150, and 300 μg/mL. A solution from one of the six standards at 100 μg/mL was used as external standard (ES) to compensate for instrumental fluctuation and other systematic errors. Samples from each standard were analyzed on HPLC and monitored by UV at 214, 220, and 254 nm. The peak area at each concentration was divided by that of ES to give the peak ratio (peak ratio = (peak area) /$(\text{peak area})_{ES}$). A plot of peak area ratio vs concentration yielded calibration curves. The correlation coefficient (R^2) was >0.99 for all calibration curves.

To validate standard calibration curves, a 100 μg/mL solution of each standard was prepared and analyzed in triplicate by HPLC with UV detection at 214, 220, and 254 nm. The determined concentrations from their corresponding calibration curves should be accurate to ±5%.

Determination of the Quantitative Purity of QC Compound. The final library, among which six QC compounds were placed, was synthesized on 96-well plates in qualification and production libraries. Each QC compound was dissolved in the well by 2 mL of acetonitrile, and the solution was transferred to a preweighed vial. The well was then rinsed with three 0.5-mL aliquots of acetonitrile, and the rinse solution was also transferred to the vial. More volatile solvent was first removed on a rotary evaporator. The vial was then lyophilized for 14 h at 14 mTorr to remove trace amount of moisture before weighing. To ensure complete moisture removal, the vial was lyophilized for two additional hours and weighed again until the relative weight change was <0.3%.

Assuming these compounds were pure, a solution of ~200 μg/mL was made using either methanol or acetonitrile. These samples were analyzed by HPLC/UV in triplicate. The concentration of QC compounds was determined from the peak area ratio relative to ES and individual standard calibration curves. The measured concentration should be within 1% from all three runs. The product quantity was determined by the actual concentration times the sample volume. The product purity was determined from the ratio of determined quantity to the total sample weight. The quantity and quantitative purity of these QC compounds were also determined by quantitative NMR (qNMR).

Quantitative NMR. Proton and fluorine NMR spectra were acquired on a JEOL (Peabody, MA) Eclipse 270 FT spectrometer with tunable probe and Stackman automatic sample changer. Delta version 3.1 software controlled the instrument and performed data processing. For the ^{1}H NMR experiments, the probe was tuned to a frequency of 270.17 MHz with acquisition parameters as follows: receiver gain 22, pulse width 10.4 μs (π/2), spectral width 4053 Hz, offset 5 ppm, digital resolution 0.25 Hz, data acquisition time 4.04 s, relaxation time 25 s, and total number of scans 32. For the ^{19}F NMR experiments the probe was tuned to 254.18 MHz with the following acquisition parameters: receiver gain 18, pulse width 10.0 μs (π/2), spectral width 50.8 kHz, offset −100 ppm, digital resolution 3.10 Hz, data acquisition time 0.32 s, relaxation delay 5 s, and total number of scans 32.

The following equation was used for quantification:

$$C_a = A_a M_a C_s n_s / n_a A_s M_s$$

where C represents the concentration, n is the number of nuclei responsible for a given peak, A is the area under the peak, and M is the weight. The subscript "s" represents the values for the standard peaks, and the subscript "a" represents the values for the analyte peaks.

QC compounds were weighed accurately (±0.02 mg), spiked with a known amount (4.41 mg) of 4,4-dimethoxybenzhydrol, and diluted to 1.00 mL with either methanol-d_4 or acetonitrile-d_3. From ^{1}H NMR data, concentration values were calculated by comparing the area of the standard resonance peaks to those of the sample.

Another issue in the development of the library synthesis protocol is the amount of trifluoroacetic acid (TFA) in the final product. Because TFA is used for cleavage of product from resin support and it is difficult to remove by vacuum pumping and lyophilization, there is always certain amount of TFA carried through to the end. The ^{19}F qNMR experiment was therefore performed on the same samples used for the proton qNMR work. In this analysis, a set of calibration solutions of fluorine standard α,α,α-trifluorotoluene were prepared, and an external calibration curve was made from α,α,α-trifluorotoluene in acetonitrile-d_3. This compound gives a single fluorine resonance at −63.1 ppm (relative to Freon $CFCl_3$). Fluorine NMR spectra of all the synthetic samples revealed one major fluorine resonance for TFA at −77.7 ppm. Concentration of TFA was determined by comparing the fluorine resonance area of the sample to the calibration curve of the standard.

Determination of Relative Purity by LC/MS/UV/ELSD. Analysis by LC/UV/ELSD/MS was performed using an API 150 EX instrument from PE Sciex (Concord, Ontario, Canada). The HPLC system consisted of a Gilson 215 liquid handler equipped with an 819 injection valve (Middleton, WI), a HP1100 vacuum degasser, binary pump, column compartment, and a diode array detector (Agilent Technologies, Palo Alto, CA). Eluent from HPLC system was split 1:5 to the mass spectrometer and a Sedex 55 (S.E.D.E.R.E., Alfortville Cedex, France) evaporative light-scattering detector. The mass spectrometer was operated in positive ion mode. The turbo ion spray conditions were as follows: temperature 400 °C, ion spray voltage 5000 V, curtain gas 12 (48 psi), and nebulizer gas 10 (40 psi). A full scan range from 150 to 800 amu in 1.5 s was used to acquire the MS data. The ELSD drift tube temperature was at 40 °C, the gain was set at 10, and the nitrogen flow rate was 3.3 L/min.

Signal from both UV_{214} and ELSD were collected through a PE Nelson 900 series interface to a Mac computer using MassChrom 1.1 at a rate of 50 data points/s. All peak areas and their qualitative peak purity for TIC, UV_{214}, and ELSD signals were automatically processed by a customized PurityScript in MultiView 1.4.

Reversed-phase HPLC was carried out on a Luna C18 column (2.0 × 30 mm, 5 μm, 100 A) from Phenomenex (Torrance, CA) at 40 °C with a flow rate of 3.0 mL/min. Two mobile phases (mobile phase A: 99% water, 1% acetonitrile, 0.1% acetic acid; mobile phase B: 1% water, 99% acetonitrile, 0.1% acetic acid) were employed to run a gradient condition from 10% B to 100% B in 3.0 min, stay at 100% B for 0.5 min, and reequilibrate at 10% B for additional 0.5 min.

The MS signal of MH^+ was used to identify product peak, the UV signal was used to assess product purity, and the ELSD signal was used to estimate product quantity. ELSD data were also used to estimate the yield of the desired product in each well.

Results and Discussion

Identity and Purity. Compound identification, the purity determination, and quantitation of the yield are three key elements for the quality control of combinatorial libraries. High-throughput identity determination in combinatorial chemistry is carried out exclusively by mass spectrometry. The resolution of a single quadrople ESI-MS instrument used for analyzing discrete combinatorial libraries is 1 Da. Assuming there is no interference of isobaric ions from impurities, a unit resolution MS measurement may be adequate to address the product identification need. However, this assumption is often invalid in reality. A more accurate mass measurement with high-throughput is needed to confirm the product formation and identify the side products. To fulfill this need, we implemented and optimized a parallel high-throughput accurate mass measurement system using a nine-channel multiplexed electrospray LC/UV/TOFMS system. Using this system, we achieved an accuracy of 10 ppm for 60% of compounds in diverse combinatorial libraries. We will report the system and our results on high-throughput accurate mass measurement separately.[5]

Relative Purity Determined by UV and ELSD. The purity is the percentage of the desired compound in a sample by weight. Impurities should include all chemical and nonchemical substances other than the desired compound. In reality, the purity is defined, measured, and reported in various forms.

The exact weight of the desired compound and its weight percentage in the sample is not always easily obtained in the combinatorial chemistry format because of the large number of compounds each present at very low amount. The most widely used method to measure the purity is based on separating components in a sample by reversed-phase HPLC and comparing the relative peak area of each component in the chromatogram as detected by UV or evaporative light-scattering detection. Purity reported using these methods is the relative purity, because these methods measure only the relative amounts of components that respond to a UV_{214} or ELSD detector on the basis of the assumption that all substances respond to the detector equally. There are at least two problems with these detection methods. First, not every component in a sample will respond to a specific detector. Second, the assumption of equal response is invalid. Because of technical limitations, this purity measurement method constitutes the vast majority of combinatorial library characterization.

Relative Purity 1: UV_{214} vs UV_{254}. Six compounds (Figure 1) selected from an alkoxyprolines library were studied for their concentration−response relationship using various detection methods (Figure 2). These compounds were synthesized on the basis of a common scaffold and, therefore, are structurally related. Calibration curves based on UV detection at 254 nm show a wide deviation from each other. This is because response factors of different compounds depend on extinction coefficients of their chromophore. The extinction coefficients of organic compounds at 254 nm are more diverse than those at 214 nm. Curves based on UV detection at 214 nm show less deviation from each other, indicating the 214 nm is closer to the common chromophore of organic compounds. In a library synthesis, most impurities are synthetic intermediates or starting materials with smaller MW and likely have less absorbance at 254 nm, as compared to the product. Purity measured at UV_{254} may overestimate

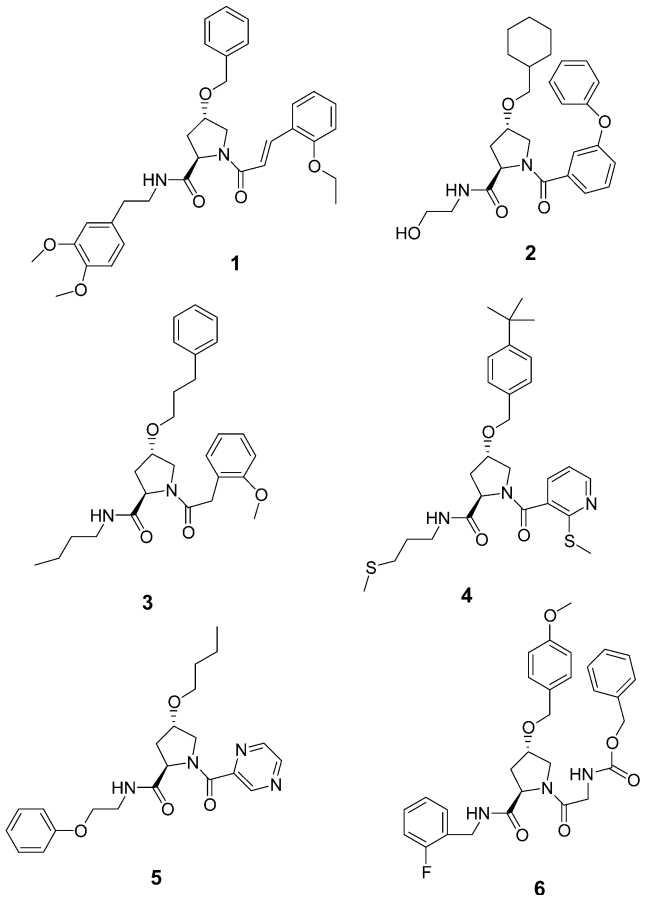


Figure 1. Chemical structures for compounds **1**−**6**.

the purity. A comparison of purity measurement at 214 and 254 nm for 70 compounds based on seven different scaffolds (from seven different libraries) is shown in Figure 3. Purity measured by UV_{254} is higher than that measured by UV_{214} for 80% of compounds in this random sampling experiment. In some cases, the purity gap is 20−50%, indicating the high variability of UV_{254} detection method.

Relative Purity 2: ELSD vs UV_{214}. In Figure 2, calibration curves based on ELSD detection show the smallest variation. These results suggest that the ELSD response factor of diverse compounds is less dependent on their chromophores[6] or structures, as compared to UV response. Because of the advantages of commercial ELSD detectors, they are widely used in the analysis of diverse organic compounds.

Although ELSD is a more generic detector for organic compounds, it also has some limitations. ELSD and UV_{214} purity determinations for 100 compounds from seven different libraries are shown in Figure 4. These data show a trend that the relative purity determined by ELSD detection is generally higher, as compared with that determined by UV_{214} (64% vs 49% for given compounds). We observed such an inconsistency in over 50 combinatorial libraries we analyzed using both ELSD and UV_{214} detection methods. By studying ELSD responses and quantitation of groups of compounds with different MW and volatility, we found that compounds with a MW < 300 generally give a response less than what is expected from their concentration.[7] Impurities in the final synthetic products are mostly starting

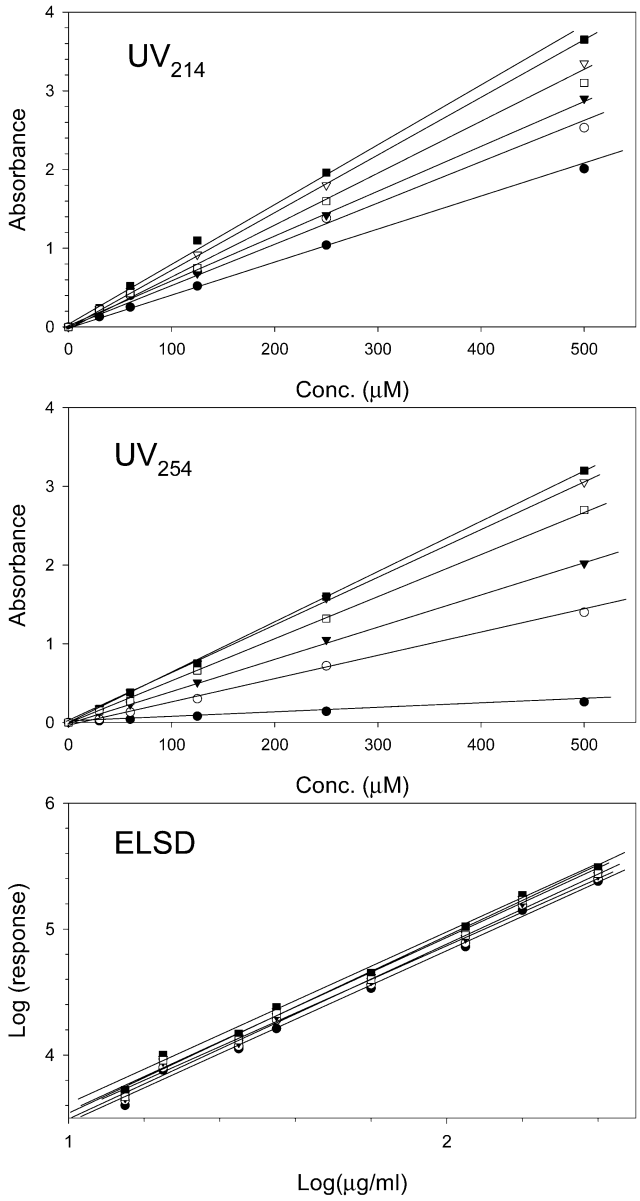


Figure 2. Concentration−response calibration curves for compounds **1**−**6**. The detection methods are UV_{214}, UV_{254}, and ELSD.

materials, decomposition product, or synthesis intermediates. These impurities generally have lower MW (more volatile) than the product. These more volatile molecules may evaporate with the solvent and result in less response. They may also form fewer scattering liquid droplets instead of solid particles after solvent evaporation. These liquid droplets scatter light poorly, and they may even absorb light instead of scattering light at certain wavelengths. Even though the ELSD response of these low-MW compounds is small, these compounds usually respond well to UV_{214} detection. On the basis of these findings and analysis, we conclude that UV_{214}, UV_{254}, and ELSD detections all have limitations. First, both UV and ELSD methods measure only the relative purity. Many impurities, such as TFA, inorganic salts, and high boiling point solvent, may adversely affect HTS, yet are undetectable by these methods. Furthermore, UV_{214} may give a wrong relative purity of the product as a result of the chromophore variation. Measurement using UV_{214} frequently underestimates the purity of compounds. On the other hand,

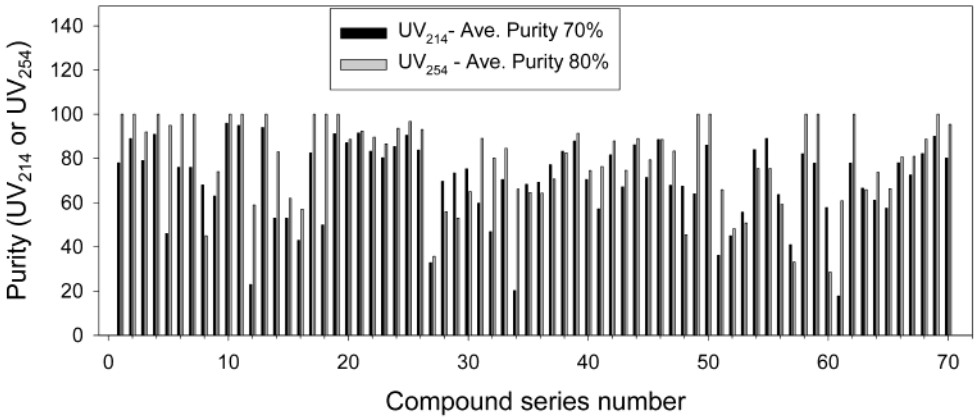


Figure 3. Relative purity determined by UV_{214} and UV_{254} detections for 70 compounds based on seven different scaffolds.

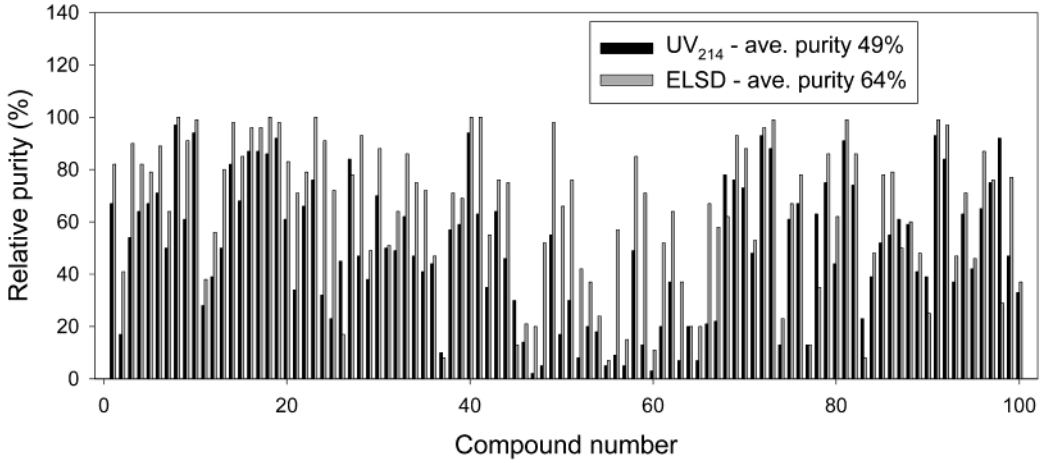


Figure 4. Relative purity determined by UV_{214} and ELSD detections for 100 compounds based on seven different scaffolds.

UV_{254} and ELSD overestimate the relative purity. A more cautious way is to describe the relative purity of a library by both UV_{214} and ELSD.

Relative Purity 3: HPLC with FIA-MS. There is no purity measurement without the identity determination first. Therefore, a purity determination is ideally performed by LC/MS analysis to facilitate separation, identity, and purity determination in the same run. Because of the low throughput and the high cost of LC/MS instruments, some laboratories use flow-injection MS and separate HPLC/UV analyses to characterize libraries. In this kind of analysis, determination of the purity is based on a very optimistic assumption that the largest peak in each chromatogram is the desired product. To evaluate the reliability of this approach, we analyzed 10 small libraries (200−400 compounds in each library) by LC/MS. The quality of synthesis for these libraries spans from high to low in terms of relative purity. Our data (Figure 5) show that the assumption that the largest peak is the product for purity determination is 95% correct for ∼20% of libraries with good quality (average relative purity ∼80%), 60% correct for 50% of libraries of average quality (average relative purity 65%) and only 25% correct for 30% libraries of low quality (average relative purity <45%).

The reliability of purity determination by assuming the largest peak to be the product is highly dependent on the quality of the library and individual compound. Applying this method for the analysis of libraries with unknown quality takes the risk of making a wrong assignment and, therefore,

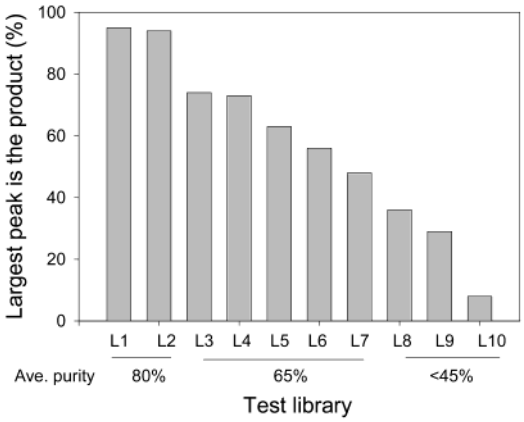


Figure 5. The success rate (%) to predict the largest peak as the product for 10 test libraries.

a wrong purity determination. The best identity/purity measurement is accomplished by LC/MS/UV_{214}. The problems of low throughput and high cost of LC/MS has now been addressed by applying parallel analysis configuration. We use two eight-way LC/MS/UV_{214} instruments to determine identity and purity of combinatorial compounds at a throughput of 4000 samples/day.[8]

Quantitative Purity (Absolute Purity). LC/MS/UV_{214} or LC/MS/ELSD measures only the relative purity. It is crucial to know the absolute purity of combinatorial library members and the difference between the relative purity and the quantitative (absolute) purity. The quantitative purity mea-

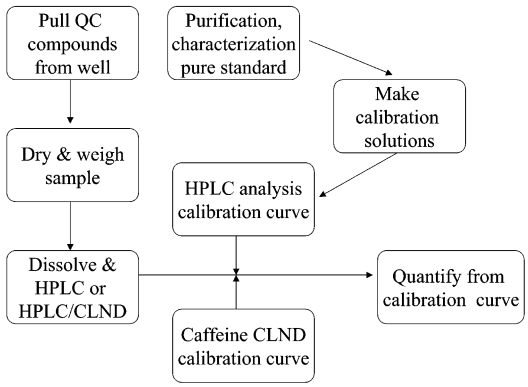


Figure 6. Flowchart for measuring the quantitative purity.

sures the percentage of the desired compound in a sample by absolute weight. The total weight of a sample is determined by weighing directly. The weight of the desired compound in a sample is obtained by LC/UV, LC/ELSD, LC/CLND, or NMR using a calibration curve generated from an authentic standard or an unrelated standard in the case of LC/CLND and NMR.

Quantitative Purity by Individual Calibration Method. We determined the absolute purity of six representative compounds in each library at various stages of library development and synthesis. We call them QC compounds. Compounds with these six structures were also synthesized separately and purified by reversed-phase and normal-phase HPLC. They are called standard compounds. Their role is to serve as authentic standard for the quantification of library compounds with the same structures.

After purification, standard compounds were rigorously characterized by ^{1}H and ^{13}C NMR and LC/MS, and their purity was higher than 99.6%, as determined by combustion elemental analysis. Standard compounds were accurately weighed to make a stock solution. The solution was then diluted to make a series of calibration solutions. These solutions were analyzed by HPLC, and a standard curve was generated on the basis of the compound peak areas in the chromatogram, as detected by UV at 214 and 254 nm. At the rehearsal and production library stages, libraries were synthesized in 96-well plates. QC compound sample was taken out of its well, and its total weight was determined after drying. The absolute weight of the target compound was deduced from the HPLC analysis of the compound and the standard curve. The quantitative purity was determined by the ratio of the pure compound weight and the total sample weight. Procedures for quantitative purity measurement are described in a flowchart (Figure 6). The quantitative purity values of QC compounds in a 1,2,5-trisubstituted benzimidazoles library (Figure 7) are shown in Table 1. In the first rehearsal library, QL#1, the quantitative purity is low, as determined by both individual calibration and qNMR methods. TFA was found to be the major impurity by ^{19}F NMR analysis. After incorporating an effective method to remove TFA, the synthesis purity (before purification) of QL#2 was much improved, as well as the final production library PL.

The above method for absolute purity determination is accurate, but requires individual authentic standard compound. The synthesis, purification, and characterization of these standards are time-consuming. Therefore, we also use quantitative NMR (qNMR) and LC/CLND to determine the quantitative purity. These methods require only a single unrelated external or internal standard.

Quantitative Purity by qNMR with a Single External Standard. The qNMR[9] is based on the fact that the peak areas of a given NMR resonance are directly proportional

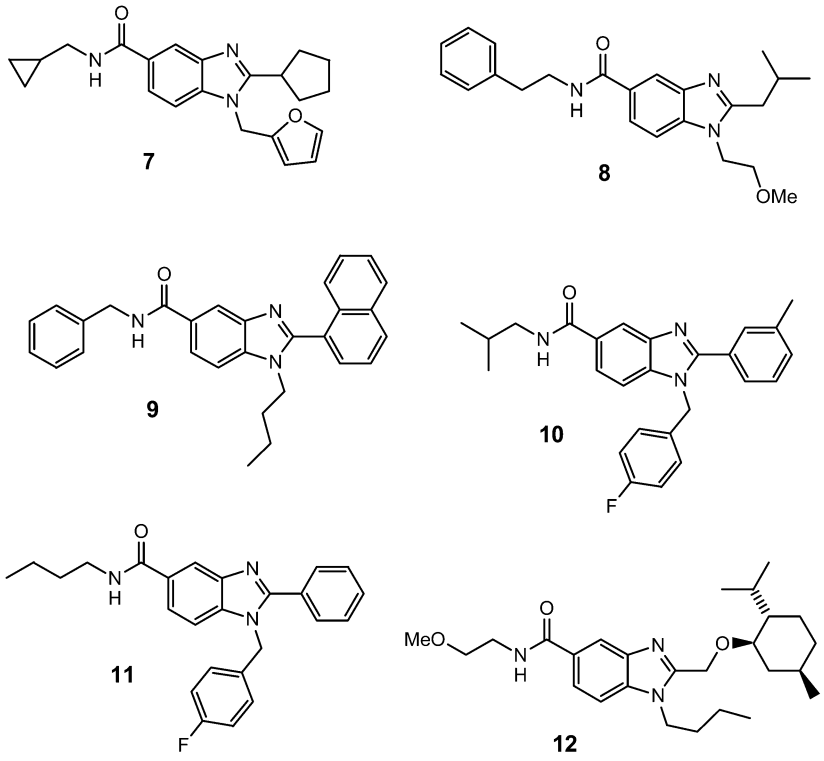


Figure 7. Chemical structures for compounds **7**−**12**.

Table 1. Quantitative Analyses of QC Compounds in a 1,2,3-Trisubstituted Benzimidazole Library[a]

	QL#1			QL#2	PL
QC compd	quant. purity (%)	qNMR purity (%)	TFA (%) by ^{19}F NMR	quant. purity (%)	quant. purity (%)
7	24.9	27.3	20.5	57.3	13
8	18.8	23.5	17.1	59.2	65.6
9	15.3	24.7	24.1	62.8	70.7
10	13.1	12.4	20.7	61.9	86.5
11	2.4	10.0	28	34.6	50.5
12	13.3	22.2	54.3	8.6	49.7

[a] Note: QL is qualification library, a small rehearsal library containing 100−200 compounds. PL is production library containing 4000−6000 compounds.

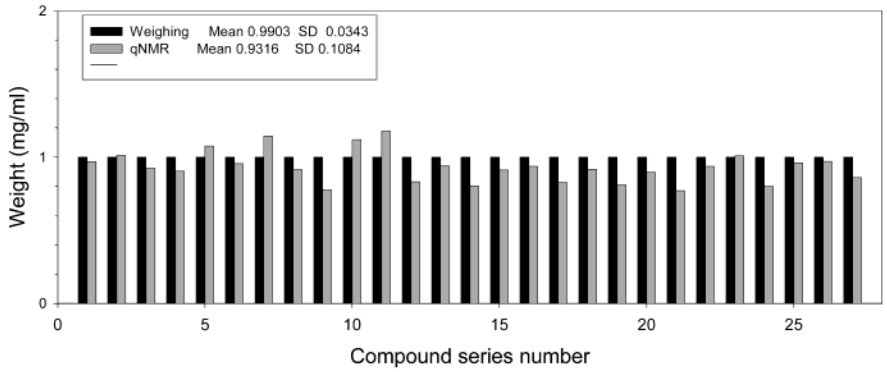


Figure 8. qNMR quantitation results for 27 compounds based on ofur different scaffolds, as compared to their weights.

to the molar amount of that nucleus in the sample. There are several advantages to using qNMR for the analysis of organic compounds: (1) The method is nondestructive; (2) besides the quantitative data, structural (identity) information of the compound is also gathered; and (3) high-throughput spectral acquisition instruments are commercially available.[3]

The main drawback of qNMR is that the spectral assignment is required. To date, there is no automated data interpretation software available. Second, some impurities may not have NMR signals and, therefore, are undetectable by this technique. QNMR can determine the quantity of a compound or the quantitative purity if the weight of the whole sample is determined. However, qNMR alone usually cannot measure the sample purity. Third, we found that the accuracy of the method is reduced if there are impurities in the sample. This is further discussed below.

For relatively pure compounds, qNMR can provide accurate quantitation results. Figure 8 summarizes data obtained from the analysis of 27 pure standard compounds (1 mg/mL) and the comparison of qNMR results (RSD 11%) to those generated by weighing (RSD 3%).

For samples that contain impurities, the assignment of the various peaks is more difficult. The overlapping peaks from impurities compromise the accuracy of quantification by the qNMR method. Table 1 shows the comparison of the qNMR results and the data from the individual calibration method. Higher values were obtained by qNMR for most samples. These results suggest that peaks from impurities may overlap with the product resonance and erroneously enlarge the product peak area used for quantitation. In combinatorial synthesis, impurities are unavoidable. Most impurities are starting materials, decomposition products, or intermediates. The NMR signals of these related compounds are more likely to overlap with NMR signals of the product. Therefore, the qNMR method should be used with caution in the analysis of combinatorial library samples.

Table 2. Quantitative Purity Determined by CLND

compd	MW	purity UV214 (%)	quantity CLND (mg)	sample weight (mg)	quant. purity (%)
A1	396.17	64.6	67.18	352.00	19.1
B1	458.18	77.63	108.63	483.00	22.5
C1	382.15	97.75	103.96	250.20	41.6
D1	424.2	90.42	139.15	300.90	46.2
E1	444.17	88.22	103.34	497.10	20.8
F1	464.14	95.6	115.85	323.90	35.8
G1	423.22	93.4	86.03	247.30	34.8
H1	477.15	37.06	91.43	314.60	29.1
A2	350.16	65.87	62.26	243.30	25.6
B2	412.18	65.25	86.69	289.20	30
C2	336.15	66.4	80.12	201.60	39.7
D2	378.19	76.83	76.13	246.60	30.8
E2	398.17	86.46	89.69	337.20	26.6
F2	418.14	94.64	79.22	271.90	29.1
G2	377.12	47.88	76.77	164.70	46.6
H2	431.14	34.08	33.60	394.10	8.5
av		73.88	87.51	307.35	30.425

Quantitative Purity Measured by LC/CLND. Drug molecules and libraries made for drug discovery generally contain nitrogen atoms (>90%). Therefore, HPLC with a chemiluminescence nitrogen detector (CLND) is a high-throughput method for quantitative analysis. CLND is a sensitive, nitrogen-specific detector. An unrelated compound is used to make an external calibration curve. On-line or direct-injection analysis gives quantitative results with a relative error of 10%.[10] When the weight of samples is obtained, the quantitative purity can be calculated from the absolute amount of the desired compound and the sample weight (Figure 6).

We determined the quantitative purity of 16 compounds in a library by LC/CLND and weighing methods (Table 2). Although the average quantitative purity was 30%, the absolute quantity of 87 mg indicates that the synthesis

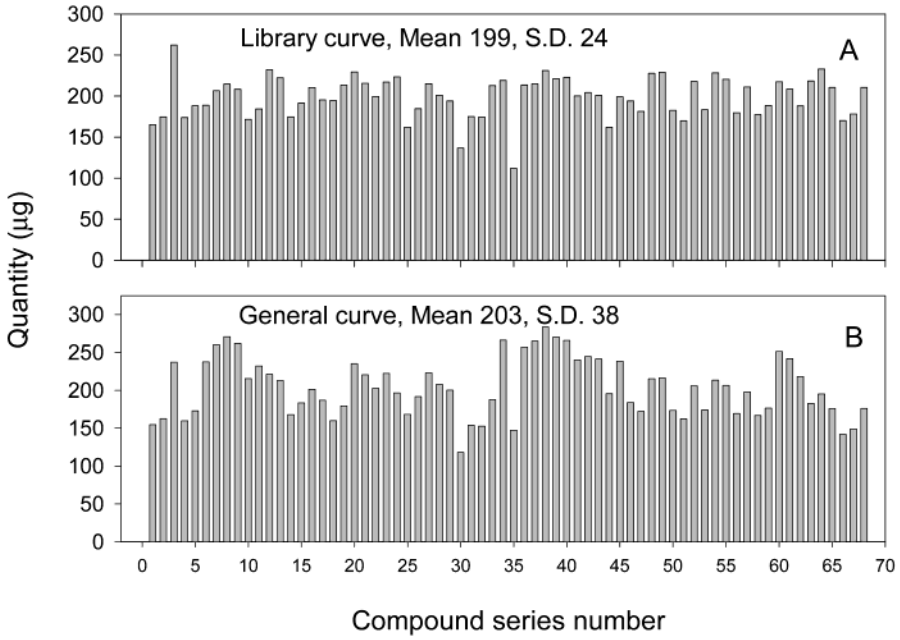


Figure 9. ELSD quantitation results for 68 compounds based on 12 different scaffolds. A. Quantitation using a scaffold-specific calibration curve made from compounds with the same scaffold. The RSD is ~10%. B. Quantitation using a general calibration curve made from all 68 compounds with different scaffolds. The RSD is ~30%.

material provides enough compounds for purification. The average of 307-mg sample weight and the average UV purity of 73% also suggest the existence of a large amount of impurities undetectable by the UV_{214} detector in this library.

High-Throughput Quantity (Yield) Determination. A high-throughput weighing station can weigh individual vials quickly if synthesis is carried out in the individual vial format. If the absolute amount of a desired compound can also be determined, the quantitative purity can then be obtained. On the other hand, if only the sample weight and the relative purity are known, it is not possible to determine the amount of the desired compound in the sample because of the presence of "invisible" impurities. The weighing method, however, cannot be used for compounds made in a 96-well plate. If it is not possible to get the sample weight, it is still informative to know the quantity of the desired compound in order to assess the reaction yield, to control the inventory, and to dilute compounds for biological screening with more reliable data. Two methods for quantifying compounds in a 96-well plate format are LC/MS/ELSD and LC/MS/CLND.

High-Throughput Quantification by LC/MS/UV/ELSD. As discussed above, the quantitative purity determined by an individual calibration curve with UV or ELSD detection (with a relative error of <5%) is the ideal information needed for each compound. However, for a combinatorial library containing ~5000 compounds in our synthesis, it is practically impossible to apply this analysis method to all compounds. Therefore, we investigated ELSD as a high-throughput method for quantitative analysis using a generic calibration curve.[6c] We carried out quantitative analysis on 68 pure compounds (>99.6%). These compounds were synthesized on the basis of 12 different scaffolds, and each compound was made at a concentration of 100 μg/mL. Quantitative measurements were performed using average standard curves made from structurally related (Figure 9A) and structurally different (Figure 9B) standard compounds. Our results show that, with these compounds, a RSD of 10% is achievable if the calibration curve is made from compounds with the same scaffold (Figure 9A) or a RSD of 30% if the calibration curve is made from unrelated compounds (Figure 9B). This method is not suitable to accurately determine the absolute weight of compounds. However, considering the volume of samples to be analyzed with this method, it can be useful for evaluation of a large number of compounds with a compromised accuracy. The ease of use, the robustness of the ELSD detector, and the compatibility with HPLC are welcomed in high-throughput synthesis laboratories.

High-Throughput Quantification by LC/MS/CLND. A previous paper has reported a linear correlation of CLND response with the amount of nitrogen for a set of commercial compounds with different numbers of nitrogen atoms.[10d] To validate our CLND instrument and protocol, we have expanded the scope of study to 15 compounds with the number of nitrogens varied from 1 to 6 in the molecule. These compounds are colchicine, diphenhydramine, and doxepin for one N per molecule; chlorpheniramine, diphenylalanine, and triprolidine for two Ns per molecule; dibucaine, *N*-(1-2-thiazolyl)sulfanilamide, 4-(dimethylamino)antipyrine, 3,8-diamino-6-phenylphenanthridine, and dibucaine for three Ns per molecule; caffeine, nalpha-(9-fluorenylmethoxycarbonyl)-L-arginine, and nialamide for four Ns per molecule; and our internal reference compounds HXZ002-A and HXZ002-B for five Ns and six Ns per molecule.

Nitrogen concentrations were varied from 0.02 to 10 mM. The plot of sample peak area ratio (peak area of sample divided by the peak area of 50 μg/mL of caffeine as the external standard) vs nitrogen concentration of the sample yielded a linear calibration curve for each of the 15 compounds (Figure 10). The slope average for 15 calibration curves is 1.036, with a relative standard deviation of 5.8%.

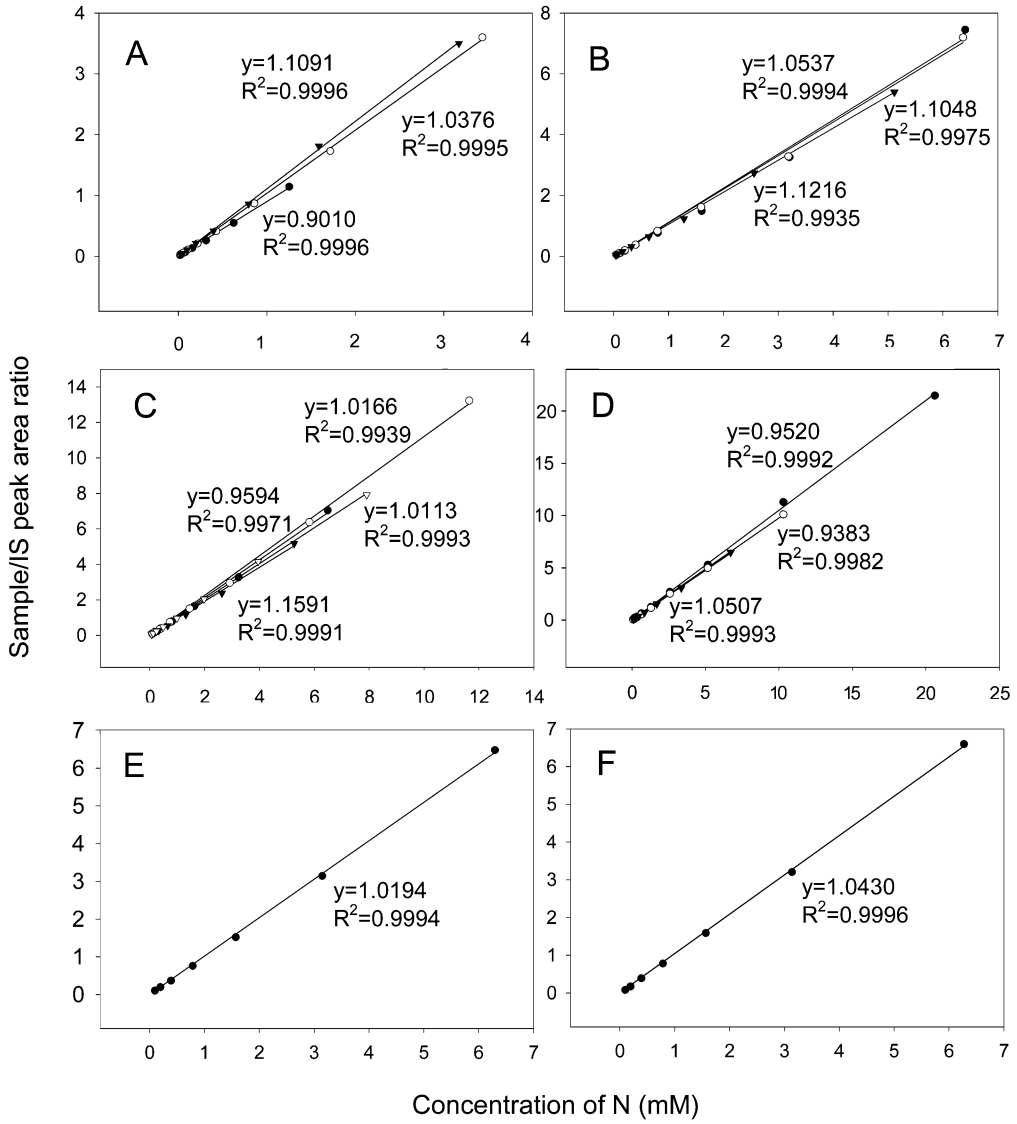


Figure 10. Concentration−response curves for compounds with different numbers of nitrogen atoms in the molecule: (A) colchicine (●), diphenhydramine (○), and doxepin (▼); (B) chlorpheniramine (●), diphenylalanine (○), and triprolidine (▼); (C) dibucaine (●), *N*-(1-2-thiazolyl)sulfanilamide (○), 4-(dimethylamino)antipyrine (▼), and 3,8-diamino-6-phenylphenanthridine (▽); (D) caffeine (●), nalpha-(9-fluorenylmethoxycarbonyl)-L-arginine (○), and nialamide (▼); (E) our internal reference compound HXZ002-A; and (F) our internal reference compound HXZ002-B. IS stands for internal standard. Curves were fitted with the equation $y = ax + b$ with the condition $b = 0$. The value for a is the slope and R^2 is the correlation coefficient for the fit.

This result demonstrated that our system could be used for quantitation over a wide range of compounds using a single external standard.

To get accurate quantities for each well, it is critical to ensure all samples are dissolved well and diluted precisely to make a daughter plate for CLND analysis, because samples in the master plate have to be diluted 2000 to 3000 times. Any error will be magnified several thousand times for quantity measurement for LC/CLND. We determined the absolute quantity of each compound in all validation and rehearsal libraries.

Gap between Relative Purity and Quantitative Purity. In previous sections, we discussed various methods to determine relative purity, quantitative purity, and the absolute quantity of compounds in combinatorial format, that is, a large number of compounds each in small amount and mostly located in 96-well plate. The relative purity, which is widely used to characterize combinatorial libraries, did not agree with the quantitative purity. Figure 11 shows the chromatograms of three compounds, **7**−**9**, and their relative and quantitative purity. Relative purity measured by UV_{214} for these three compounds appears to be 30−40% higher than their quantitative purity. We have seen this inconsistency in almost every compound when both relative and quantitative purities are measured. As shown in Figure 12, we analyzed 50 QC compounds based on 12 different scaffolds by both quantitative purity determination method (individual calibration) and HPLC/ UV_{214} relative purity method. These data show that the relative purity is in general 20−40% higher than the quantitative purity (also see Table 2). Because the quantitative purity is based on weight, therefore, these data suggest that about 20−40% of impurities by weight in each sample do not give sufficient UV_{214} response and are, therefore, undetected or underestimated. Without quantitative purity determination, the presence of these impurities is practically unknown. These "invisible" impurities (for ex-

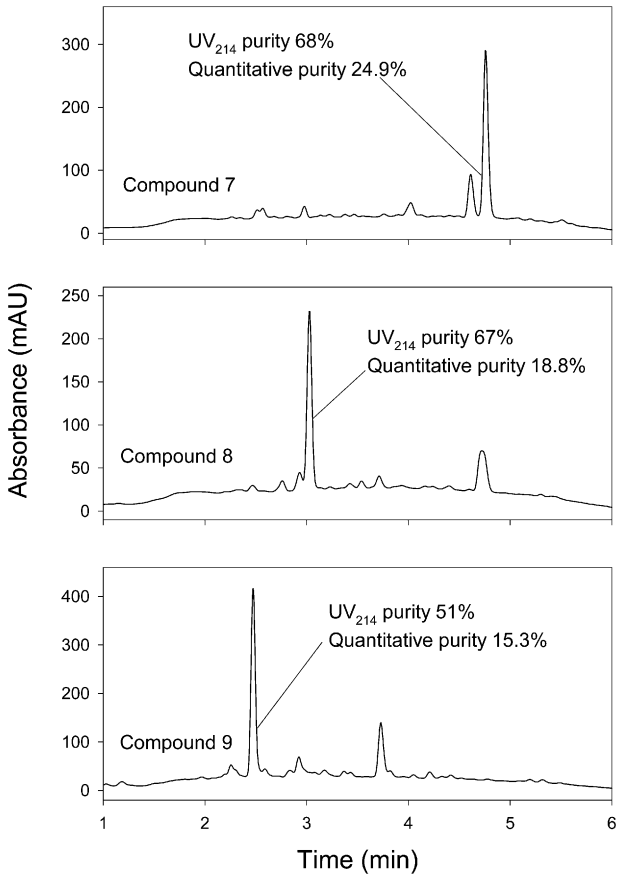


Figure 11. Chromatograms for compounds **7**−**9** using UV_{214} detection. Their relative purities are compared to quantitative purity.

ample, TFA) can either cause false positive or false negative responses in biological screenings or cause a wrong calculation of experimental concentration and, therefore, wrong assay results.

These "invisible" impurities can be removed by purification. We have used high-throughput reverse phase chromatography to purify all compounds at a throughput of 80 000/quarter. We determined the quantitative purity (LC/CLND/weighing) and relative purity (UV_{214}) for a randomly selected set of compounds (76 compounds) after purification (Figure 13). The average quantitative purity is 88.5%, and the average relative purity is 89.5%. The closeness between the quantitative purity and the relative purity has provided compelling evidence that purification is absolutely required for all compounds, even for synthesis product with a high relative purity (>90%).

Possible Sources for Invisible Impurities. Among numerous possibilities, we focused on the following sources of impurity: TFA, water, plastic extracts by organic solvents, high-boiling point solvent, inorganic salts, and resin extracts.

TFA. To reduce the log *P* value, incorporate hydrogen-bond acceptors, and render a druglike structure in the molecule, combinatorial library members often have basic nitrogen atoms that can bind acid molecules. In addition, acid is frequently used in the synthesis procedure. For example, in solid-phase reactions, TFA is often used to cleave synthesis product from resin. Assuming a 1:1 association with TFA (MW 115), the quantitative purity of a compound with a MW of 400 would be reduced by 22%, although the relative purity is not affected. TFA is very difficult to remove by standard evaporation and lyophilization procedures. It could be a contributor to the inconsistency between HPLC purity and the quantitative purity.

We investigated the TFA problem using two different approaches: (1) directly quantify TFA content in library compounds when TFA was used in the synthesis and (2) determine the remaining amount of TFA after adding TFA to samples and drying the samples by overnight lyophilization according to our protocols.

The amount of TFA in six compounds from a 1,2,5-trisubstituted benzimidazole testing library (Figure 7) was quantified by ^{19}F NMR. An external calibration curve made from α,α,α-trifluorotoluene was used in this analysis. The content of TFA in these samples ranged from 17 to 54%, with an average of 27.5% (Table 1). This example shows that the determination of quantitative purity is important for identifying potential purity problems in a library. If impurities are invisible by UV_{214}, ^{19}F NMR can determine the presence of TFA and quantify the absolute amount of TFA in a sample.

In the second experiment, compounds **8**, **10**, **13**, and **14** (Figures 7 and 15) were treated with 10% TFA in methanol and vacuum-dried and lyophilized for 24 h. The weights before and after this treatment were determined. The retained

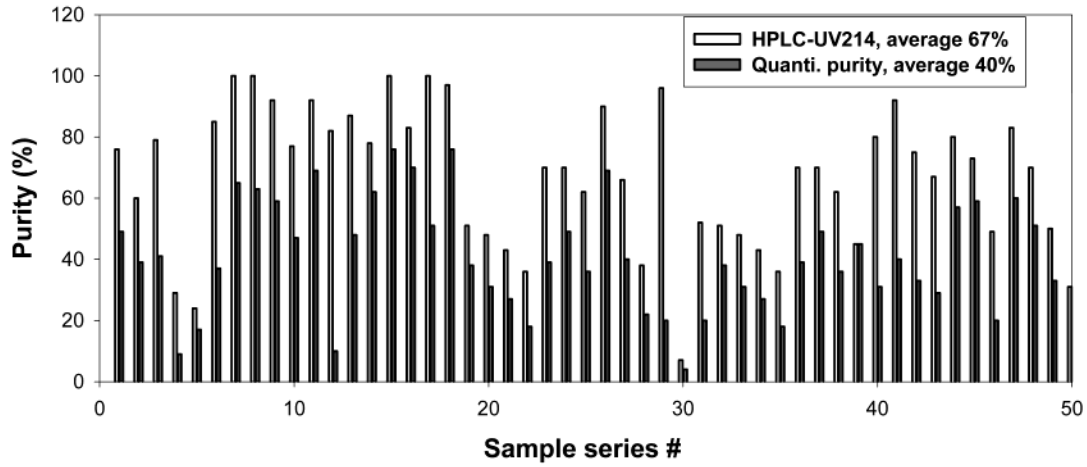


Figure 12. Comparison of the relative purity to quantitative purity for 50 unpurified compounds based on 12 different scaffolds.

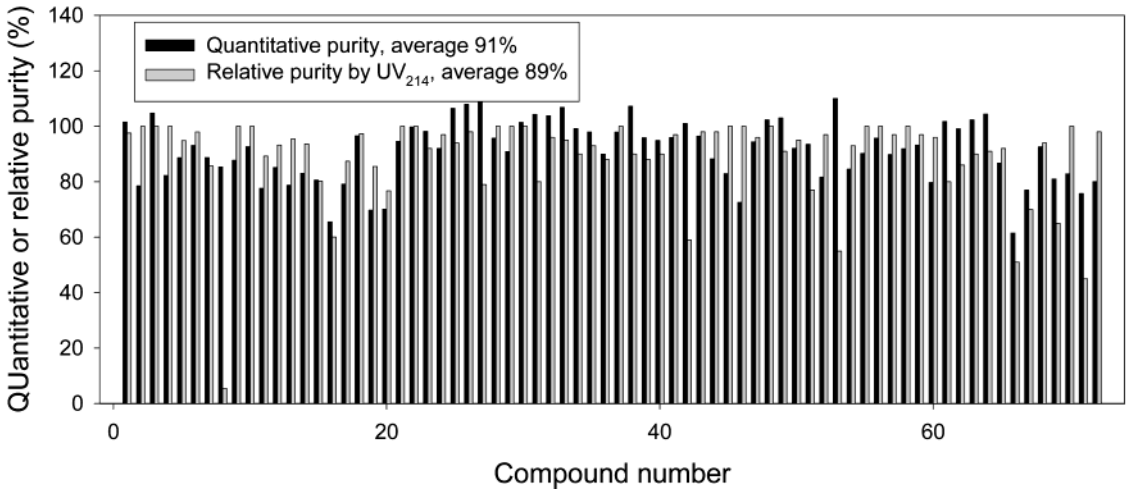


Figure 13. Comparison of the relative purity to quantitative purity for 76 purified compounds based on two different scaffolds.

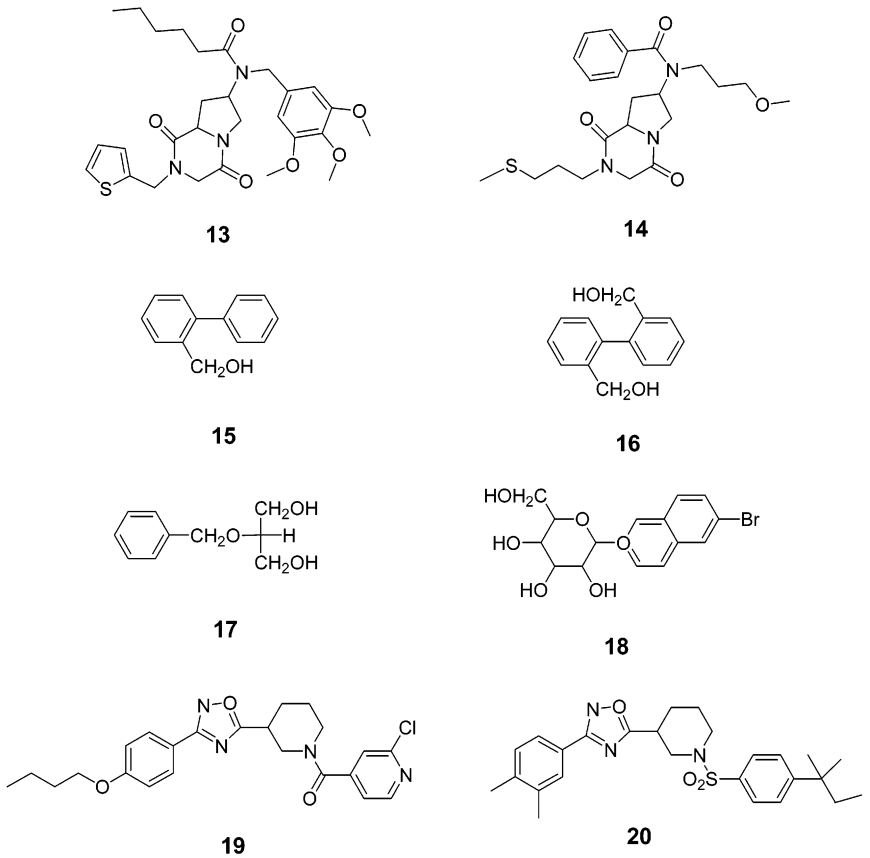


Figure 14. Chemical structures for compounds **13**−**20**.

TFA ranged from 17 to 60%, with an average of 42 wt %. This result indicates that whether compounds contain basic functional group or not, TFA will not be completely removed.

We also investigated the possibility of replacing TFA with an excess amount of acetic acid. Five-fold acetic acid was added to samples contaminated with TFA, and the samples were lyophilized for 12 h. After these treatments, an equivalent of TFA still remained in the sample, as determined by ^{19}F NMR and elemental analysis. However, acetic acid and formic acid can be removed by the same treatment, as confirmed by ^{1}H NMR.

Water. Water is a possible source for the extra weight, yet not detectable by UV_{214} or ELSD in HPLC or LC/MS analyses. Water may come from aqueous reaction steps, reversed-phase purification, or from routine sample handling and storage. Four compounds, **15**−**18** (Figure 14) were incubated with a methanol/water (1:1) solution. After drying and weighing, there is an average increase of 2.2% in weight. Therefore, moisture can be efficiently removed by the lyophilization treatment.

Plastic. Another "invisible" impurity is the plastic residues extracted from 96-well plates, pipet tips, or other plastic sources by organic solvent used in synthesis and transfer

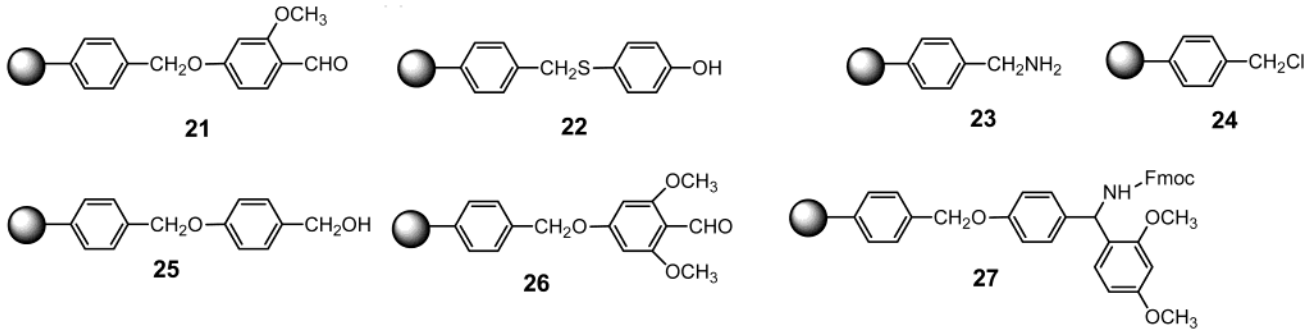


Figure 15. Chemical structures for resins **21**−**27**.

Table 3. Weight of Solvent Extracts from 96-Well Plates

solvent	Beckman (mg)	E&K Scientific (mg)	Whatman Unifilter (mg)
DCM	2.12	0.28	2.12
DCM + 10% TFA	2.25	0.5	1.69
THF	4.33	0.52	1.08
THF + 10% TFA	3.87	0.58	1.55
toluene	4.62	0.92	2.49
toluene + 10% TFA	3.75	1.36	3.85
chloroform	4.25	0.56	2.65
chloroform + 10% TFA	4.59	0.68	2.88

processes. We specifically investigated the 96-well plate as a source for plastic contaminants. Pure compounds **19** and **20** (>99.6%) were weighed (4.5 and 5 mg) and quantitatively transferred into a Beckman 96-well plate, and the plate was shaken for 14 h in chloroform. Compound solutions were dried and the obtained weights were 6.6 and 7.6 mg. The quantitative purity was determined to be 75 and 66%, which were much lower than the expected 99.6%. To study the tolerance of plastic plates to normal organic solvents, we treated three different plates from different manufacturers with DCM, THF, toluene, and chloroform. Because TFA and amine are often used in resin cleavage and reactions, we also tested the effect of adding TFA and *n*-butylamine to the above solvents. Solvents were added to wells in a 96-well plate. The plate was covered with aluminum foil and kept for 24 h, a frequently used time period for combinatorial synthesis. All extracted materials were transferred to a preweighed vial for drying and weighing. After drying, white powder was clearly visible in every vial. The weights of the extracted materials are listed in Table 3.

The order of aggressiveness of solvents is chloroform ≈ toluene > THF > DCM. The addition of TFA and *n*-butylamine (not shown) did not significantly increase the extraction of materials from plate. To avoid generating such "invisible" impurities, one should select solvent-resistant plates and avoid using plastic-erosive solvents.

Solvent. Solvent with a high boiling point may be hard to remove completely by evaporation and lyophilization processes. We tested DMA, DMF, NMP, dioxane, and DMSO (bp's: 166, 153, 93, 101, and 189 °C, respectively). The average of solvent residue after vacuum-drying and lyophilization was 0.75 mg or 3% of the relative weight of the expected final product (25 mg). The contribution of a high boiling point solvent to "invisible" impurities is minor under our sample treatment conditions.

Table 4. Weight of Resin Wash-Out

resin	DCM (mg)	DCM + 10%TFA (mg)
21	0.76	4.94
22	0.41	6.09
23	0.62	4.76
24	1.16	1.26
25	0.56	1.15
26	1.01	5.7
27	1.15	15.76

Inorganic Salt and Catalyst. A synthesis reaction was carried out in the presence of KF, Pd, and TFA. LC/MS analysis showed that the purity of the product was 89% by UV_{214} detection. To reveal the "invisible" impurities, the fluorine and Pd content were quantified by elemental analysis. The F percent was 17%, and the Pd percent was 0.34%. Although the Pd is negligible, the presence of F is significant. In addition to TFA, the KF salt may also be present in the compound on the basis of the high percentage of F in this sample and the low percentage of TFA used in the reaction.

We often discovered the presence of inorganic salts in the sample by analyzing materials eluted in the solvent front during HPLC analysis and purification of combinatorial library compounds. This is a reaction-dependent problem and should be considered when necessary.

Impurities from Resin. Resins from various suppliers often contain impurities.[11] These impurities may constitute a source for "invisible" impurities in the final library compounds in solid-phase synthesis. To investigate this possibility, seven commonly used resins, **21**−**27** (Figure 15), were treated with DCM and DCM + 10% TFA. We used 100 mg of resin for each study because this is the amount we often use for library synthesis with the expectation of ~25 mg product (~MW 500) in the end. For all resins, the larger amount of impurities was washed out when TFA was used with the organic solvent. The extracted materials ranged from 1 to 15 mg (Table 4), with an average of 5.7 mg, which is 23% of the expected amount of synthetic product. Therefore, extracts from resin can be a significant source for impurities from solid-phase synthesis compounds. To avoid impurities from resin, reaction resins should be washed with solvent and acid before use.

In summary, one or more of above impurities may exist in the final synthetic product. The existence of "invisible" impurities will cause the inconsistency between the relative purity and the quantitative purity. The amount of these

"invisible" impurities can be determined by measuring quantitative purity of these compounds. To remove "invisible impurities", all compounds, including those having a synthesis purity of >90%, should be purified by high-throughput purification methods.

By determining the quantitative purity at an early phase of synthesis development, "invisible impurity"-related quality issues can be identified early on, and synthesis and treatment procedures can be modified to remove these impurities. During the development of a synthesis protocol for a 1,2,5-trisubstituted benzimidazole library, we measured the quantitative purity for six compounds (**7**−**12**) and found that although the relative purity by UV_{214} was 50−80%, the quantitative purity was <20%. Quantitative ^{19}F NMR was used to assay the amount of TFA in the sample. Experimental results (Table 1) show that TFA contributes to more than 20% of the impurities. TFA was used in the intermediate synthesis steps. To remove residue TFA, more thorough base washing steps were incorporated. The quantitative purity of the same six QC compounds in a new rehearsal library and the production library was measured. The average quantitative purity before purification was increased to 48 and 56%, respectively (Table 1), and the relative purity (UV_{214}) was over 80%.

Conclusion

A successful combinatorial synthesis should be examined by the characterization of identity, purity, and quantity (yield) of all library members. The identity can be determined by MS detection, preferably accurate mass analysis. The quantity can be determined by using techniques such as LC/MS/CLND or LC/MS/ELSD. Purity measured by UV_{214} or ELSD is only relative purity and does not represent the real composition of the sample. "Invisible" impurities, such as TFA, plastic extract, inorganic salts, and resin extracts, can add to the sample weight, yet not be detectable by UV_{214} or ELSD detectors. The quantitative (absolute) purity can be determined by an individual calibration method (low throughput), qNMR (medium throughput), and LC/MS/CLND (high-throughput) with the high-throughput weighing methods.

The yield and quantitative purity are not only the ultimate quality measure of a combinatorial library, they also determine the success or failure of the high-throughput purification process. Low-purity compounds (<20%) and compounds with a high purity but low quantity rarely give enough compound after purification. Yield and purity can be improved early on in the combinatorial synthesis optimization process. Validated chemistry, proper selection of building blocks (reagents), crucial analytical measurements, and high-throughput purification will help produce compounds with a high yield and purity.

Acknowledgment. We gratefully acknowledge the previous and current members of our team: Doug Livingston, Jiang Zhao, Jianmin Pan, Jason Cournoyer, Duayne Tokushige, Melody Wen, Audie Cabero, Michael Demee, Teresa Sierra, Kyle Leopold, Birong Zhang, Thutam Hopkins, Jeffrey Dener, Cuong Ly, Pascal Fantauzzi, Mark Gleason, Tim Lease, Elaine Krueger, Dean Dragoli, Eric Zylstra, Hyo Lee, John Munger, Barry Bunin, Todd Sommers, Guy Breitenbucher, Matt Plunkett, and Kumar Gadamasetti for their contributions and enthusiasm to build and improve our quality systems for combinatorial synthesis and analysis.

References and Notes

(1) (a) Gordon, E. M.; Barrett, R. W.; Dower, W. J.; Fodor, S. P.; Gallop, M. A. *J. Med. Chem.* **1994**, *37*, 1385. (b) Thompson, L. A.; Ellman, J. A. *Chem. Rev.* **1996**, *96*, 555. (c) DeWitt, D. H.; Czarnik, A. W. *Acc. Chem. Res.* **1996**, *29*, 114. (d) Balkenhohl, F.; Bussche-Hunnefeld, C.; Lansky, A.; Zechel, C. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 2288. (e) Lam, K. S.; Lebl, M.; Krchnak, V. *Chem. Rev.* **1997**, *97*, 411.

(2) (a) Leznoff, C. C. *Chem. Soc. Rev.* **1974**, *3*, 65. (b) James, I. W. *Mol. Diversity* **1996**, 2, 175. (c) Hermkens, P. H. H.; Ottenheijm, H. C. J.; Rees, D. *Tetrahedron* **1996**, 52, 4527. (d) Brown, R. *Contemp. Org. Synth.* **1997**, *4*, 216. (e) Czarnik, A. W.; DeWitt, S. H. *A Practical Guide to Combinatorial Chemistry*; American Chemical Society: Washington, DC, 1997.

(3) (a) Keifer, P. A. *Drug Discovery Today*, **1997**, *2*, 468. (b) Keifer, P. A. *Drugs Future* **1998**, *23*, 301.

(4) (a) Weller, H. N. *Mol. Diversity* **1999**, *4*, 47. (b) Kyranos, J. N.; Cai, H.; Zhang, B.; Goetzinger, W. K. *Curr. Opin. Drug Discovery Dev.* **2001**, *4*, 719. (c) Edwards, C.; Hunter, D. J. *J. Comb. Chem.* **2002**, in press. (d) Isbell, J.; Xu, R.; Cai, Z.; Kassel, D. B. *J. Comb. Chem.* **2002**, *4*, 600.

(5) Fang, L.; Demee, M.; Cournoyer, J.; Sierra, T.; Young, C.; Yan, B. *Rapid Commun. Mass Spectrom.* **2003**, in press.

(6) (a) Hsu, B. H.; Orton, E., Tang, S.Y.; Carlton, R.A. *J. Chromatogr., B* **1999**, *725*, 103. (b) Kibbey, C.E.; *Mol. Diversity* **1995**, *1*, 247. (c) Fang, L.; Wan, M.; Pennacchio, M.; Pan, J. *J. Comb. Chem.* **2000**, *2*, 254.

(7) Fang, L.; Pan, J.; Yan, B. *Bioeng. Biotechnol. Comb. Chem.* **2001**, *71*, 162.

(8) Fang, L.; Demee, M.; Sierra, T.; Zhao, J.; Tokushige, D.; Yan, B. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 1440.

(9) (a) Rabenstein, D. L.; Keire, D. A. Marcel Dekker: New York, 1991, pp 323−369. (b) Maniara, G.; Rajamoorthi, K.; Rajan, S.; Stockton, G. W. *Anal. Chem.* **1998**, *70*, 4921. (c) Godejohann, M.; Preiss, A. *Anal. Chem.* **1998**, *70*, 590. (d) Drew, M.; Orton, E.; Krolokowski, P.; Salvino, J.; Kumar, V. *J. Comb. Chem.* **2000**, *2*, 8. (e) Ellis, D. A.; Martin, J. W.; Muir, D. C. G.; Mabury, S. A. *Anal. Chem.* **2000**, *72*, 726. (f) Gerritz, S.; Sefler A. *J. Comb. Chem.* **2000**, *2*, 39. (g) Pinciroli, V.; Biancardi, R.; Colombo, N.; Colombo, M.; Rizzo, V. *J. Comb. Chem.* **2001**, *3*, 434. (h) Silvestre, V.; Goupry, S.; Trierweiler, M.; Robins, R.; Akoka, S. *Anal. Chem.* **2001**, *73*, 1862. (i) Hany, R.; Rentsch, D.; Dhanapal, B.; Obrecht, D. *J. Comb. Chem.* **2001**, *3*, 85.

(10) (a) Fitch, W. L.; Szardenings, A. K.; Fujinari, E. M. *Tetrahedron Lett.* **1997**, *38*, 1689. (b) Fujinari, E. M.; Manes, J. D.; Bizanek, R. *J. Chromatogr., A.* **1996**, *743*, 85. (c) Fujinari, E. M.; Manes, J. D. *J. Chromatogr., A*, **1994**, *676*, 113. (d) Taylor, E. W.; Qian, M. G.; Dollinger, G. D. *Anal. Chem.* **1998**, *70*, 3339. (e) Yurek, D. A.; Branch, D. L.; Kuo, M. *J. Comb. Chem.* **2002**, *4*, 138.

(11) MacDonald, A. A.; DeWitt, S. H.; Ghosh, S.; Hogan, E. M.; Kieras, L.; Czarnik, A. W.; Ramage, R. *Mol. Diversity* **1995**, *1*, 183.

CC030008F