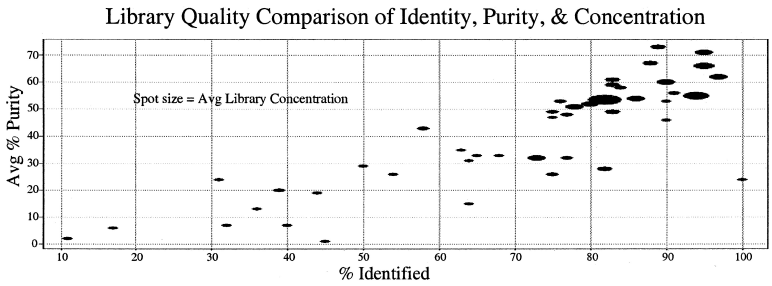
JOURNAL OF combinatorial CHEMISTRY

Article

Development of a System To Evaluate Compound Identity, Purity, and Concentration in a Single Experiment and Its Application in Quality Assessment of Combinatorial Libraries and Screening Hits

David A. Yurek, Derek L. Branch, and Ming-Shang Kuo

J. Comb. Chem., **2002**, 4 (2), 138-148• DOI: 10.1021/cc010028m • Publication Date (Web): 29 January 2002



More About This Article

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

- Supporting Information
- Links to the 10 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

Development of a System To Evaluate Compound Identity, Purity, and Concentration in a Single Experiment and Its Application in Quality Assessment of Combinatorial Libraries and Screening Hits

David A. Yurek, Derek L. Branch, and Ming-Shang Kuo*

Discovery Technologies, Pharmacia Corporation, Kalamazoo, Michigan 49001

Received May 31, 2001

The development and use of a new assay system for the simultaneous determination of identity, purity, and concentration of sample components from combinatorial libraries produced by parallel synthesis are described. The system makes use of high-performance liquid chromatography with UV/vis photodiode array (PDA), evaporative light scattering (ELSD), chemiluminescent nitrogen (CLND), and time-of-flight mass spectrometer (TOFMS) detectors (HPLC−PDA−ELSD−CLND−TOFMS). Although these detectors have previously been utilized separately for the analysis of combinatorial chemistry libraries, the use of TOFMS along with CLND provides a synergistic combination enabling target and side-product structures to be identified and their concentrations and purities determined in a single experiment from a solution containing microgram levels of material. The CLND was found to give a linear response based on the number of moles of nitrogen present. Therefore, if the number of nitrogens per molecule is known, the concentration of each nitrogen-containing sample component may be determined utilizing an unrelated co-injected standard. A molecular formula for an impurity may often be calculated from the exact mass determined by the TOFMS and knowledge of the chemistry involved. Thus, if the sample components contain nitrogen, the concentration of every identified HPLC peak may be determined even in the absence of primary standards. This combination of detectors enabled the characterization of both target compounds and byproducts in combinatorial libraries, allowing the optimization of library synthetic procedures. This system was also used to survey the quality of libraries, enabling the selection of the best libraries for screening. This method also facilitated the characterization of samples from combinatorial libraries found as hits in high-throughput screening to establish the potency of the leads based on their actual concentration. In addition, concentrations and potencies of impurities were determined after identification of their structures, utilizing exact mass data, determination of charge states, and knowledge of the synthetic chemistry.

Introduction

In the transition of screening labs from the utilization of traditional compound libraries containing well-characterized compounds to combinatorial libraries produced by parallel synthesis, it is often mistakenly assumed that the latter compounds are structurally correct, pure, and uniformly present at a specified concentration. However, as a direct result of the use of a single set of synthetic conditions across a wide range of substituents, target compounds are not produced in the same yield and purity across the library. Two basic philosophies exist in dealing with this fact: (1) purify, dry, and weigh all target library compounds produced prior to screening; (2) screen unpurified libraries using various procedures to characterize the active leads. Both these philosophies have merit, but both involve time-consuming processes. The up-front purification approach requires that each of the thousands of library compounds be purified, dried, weighed, and dispensed to provide screening samples of uniform purity and concentration. The greatest advantage of this approach is that screening hits are directly comparable for potency and derivation of structure−activity relationships (SAR). The purified compounds may be used for multiple screens if sufficient compound is obtained. However, since the majority of screens have a hit rate of <0.1%, most of these compounds will never be discovered as a hit, making the amount of time and effort required to purify all these samples hard to justify. The second approach of screening unpurified samples requires that the active component(s) must be identified and the individual potency established for each screening hit in a rapid fashion. Since the compounds do not have the same purity and concentration, the screening data are not directly comparable across the library. Target compounds may not be assumed to be the active component for identifying possible templates without further characterization. Purifying compound libraries prior to screening or assessing the parameters of identity, purity, and concentration after hits are identified can potentially add weeks to months of additional work to an increasingly shortening timeline. As described here, a rapid method developed to obtain identity, purity, and concentration information on combinatorially produced compounds has been implemented to shorten this process.[1,2] Used prior to screening or purification on a statistically relevant number of random samples, this method allows focusing of effort on higher quality libraries. Obtaining this information after screening unpurified combinatorial chemistry libraries enables prioritization of the screening leads based on purity and concentration. When non-target components are active, the method provides

10.1021/cc010028m CCC: $22.00
Published on Web 01/29/2002

structural and concentration information to assist in their identification and characterization as new leads and to enable better design in the synthesis of the next round of lead candidates.

Historically, mass spectrometry, utilizing a neat injection of the product mix into the instrument (flow injection)[3] or analysis of the effluent from high-performance liquid chromatography (HPLC), is routinely employed to identify the synthetic products of combinatorial chemistry. Relative purity is often estimated by HPLC with either UV or evaporative light-scattering detection (ELSD), and concentration estimates are calculated using the assumption that the detector has observed all sample components with a similar response, or from a response curve generated using primary standards. ELSD and especially UV detectors do not provide a uniform response to different compound classes, making the determinations of purity, yield, or concentration only rough estimates. This estimate is further compromised when the sample contains water, solvents, and other materials undetectable by the detectors because a relative response is used. Estimation of library compound concentration using ELSD has recently been presented;[4] however, although better than UV, this technique still suffers from a lack of consistent response to different compound classes. Generation of primary standards to estimate the purity of a combinatorially produced compound requires a statistically relevant number to estimate the yield for each library. Because of the difficulties associated with production of this number of primary standards, often too few are utilized, making the calculation of average yield uncertain.[4] Recently, a chemiluminescent nitrogen detector (CLND)[5] became available that provides a linear response to the absolute amount of nitrogen from most nitrogen-containing analytes. Except for a few rare types of structures (such as azides, tetrazoles, and others that yield N_2 on combustion[6]), the signal produced in the CLND is proportional to the number of moles of nitrogen present. This detector has been shown to allow rapid determination of pure compound concentrations using flow injection[7] and has been utilized with GC, LC, and supercritical fluid chromatography.[8] The sample concentration is determined by comparison to the response of a well-characterized sample as an internal or external standard. In addition, when the measured concentration is compared to that expected from sample preparation, this method accounts for the presence of other materials unobserved by the detectors but still contained in the sample (water or solvent due to incomplete drying, hydrates, solvates, SLE material, etc.), giving an actual concentration rather than a relative purity. Other techniques attempting to address the issues of identity, purity, and concentration in an absolute rather than a relative fashion include DI- (direct injection), FIA- (flow injection analysis), and LC−NMR methods.[9] Although these suffer from automated data processing and interpretation issues, progress is rapid in this field and should provide data complementary to hyphenated LC techniques.

The CLND therefore functions as a nearly universal detector for nitrogen-containing compounds and enables the concentration of a sample component to be determined if the number of nitrogens per molecule is known, without the requirement for primary standards. For the target compounds of a combinatorial parallel synthesis library, the number of nitrogens per molecule is known. However, unless the library is purified or contains only single compounds, many other components are often present because of incomplete or side reactions. Without the identification of the molecular formula, no concentration can be derived for these impurities. The use of a time-of-flight mass spectrometer (TOFMS) enables the charge states and exact masses (within 5−10 ppm) to be determined for all ionizable components of a sample. By use of the exact masses and knowledge of the synthetic chemistry, a molecular formula (and the number of nitrogens per molecule) may often be determined for impurities.[3] Although unit mass determination may often be adequate to identify impurity structures when the chemistry is well understood, the superior mass resolution of the TOF gives higher assurance of the correct molecular formula, provides a set of possible empirical formulas when the chemistry goes astray, and discriminates multiply charged ions. Thus, the concentration of each nitrogen-containing component resolved by the HPLC system and detected by the CLND and MS detector may be determined. Therefore, the described system provides a synergistic combination enabling target and impurity structures to be identified and their concentrations and purities determined in a single experiment from a solution containing microgram levels of material without the need for primary standards. If the compound does not contain nitrogen, the relative purity and concentration may still be estimated from the UV or ELSD data. However, since most compounds of pharmaceutical interest contain nitrogen, the described system should be applicable to the majority of analytes. In this report, the development and application of this system are described for leads generated from combinatorial compound libraries.

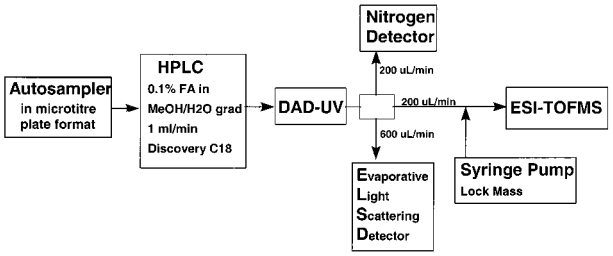


Figure 1. Flow diagram of instruments.

Results and Discussion

Assay Development. A schematic of the overall setup is depicted in Figure 1. Development of the assay required optimization of the individual detectors as well as the HPLC conditions. To accommodate the CLND, a new HPLC system had to be deployed and used exclusively with non-nitrogen-containing solvents and buffers.[10] Caffeine was used as an internal standard for all detectors and as a calibrant for the CLND response. To split the flow to the various detectors, a custom-built LC Packings flow splitter was purchased.[11] This unit maintains the same split ratios throughout the gradient run, allowing peaks in one detector to readily be correlated with a peak in another detector, a crucial requirement for proper interpretation of the data. Figure 2 shows

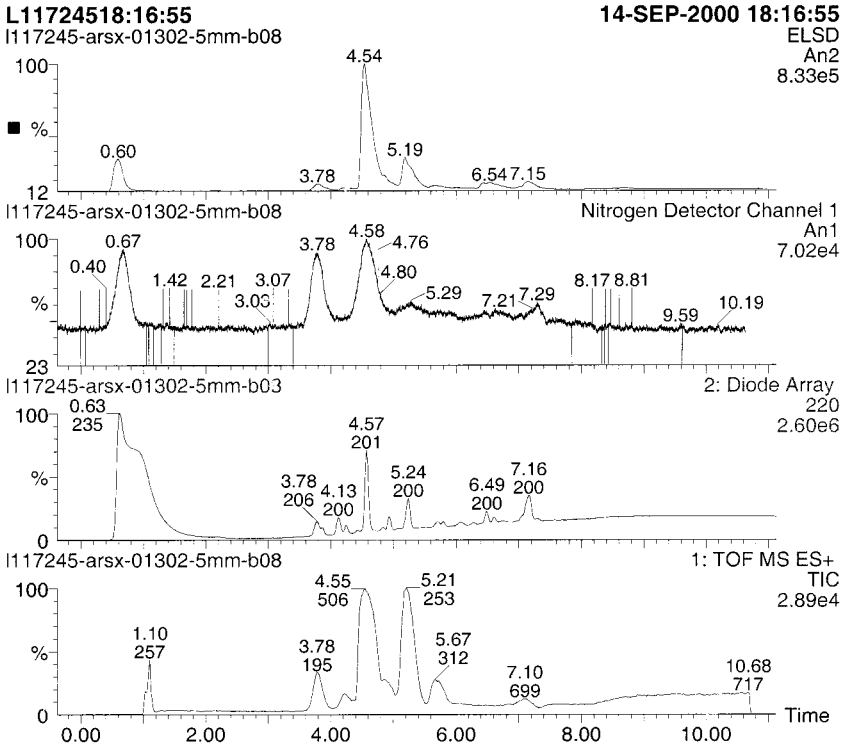


Figure 2. Typical chromatogram traces for a single sample.

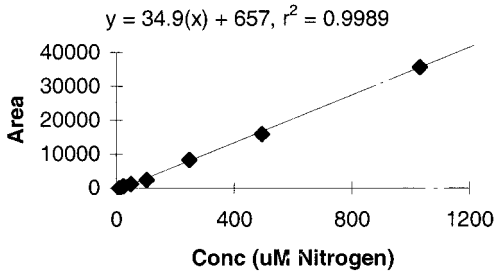


Figure 3. CLND response curve for caffeine.

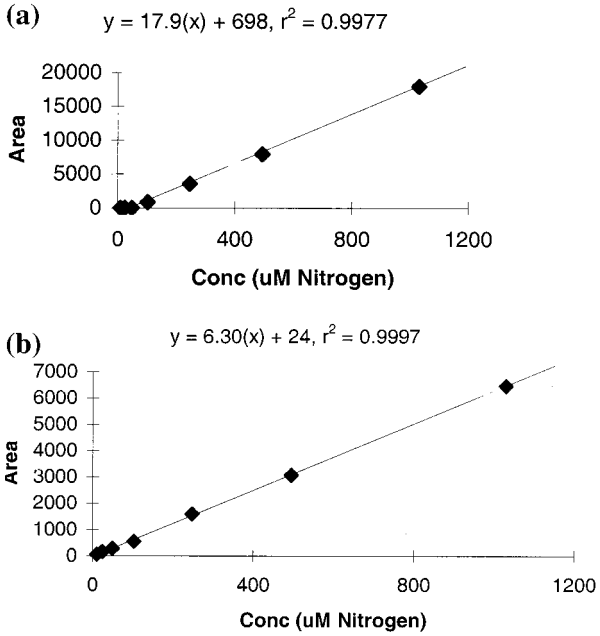


Figure 4. Caffeine response curves for UV at (a) 220 and (b) 254 nm.

typical chromatograms for a real combinatorial sample. A comparison of the peaks in the traces for each signal shows the variation in response for these detectors. Resolution and sensitivity losses are apparent when the UV trace is compared with the post-splitter detectors. The CLND trace reflects the low concentration of the sample. However, the MCP detector used in the TOFMS is often saturated by the eluting components. Therefore, the automated calculation of the exact mass is set to only utilize signal below 1000 cps. To adjust the TOFMS calibration, erythromycin was used as a lock mass compound and added via a T-fitting placed between the splitter and the TOFMS. Some degradation of erythromycin over time was seen due to loss of water, but this could be accommodated by using the mass for the dehydrated product as the lock mass.

Sensitivity and Linearity. Responses of the various detectors have already been published (vide supra), but to facilitate comparison between them, caffeine samples were used to generate standard curves. For the CLND (Figure 3), levels down to 25 μM nitrogen (10 ng injected onto the column) could be detected. The linear range of the instrument extended to about 1000 μM nitrogen. As expected, the UV data gave good linear relationships between area and concentration (Figure 4). The ELSD response (Figure 5) was linear when plotted in a log/log fashion. However, the ELSD was found to have differing responses to various compound classes (data not shown),[4] consistent with reported ELSD experimental results.[12−16] Caffeine response curves for the TOFMS were generated for the total ion chromatogram and for the extracted ion trace for 195 ± 0.5 amu (Figure 6). As expected, these have a linear region, but amounts above this region saturate the detector.

Assay Validation. Samples of individually synthesized and purified compounds taken from a combinatorial library

Table 1. Assay Validation Data

sample PNU no.	no. of nitrogens	CLND area of caffeine (258 uM, 4 N)	CLND area of compound	calcd concn (uM)	% error (vs 250 μM)	amount injected (μg)	ELSD area
184 346	6	28 886	46 136	275	9.9	4.1	164 226
184 445	6	23 391	40 072	295	17.9	3.6	122 178
241 977	6	20 446	31 304	263	5.3	2.9	75 471
242 264	6	22 422	32 199	247	1.2	3.8	127 354
244 263	5	25 088	29 010	239	4.5	2.8	74 459

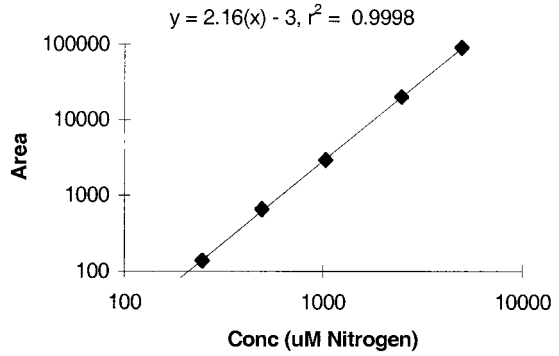


Figure 5. Caffeine response curve for ELSD.

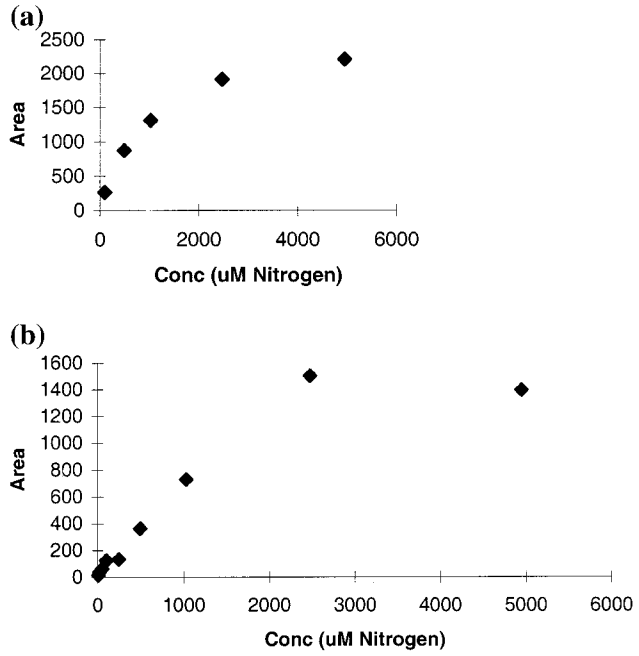


Figure 6. Caffeine response curves for (a) TOFMS total ion chromatogram (TIC) and (b) 195 amu extracted ion chromatogram (XIC).

were used for validation purposes. Solutions of these compounds were prepared at 250 μM in the internal standard (caffeine) solution and injected onto the system. Agreement of the concentration derived from the CLND data with the theoretical concentration was excellent, as shown in Table 1. The correlation between amounts injected (in micrograms) and ELSD area also was good ($r^2 = 0.986$).

Data Analysis. Results were analyzed using the Micromass OpenLynx Diversity software. This program has many variables to optimize for the analysis and the reporting of data from these experiments, which are contained in *.olp and *.ors files, respectively. Initially, the *.olp program translates the data to a text file (*.rpt) containing all of the raw information from the run. Items of interest are then specified in the *.ors file and reports generated. However, even with the flexibility of the vendor package, a report containing all of the relevant information for our analysis could not be obtained. Therefore, a Visual Basic program was written in-house to extract the exact mass of the expected ion and any significant impurities, and the area and area percent for these peaks in the UV220, ELSD, and CLND traces. The data are downloaded to an Excel template for further processing and generation of summary tables. An example of a summary table (Table 2) shows average values for the percent of compounds found, UV and ELSD purity, percent of the expected concentration, percent of library compounds in each purity quartile, and a partial listing of individual sample data highlighting the target compound and the two largest impurities and their exact masses.

The use of the TOFMS allowed the identity of expected compounds to be confirmed in a more rigorous fashion than with a quadrupole instrument and facilitated the identification of any impurities found through the generation of possible molecular formulas along with consideration of the chemistry involved. The CLND also assisted the identification of unknown peaks by indicating the presence or absence of nitrogen.

Calculation of Concentration Using CLND. Caffeine was chosen as an internal standard because it is a well-characterized, easily soluble, commercially available compound that elutes early in the HPLC run. Any similar nitrogen-containing compound could be utilized. The compound gives a response in all the detectors, facilitating alignment of the chromatograms. In addition, its response in the ELSD is nearly negligible when routine concentrations of library compounds are assayed, enabling the integration of the entire chromatogram for calculation of purity. The concentrations of sample components were calculated from the CLND areas using the following equation:

$$\text{ConcSample} = \frac{(\text{SampleArea})(\text{IntStdConc})(\text{IntStd\#N})}{(\text{IntStdArea})(\text{Sample\#N})}$$

Exact masses of peaks other than the target compound were used to generate molecular formulas using the software. After consideration of the chemistry involved in the synthesis of the samples, this allowed the structures to be solved and the number of nitrogens per molecule to be determined for the significant impurities. After the number of nitrogens for an impurity was determined, the concentration was then calculated using the above equation. An automated check for compounds co-eluting with the caffeine internal standard was accomplished by comparing the CLND area for the indi-

Table 2. Summary Table for Library Analysis

Combi-Library Analysis																			
Library Name																			
Identity % Found by MS		97																	
					Quality by Quartiles														
Quality Parameters	Avg	Std Dev	<5%	5-24%	25-49%	50-74%	75-100%	>100 %				Total ELSD Area	2.8E+06						
Purity by ELSD Area %	62	29	5	8	20	17	50					Avg ELSD Area	31391						
Purity by UV220 Area %	68	25	23	44	31	1	1	0				Std Dev	41926						
% of Expt'd Conc by CLND	18	16																	
	Target				Impurity 1								Impurity 2						
Compd #	Target Found	% Purity ELSD	% Purity UV220	% Conc Exptd	Ret. Time	% Purity ELSD	Exact Mass 1	%BPI Mass 1	Exact Mass 2	%BPI Mass 2	Exact Mass 3	%BPI Mass 3	Ret. Time	% Purity ELSD	Exact Mass 1	%BPI Mass 1	Exact Mass 2	%BPI Mass 2	Exact Mass 3	%BPI Mass 3
1	5.4	58	35	11	6.39	22	310.145	100	171.0849	63	309.1647	45								
2	6.94	94	59	18																
3	5.62	31	29	15	6.14	56	431.17[illegible]	100	296.1352	27	157.0742	16								
4	5.96	93	66	24																
5	5.39	45	49	9	5.08	22	371.174	38	416.195	18	438.1744	100	6.44	11	422.1723	14	461.1851	100	734.469	30
6	5.34	30	33	14	5.58	65	564.2044	100	525.1959	75	476.1519	69	6.27	5	735.[illegible]3	100	791.2264	90	769.2466	58
7	4.05	9	35	3																
8	5.73	0	54	1	4.39	7	317.123	100	250.1229	36	317.3313	24	8.4	[illegible]	413.2656	100	735.4644	99	261.1363	63
9	NF	0	0	0	4.37	20	317.1223	100	250.1234	27	317.3327	15	8.64	8	735.4617	100	413.2653	73	261.1343	71
10	6.25	29	74	2																
11	NF	0	0	0																
12	5.71	65	35	9	8.63	1	735.462	100	261.1369	95	734.8972	78								
13	7.06	80	85	7																
14	6.36	16	63	1																
15	7.12	40	84	1	4.42	10	317.1214	100	250.1274	34	317.3605	13								
16	4.89	76	47	16																
17	5.34	90	99	48																
18	5.9	82	78	21																
19	5.22	58	100	8	6.3	1	735.4611	100	516.2004	63	217.1173	51								
20	4.73	56	82	12	8.7	3	261.1371	83	735.4684	78	217.1079	70								
21	5.36	21	46	2																
22	4.1	61	57	17	5.36	6	735.4661	100	734.831	64	621.234	40								
23	NF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	4.49	35	43	12	4.81	34	528.2036	100	506.2314	54	735.4591	38								
25	6.25	77	54	8																
26	4.4	19	86	5	5.79	10	735.4643	100	601.2476	40	734.9562	37								
27	6.48	92	59	25																
28	6.01	88	92	35																
29	5.48	61	90	16	6.98	24	490.2634	100	512.2609	99	425.2089	69								
30	5.75	28	58	3	6.76	28	457.2025	100	735.4746	29	734.8147	27								

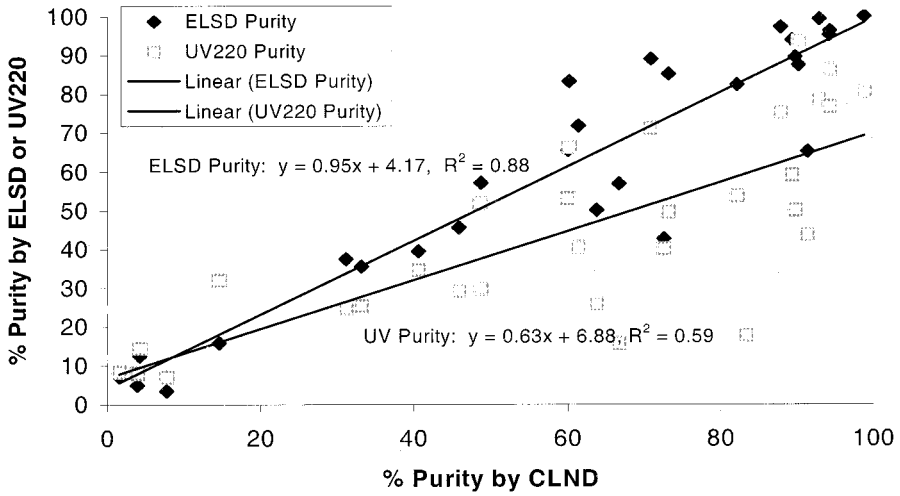


Figure 7. Comparison of expected compound purities calculated from ELSD or UV at 220 nm with CLND.

vidual run with the average area seen for the entire plate and/or control samples.

Evaluation of Purity Using CLND, ELSD, and UV at 220 nm. Comparison of purity determinations using CLND, ELSD, and UV are possible using the data available from these samples. In Figure 7, a plot of percent CLND purity (determined using concentrations calculated from the CLND area for all significant sample components, taking into account their molecular formulas) vs relative purity (derived from ELSD or UV (220 nm) area percents) is shown for a set of 35 library samples. The use of relative ELSD area percent to determine purity correlates well with the absolute CLND-derived purity, making the use of this detector beneficial[4] for purity determinations of these similar combinatorial library samples ($r = 0.939$). Conversely, poor correlation is seen between the area percent for UV at 220 nm and the CLND-derived purity. However, because of superior sensitivity and resolution, the UV trace often indicates the presence of peaks that are not observed by the other detectors, resulting in lower relative purity results from this detector.

Statistics. The analysis of large sample sets with statistically valid sampling allows valuable information to be obtained without testing the entire population. For a normal distribution, the accuracy of the calculated average improves with a larger number of random samples taken from any population, but it plateaus after sufficient samples are tested. This observation is the basis of the *t* distribution, which takes into account the standard deviation of a sample set, the degrees of freedom, and the desired confidence level to predict the variance from the true average value. The typical confidence level recommended for analysis is 95%. The variance from the true average should be minimized, typically within 10% (i.e., a total range of 20%). For this work, adequate numbers of samples must be assayed to determine the standard deviation of the average purity and average percent of expected concentration values. The actual number of samples to be assessed is dependent on the standard deviation of the averages and the confidence limits desired for the final average values. However, even with a maximum standard deviation of 0.5, the calculated variation at the 95% confidence limit changes only slightly after a sufficient number of samples have been tested. Therefore, testing one plate of randomized samples (88) for each library is sufficient to generate averages within a variation of 10%. Therefore, the calculated averages are expected to be within a range of 20% of the actual average for the library 95% of the time.

Assay Difficulties. Several types of problems associated with these libraries were encountered. Sample solubility must be taken into account because the system can only assess the quality of what has been injected onto the HPLC column. Precipitation of sample components is difficult to observe because of the low sample concentrations used, the opaque, covered injection plates, and the number of samples. Often, samples are stored at low temperature prior to injection, further providing opportunity for component precipitation. Therefore, a judicious choice of injection solvent must be made on the basis of the expected lipophilicity of the library. DMSO may be used as the injection solvent for small injection sizes; however, this solvent tends to broaden the early-eluting polar peaks and gives a large UV peak at the void volume. Occasionally even DMSO fails to dissolve certain samples.

The use of CLND to determine the concentration of sample components relies on the separation of these components and the internal standard from other sample constituents. Therefore, peaks that co-elute with the internal standard or the peaks of interest are a major impediment to automated integration and data processing. Internal standard peak areas may be utilized to detect compounds that co-elute with the internal standard because the area for the internal standard is fairly constant barring the presence of co-eluting components. Adequate chromatographic resolution must be achieved to produce reliable data despite the desire to run the separation gradient as quickly as possible. In addition, because of the splitting of the column effluent, resolution losses occur between the UV and other detectors. This loss is minimized with the use of the LC Packings splitter, but it is still significant. As controls, samples containing only the internal standard and a known pure test compound at various concentrations may be used to verify the proper functioning of the system.

Because of the stringent nature of the determination of concentration via the CLND, the values reported were often

much lower than expected by the synthetic chemists. Differences were found to be attributable to the presence of sample impurities. These included not only the side products of the synthesis that may be observed by the detectors used in this system but also the presence of water and solvents (due to incomplete drying and the hygroscopic nature of some samples) that are invisible to these detectors. These components elute in the void volume of the column and are not included in the UV purity determination or MS data, are too volatile to be seen with the ELSD, and do not contain nitrogen required by the CLND. Therefore, although the relative area percent purity may be high by ELSD or UV, the CLND still may give a low value for the concentration. To prove the presence of these solvents requires the use of NMR, sample drying, and reweighing. Often, a combination of reaction side products and retained solvent is seen, making the observed purity difficult to reconcile with the reported concentration. In addition, the presence of inorganic components also lowers the observed concentration. These impurities usually are derived from the synthetic processes and include salts, resins, silica, and diatomaceous earth. In addition, plate and filter materials used in the synthesis often make their way into the final sample. All these components can be dissolved by DMSO and yet not be detected. However, since the CLND-derived concentration is determined for a pure HPLC peak, all sample impurities are presumably excluded, resulting in a reliable determination of the concentration of the target compound or the identified impurity.

Applications. (1) Quality Surveys of Combinatorial Libraries. The described system has been utilized to verify the quality of commercially and in-house derived combinatorial libraries using statistical sampling methods. The resulting data enable the selection of the highest quality libraries for screening and suggest a proper dilution for the samples based on the average concentration determined for the library. The results of these surveys also provide a check of quality protocols used by commercial library vendors, enabling the selection of higher quality producers (Figure 8).

(2) Identification of Impurity Peaks and Side Reactions as an Aid to Optimizing Library Synthesis. Identification of byproducts and quenched reagents found in the production of a library may assist the development of better synthetic procedures. As an example (Figure 9), the benefits of mass accuracy are seen for a sample that contained the target compound (expected $m + H^+ = 542.1571$, observed $m + H^+ = 542.1620$) plus an impurity ($m + H^+ = 302.1358$). Although the absence of chlorine could be observed by either a quadrupole MS or a TOFMS detector, several impurity structures could be proposed on the basis of the observed mass, the synthetic chemistry, and the expected impurity profiles that are identical at unit mass resolution. The exact mass data enabled the selection of one of the structures as the most consistent with the expected chemistry and provided information for the optimization of the synthesis.

In an additional example (Figure 10), a library was produced utilizing solid-phase techniques and the linear addition of diversity element linkers. One of the moieties

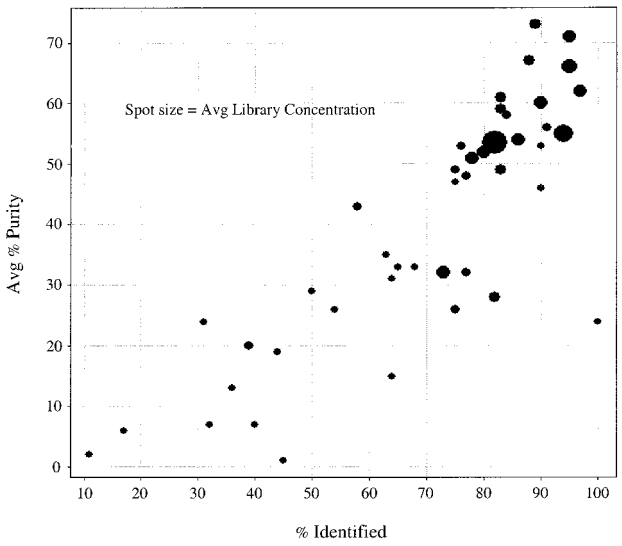


Figure 8. Library quality comparison of identity, purity, and concentration.

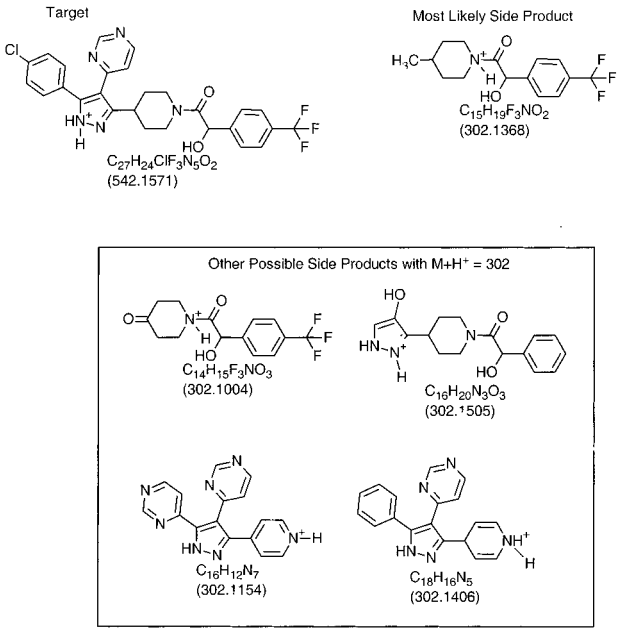


Figure 9. Benefits of high resolution.

used for the third linker, a trimethylenedipiperidine, was found to give side products attributable to the synthesized compound growing large enough to span across sites on the bead. After cleavage from the bead, compounds containing two of the initial linker plus two of the second linker plus one trimethylenedipiperidine moiety were observed. Use of the exact mass data from the TOFMS enabled the rapid identification of these byproducts with errors in the determined versus calculated masses of <10 ppm. This type of side reaction was exacerbated for one of the second linkers that inadvertently contained two possible sites for reaction with the third diversity element (see Figure 11). On addition of the trimethylenedipiperidine, one or both of the sites react. As stated above, because of the size of the third linker, the free end may cross-link to an adjacent site. We observed up to four sites reacting to create polymer-like compounds. Because of the presence of several amines in these compounds after cleavage from the resin, multiply charged ions

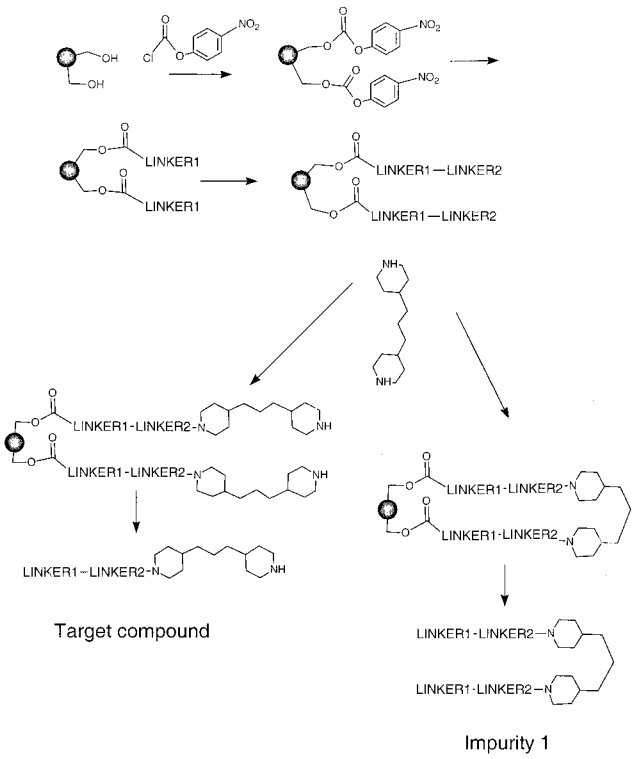


Figure 10. Combinatorial chemistry reaction scheme.

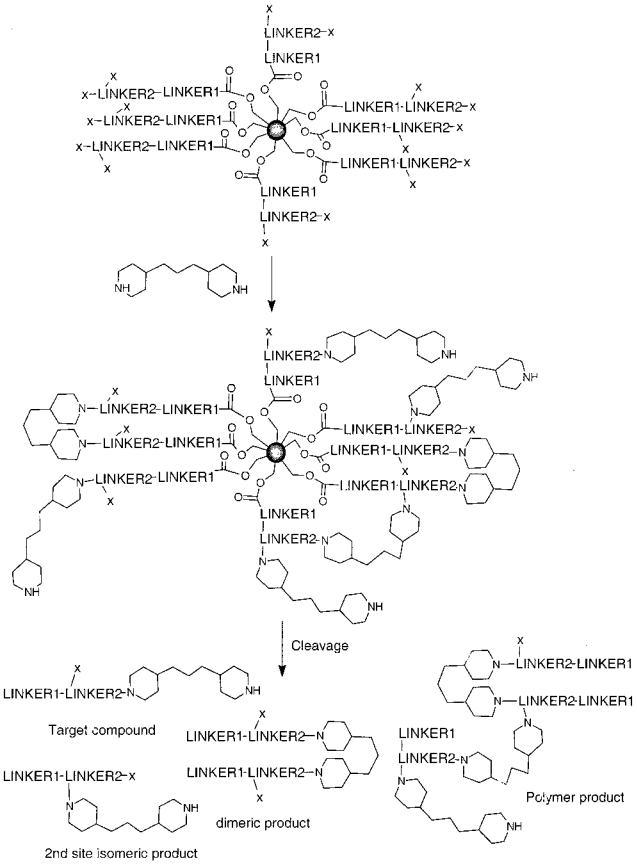


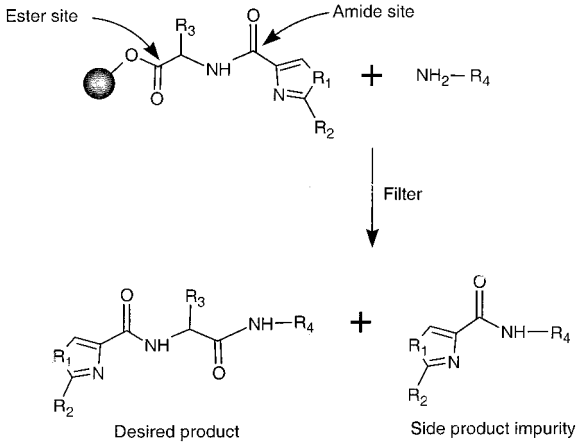
Figure 11. Another combinatorial chemistry reaction scheme.

were predominant in the TOFMS. Identification of impurity components was facilitated by the use of exact masses and the ability of the TOF to identify the charge state. The presence of double and higher charge states is difficult to ascertain using quadrupole type instruments where resolution is limited to 1 amu. This information was critical to the identification of these sample components. Identification of these side reaction pathways indicated that the use of short third linkers and mono-site second linkers would provide better yield and purity of the target compounds.

(3) Prioritization and Identification of Leads from Combinatorial Libraries. The utility of the described system is exemplified in the assessment of hits from the screening of combinatorial libraries. As an example, results from the analysis of a collection of hits from a combinatorial library showed that 44% of the samples contained the expected compound at levels above the detection limit of CLND (5 μM). Concentrations of the target compounds ranged from 7 to 325 μM, or 1.4−65% of the expected 500 μM of the diluted library. This variation emphasizes that concentration information is crucial for prioritization of these leads. Structures of impurities in several samples could also be identified, and the concentrations were determined. For instance, two samples were found to contain ions from the expected compound, but each also contained a sizable impurity. The difference between the mass of the impurity and the expected compound from each sample was calculated to be a loss of 68.0744 and 68.0727, respectively. This mass is equivalent to replacement of a piperazine by a hydroxyl ($-85.0766 + 17.0027 = -68.0739$) functional group. The exact masses determined for both the expected compounds and the impurities differed from those calculated from the respective formulas by 1−2 ppm, assisting in the unambiguous assignment of their structures. This assignment enabled the concentration of each impurity to be determined from its CLND peak area. The first sample has a target compound concentration of 71 μM and an impurity concentration of 39 μM, while the second sample contains the target compound at 327 μM and the impurity at 55 μM. Therefore, since these samples were nearly equivalent in activity in the screen (39% and 43% inhibition, respectively), the data suggest that the first sample is actually 4−5 times more active than the second if the target structures containing piperazine are responsible for the activity. However, more importantly, the equivalence in activity and similarity in impurity structures and concentrations may indicate that the hydroxyl byproducts are a possible source of activity in the assay. Once the mechanism of byproduct formation is established, one can quickly search for the presence and concentrations of the byproducts in other library samples (both active and inactive) to delineate the bioactivities observed for this series of compounds. Samples that contain pure target or byproduct greatly facilitate this process. Thus, the identity of the active sample component and an initial SAR may be determined. However, because of the similarity in structure of target and impurity compounds, often both may be active, requiring fractionation and reassay to identify the contribution from each component.

As a second example, the analysis of a set of hits (target concentration of 2.5 mM) from a different screen showed that all of the samples contained the target compound at levels detectable by the CLND. Concentrations of the expected compounds varied from 1 to 1361 μM or 0.04% to 54% of the expected concentration. This variation again

Scheme 1



emphasizes that concentration information is crucial for prioritization of these leads. Furthermore, the common presence of side products containing one structural motif was seen for all but four of the samples. After this discovery, a search of the library synthesis database was made on these samples. Despite the screening of several library chemistries, all the active compounds in this set were from the same parallel synthesis solid-phase protocol and all but six were synthesized in the same column of the synthesis plate. In this column, the final synthetic step is the formation and cleavage of the product from the resin, displacing the ester linkage to the resin by addition of an amine (R_4 in Scheme 1). A scenario for the formation of the side products consistent with the high-resolution MS results involves reaction of the amine with an internal amide linkage instead, displacing not only the resin but also the center portion (R_3) of the desired structure between the ester and amide linkages.

The facts that these active samples are mainly from the same column of the synthesis plate involving a specific amine, contain similar impurities at significant concentrations, and that several of the samples contained >90% pure side product indicate that the screening activity is derived from the similar side products and not from the target compounds. Since the concentrations of these compounds could be determined, initial structure−activity relationships could be proposed, enabling the chemists to focus their efforts on the most potent template. This finding underscores the need to have a powerful analytical system to unravel hits in the screening of impure libraries.

Conclusions

As the vast number of samples from combinatorial libraries exerts its leverage through high-throughput screening, it is important to remember the assumptions made to produce them and the impact of these presuppositions on prioritization of the hits found. Unless the library products are individually purified and weighed, it is difficult to determine the SAR from the screening data. However, since the majority of these samples will never be active in a screen, several companies screen the impure combinatorial chemistry products and characterize samples only from the active libraries. In this approach, the screening data are not directly comparable across the library because the compounds do not have the same purity and concentration. In addition, without further characterization, the target compounds may not be assumed to be the active component for the purpose of identifying possible templates. Purifying compound libraries prior to screening or assessing the parameters of identity, purity, and concentration after hits are identified can potentially add weeks to months of additional work to an increasingly shortening timeline. In this report, the development and application of a new assay system for the simultaneous determination of identity, purity, and concentration of sample components are described. The system makes use of high-performance liquid chromatography with photodiode array (PDA), ELSD, CLND, and TOFMS detection. The use of a TOFMS along with a CLND provides a synergistic combination enabling target and side-product structures to be identified and their concentrations and purities determined in a single experiment from a solution containing microgram levels of material without the need for primary standards. The CLND was found to give a uniform response based on the number of moles of nitrogen present, enabling the concentration of nitrogen-containing components to be determined utilizing an unrelated co-injected standard. The use of ELSD area percent to estimate purity was found to correlate well with concentrations determined by actual weight or by the CLND for similar compounds. Purity percentages obtained from UV at 220 nm were not found to correlate as well with the actual concentrations. However, although not demonstrated in this study, the availability of the UV/vis PDA could be beneficial because it acquires absorption spectra for all HPLC peaks. This detector does not consume sample and provides an indication of peak broadening in the system.

The technology described in this report provides a means to survey the parameters of correct identity, purity, and concentration for combinatorial libraries, enabling the retention and screening of only the libraries of good quality. The average concentration of a library may also be used to adjust its dilution prior to screening. The system also facilitates the rapid identification of side products to guide the optimization of synthetic chemistry procedures. The most powerful use of this system is to rank screening lead structures on the basis of the identity, purity, and concentration of the sample components. High-purity leads may be graded by potency using the absolute concentration determined from the CLND response, providing initial SAR information. The pure compounds should be easily resynthesized by repeating the library chemistry. Leads from impure samples may be prioritized on the basis of the concentration of the sample component (target or impurity compounds) suspected to give activity in the assay. It is likely that sample impurities are responsible for a portion of the activity found in screening unpurified libraries. Their activity may be verified by identification of library samples (active or inactive) containing a single component or by fractionation of the sample into individual components and reassay. The use of TOFMS facilitates the unambiguous identification of sample constituents, enabling their absolute quantitation by CLND without the requirement of primary standards. The TOFMS used for this work is also capable of acquiring data

Table 3. Parameters for Positive Ions

capillary (V)	sample cone (V)	extraction cone (V)	source temp (°C)	desolvation temp (°C)	nebulizer gas flow	desolvation gas flow	pusher cycle time (μs)
3500	35	10	80	100	71	497	50

Table 4. Instrument Settings for CLND Data

detector vacuum, Torr	reaction chamber pressure, Torr	makeup helium flow, mL/min	inlet helium flow, mL/min	inlet oxygen flow, mL/min	ozone production, mL/min	low-pressure set point, psi	high-pressure set point, psi	sensitivity	gain
22	42	57	200	132	25.4	3	40	×50	high

from multiple LC systems using a multiple inlet source[17] as a further improvement in throughput.

Experimental Section

Sample Preparation. An internal standard solution (DMSO or a solution of 49.5% MeOH, 49.5% water, and 1% DMSO containing 258 uM caffeine) was added to an aliquot of a 2.5 mM DMSO solution of the sample in a 96-well polypropylene plate. Alternatively, the DMSO was removed from a larger volume of the 2.5 mM samples by evaporation in a vacuum centrifuge and the sample was reconstituted in an equal or lesser volume of the internal standard solution. The concentrations of the samples used for sample preparation are based on the sample weight reported by the supplier. After sample mixing, the plates were heat-sealed with a polypropylene/aluminum foil cover to create an airtight seal. Samples were stored at −20 °C and warmed and mixed on a shaker prior to analysis.

HPLC−PDA−ELSD−CLND−TOFMS Analysis Parameters. Figure 1 is a flow diagram of the instrumentation for the assay. The sample was supplied to the system with a Waters 2700 sample manager. The HPLC gradient was supplied by a new Waters Alliance 2690 system with a column heater set at 60 °C. Flow proceeded from the column through a Waters 996 photodiode array (PDA) detector to a custom-built LC Packings flow splitter. In this module, the 1 mL/min flow was split 200/200/600 to the Micromass LCT time-of-flight mass spectrometer (TOFMS), the Antek 8060 chemiluminescent nitrogen detector (CLND), and a Sedex 55 evaporative light-scattering detector (ELSD), respectively. Initial attempts utilized a Harvard syringe pump to add a 20 μL flow of a 0.2 μg/mL solution of erythromycin in methanol to the flow going to the TOFMS to provide a lock mass for adjustment of the calibration of the instrument for exact mass determinations. Later versions used a 20−3 μL/min flow gradient of erythromycin provided by a Varian 9010 pump to give a constant intensity lock mass signal of approximately 100 cps. The initial minute of elution was diverted away from the TOFMS to avoid contamination by DMSO and other void volume components. The void volume was observed by all other detectors. The TOFMS, HPLC pump, and PDA detector were all controlled by the Micromass Masslynx data system. Data from all the detectors were accumulated by this data system and manually aligned to reference each trace to the TOFMS data, utilizing the signal for the internal standard in each detector.

An elution gradient of water and methanol containing 0.1% formic acid was used. Under the chromatographic conditions, the caffeine internal standard eluted at 3.8 min.

LCT Conditions. The TOFMS was calibrated by infusion of a PEG or polyalanine standard solution to calibrate the instrument over a mass range of 130−1000 amu prior to sample introduction. General instrument conditions were utilized for the samples that would detect all positive ions formed without fragmentation. The parameters for the positive ions are shown in Table 3. Ions were monitored from 100 to 1200 amu with a cycle time of 1.05 s. Therefore, 20 000 scans were summed each second as one data point.

PDA Conditions. The detector monitored wavelengths from 200 to 700 nm. A wavelength of 220 nm was selected for analysis of purity.

ELSD Conditions. The nitrogen pressure was set at 40 psi, the temperature was set to 40 °C, and the gain was set at a value of 10. The baseline signal was adjusted with the manual knob to give a value of 25 mV so that the Masslynx data system would adopt a 1 V full-scale setting.

CLND Conditions. The chemiluminescent emission is specific and proportional to the amount of nitrogen in the eluted sample. The instrument achieves this by combusting the column effluent in a 1050 °C furnace in the presence of O_2. All effluents are converted to CO_2, NO, water, and other oxides. Water is then removed in a drying tube, and the remaining gases reacted with ozone. This reaction with NO produces NO_2*, a high-energy molecule that relaxes to NO_2, yielding a photon. The light produced is measured with a photomultiplier and is directly proportional to the amount of nitrogen in the eluant, thus providing a concentration of nitrogen. Conversion of this to the number of moles of nitrogen and the number of moles of compound is then straightforward. The vendor has determined[14] that a few rare types of nitrogen-containing compounds give N_2 and NO under these conditions, thus allowing only $^1/_3$ of the nitrogen to be quantified. Since these rare structures are likely to be produced only by direct synthesis and not as side products, their different modes of reaction may easily be accounted for in the calculation of compound concentration. The linearity of response to nitrogen permits any nitrogen-containing compound to be used to produce calibration curves that are applicable to the quantitation of any nitrogen-containing compound after applying a factor accounting for differences in the number of nitrogens in the standard and sample. The instrument settings used are shown in Table 4.

Data Analysis. Micromass OpenLynx Diversity software (*.olp program) was used to automatically perform integration of all of the chromatograms, providing areas from the CLND trace for concentration calculations and area and area percents for the UV 220 nm and the ELSD chromatograms for purity determinations. The program also automatically provided an exact mass measurement for every peak observed in any of the detectors, utilizing the lock mass signal to adjust the calibration curve. The program identified peaks due to the expected sample components by comparison of the expected molecular weights with that obtained from exact mass measurements. The OpenLynx program generated a file (*.rpt) that could be viewed in the vendor Diversity Browser program. Final reports, generated by extraction of relevant data from this file by a Visual Basic program written in-house, reduced the data to an Excel file that included areas from the CLND trace and area and area percents for the UV 220 nm and the ELSD chromatograms as well as the exact mass determined for each peak. The concentrations of sample components were calculated from the CLND areas using the following equation:

$$\text{ConcSample} = \frac{(\text{SampleArea})(\text{IntStdConc})(\text{IntStd\#N})}{(\text{IntStdArea})(\text{Sample\#N})}$$

Exact masses of peaks other than the target compound were used to generate molecular formulas using the software. After manual consideration of the chemistry involved in the synthesis of the samples, this allowed the structures to be solved and the number of nitrogens per molecule to be determined for the significant impurities. After the number of nitrogens for an impurity were input into the Excel spreadsheet, the impurity concentration was then determined using this value and the CLND area reported for the peak and the internal standard using the above equation. An automated method of identifying impurities requiring input based on starting materials, known impurities, synthetic chemistry, and expected side reactions should be possible for these libraries. An automated check for compounds coeluting with the caffeine internal standard was accomplished by comparing the CLND area for the individual run with the average area seen for the entire plate and/or control samples. Control samples containing only the internal standard and a known pure test compound at various concentrations may be used to verify the proper functioning of the system.

References and Notes

(1) Yurek, D. A.; Branch, D. L.; Kuo, M. S. Development of an Analytical System to Evaluate Compound Identity, Purity, and Concentration in One Experiment Using HPLC−DAD−ELSD−CLND−TOFMS. Poster presented at the AAPS meeting in New Orleans, November 1999.

(2) Yurek, D. A.; Branch, D. L.; Kuo, M. S. Development and Application of a System To Evaluate Compound Identity, Purity, and Concentration in One Experiment Using Liquid Chromatography with UV/Vis Diode Array, Evaporative Light Scattering, Chemiluminescent Nitrogen, and Time-of-Flight Mass Spectrometry Detection (LC−PDA−ELSD−CLND−TOFMS). Poster presented at the ASMS meeting, June 2000.

(3) Gorlach, E.; Richmond, R.; Lewis, I. High-Throughput Flow Injection Analysis Mass Spectroscopy with Networked Delivery of Color-Rendered Results. 2. Three-Dimensional Spectral Mapping of 96-Well Combinatorial Chemistry Racks. *Anal. Chem.* **1998**, *70*, 3227−3234.

(4) Fang, L.; Wan, M.; Pennacchio, M.; Pan, J. Evaluation of Evaporative Light-Scattering Detector for Combinatorial Library Quantitation by Reversed Phase HPLC. *J. Comb. Chem.* **2000**, *2*, 254−257.

(5) Taylor, E. W.; Qian, M. G.; Dollinger, G. D. Simultaneous On-Line Characterization of Small Organic Molecules Derived from Combinatorial Libraries for Identity, Quantity, and Purity by Reversed-Phase HPLC with Chemiluminescent Nitrogen, UV, and Mass Spectrometric Detection. *Anal. Chem.* **1998**, *70* (16), 3339−3347.

(6) Gorimar, T. S. Total nitrogen determination by chemiluminescence. Application News, Volume 2, Chapter 6, Antek Instruments, Houston, 1981.

(7) Shah, N.; Gao, M.; Tsutsui, K.; Lu, A.; Davis, J.; Scheuerman, R.; Fitch, W. L. A Novel Approach to High-Throughput Quality Control of Parallel Synthesis Libraries. *J. Comb. Chem.* **2000**, *2*, 453−460.

(8) Combs, M. T.; Ashraf-Khorassani, M.; Taylor, L. T.; Fujinari, E. M. Optimization of Chemiluminescent Nitrogen Detection for Packed-Column Supercritical Fluid Chromatography with Methanol-Modified CO_2. *Anal. Chem.* **1997**, *69*, 3044−3048 and references therein.

(9) Keifer, P. A.; Smallcombe, S. H.; Williams, E. H.; Salomon, K. E.; Mendez, G.; Belletire, J. L.; Moore, C. D. Direct-Injection NMR (DI-NMR): A Flow NMR Technique for the Analysis of Combinatorial Chemistry Libraries. *J. Comb. Chem.* **2000**, *2* (2), 151−171 and references therein.

(10) A Hewlett-Packard 1100 HPLC system was flushed in an attempt to remove the nitrogen-containing solvents and buffers. However, despite repeated washings according to Antek's instructions, the presence of nitrogen-containing components was still apparent in the rolling baseline and drop in signal on a loop injection of methanol.

(11) Initially, t-fittings with or without flow adjustment valves were installed. However, these resulted in significant peak broadening and gave poor performance over the changing pressure conditions encountered in the gradient run, resulting in shifting retention time differences among the various detectors.

(12) Henry, C. In the Spray with ELSDs. *Anal. Chem*. **1997**, September, 561A−563A.

(13) Mourey, T. H.; Oppenheimer, L. E. Principles of Operation of an Evaporative Light Scattering Detector for Liquid Chromatography. *Anal. Chem.* **1984**, *56*, 2427−2434.

(14) Hopia, A. I.; Ollilainen, V.-M. Comparison of the Evaporative Light Scattering Detector (ELSD) and Refractive Index Detector (RID) in Lipid Analysis. *J. Liq. Chromatogr*. **1993**, *16* (12), 2469−2482.

(15) Kermasha, S.; Kubow, S.; Goetghebeur, M. Comparative High-Performance Liquid Chromatographic Analyses of Cholesterol and Its Oxidation Products Using Diode-Array Ultraviolet and Laser Light-Scattering Detection. *J. Chromatogr. A* **1995**, *685* (2), 229.

(16) Chaves das Neves, H. J.; Morais, Z. B. HPLC Assay of Underivatized Free Amino Acids with Column Switching and Evaporative Light-Scattering Detection. *J. High Resolut. Chromatogr.* **1997**, *20* (2), 115.

(17) Micromass product literature. Cottee, F.; Haskins, N.; Bryant, D.; Eckers, C.; Monte, S. The use of accurate mass measurement by orthogonal time-of-flight mass spectrometry in pharmaceuticals research. *Eur. J. Mass Spectrom.* **2000**, *6* (2), 219−224.

CC010028M