Molecular Diversity, **5:** 35–49, 2000.
KLUWER/ESCOM

Analysis of libraries encoded with GC tags: Compound elution, tag decode analysis, and statistical sampling analysis

Roland E. Dolle[1,*], Joan Guo[1], Wenni Li[1], Ning Zhao[2] & James A. Connelly[3]
[1] *Pharmacopeia Inc., CN5350, Princeton, NJ 08540–5350, U.S.A.;* [2] *Bristol-Myers Squibb, P.O. Box 5400, Princeton, NJ 08543, U.S.A.;* [3] *Selectide Corp., 1580 Hanley Blvd., Tucson, AZ 85737, U.S.A.*
(Author for correspondence, present address: Adolor Corporation, 371 Phoenixville Pike, Malvern, PA 19355, U.S.A., E-mail: roland@adolor.com)*

Received 30 August 2000; Accepted 8 January 2001

Summary

Libraries encoded with electrophoric tags present a unique challenge with respect to library quality control and characterization. Libraries are prepared on Tentagel resin in 200-fold redundancy wherein each resin particle contains one compound per one tag set. The amount of compound present on the bead is ca. 200–500 pmole while tag levels are estimated at 0.5–1 pmol/bead. Several quality control protocols have been developed in order to accurately estimate bead yield and purity for the entire library, ensure high tag fidelity, and to determine the overall performance of individual synthons. This review provides a unique, collective portrait of Pharmacopeia's approach in assessing the quality of libraries prepared using its molecular encoding technology.

Introduction

Since its inception in 1993, Pharmacopeia has utilized encoded split synthesis methods to create libraries of small molecules for biological screening [1]. A binary encoding protocol is employed to index library members relying exclusively upon electrophoric molecular tags as originally described by Still [2]. In this scenario, as sets of synthons are serially combined during synthesis, accompanying binary sets of tags are also attached to the solid-support (Figures 1 and 2). Compound synthesis occurs through initial chemical attachment of the synthon set to an acid- or photochemically sensitive linker, while oxidatively labile tags are incorporated directly into the bead matrix by carbene insertion. The orthogonal linkage strategy allows for the independent release of compound and tag molecules allowing off-bead assays [3]. Because each tag set is uniquely associated with a given set of synthons, the identity of the final compound on the bead can be inferred by decoding the bead, i.e., GC/ECD (electron capture detection) analysis of the detached tags (Figure 3). Application of the electrophoric-based tagging protocol to encode libraries from our laboratories [1,4] as well as from other laboratories [5], have been published.

Encoded libraries are prepared in 200–300 fold redundancy, meaning that there are 200–300 bead copies, each of which has bound to it a single compound. This ensures an ample supply of the library for many assays, and more importantly, ensures reliable statistical sampling of library compounds during biological evaluation [3]. Thus a typical 50,000 member library is composed of 10–15 million beads, wherein each resin particle contains one compound and one tag set. The amount of compound present on the bead is on the order of ca. 500 pmole/bead while tag levels are estimated at 0.5–1 pmole/bead. The large redundancy, picomole amounts of compound, trace levels of tags, and the necessity to efficiently elute compounds off the beads for solution-based biological testing, pose significant challenges to the analytical chemist assessing library quality. We disclose for the first time the analytical protocols employed in our laboratories to address these analytical issues. Specifically discussed herein are the applications of (i) on-line HPLC/UV/ELSD/MS to establish compound identity, quantity, purity and optimal conditions for eluting

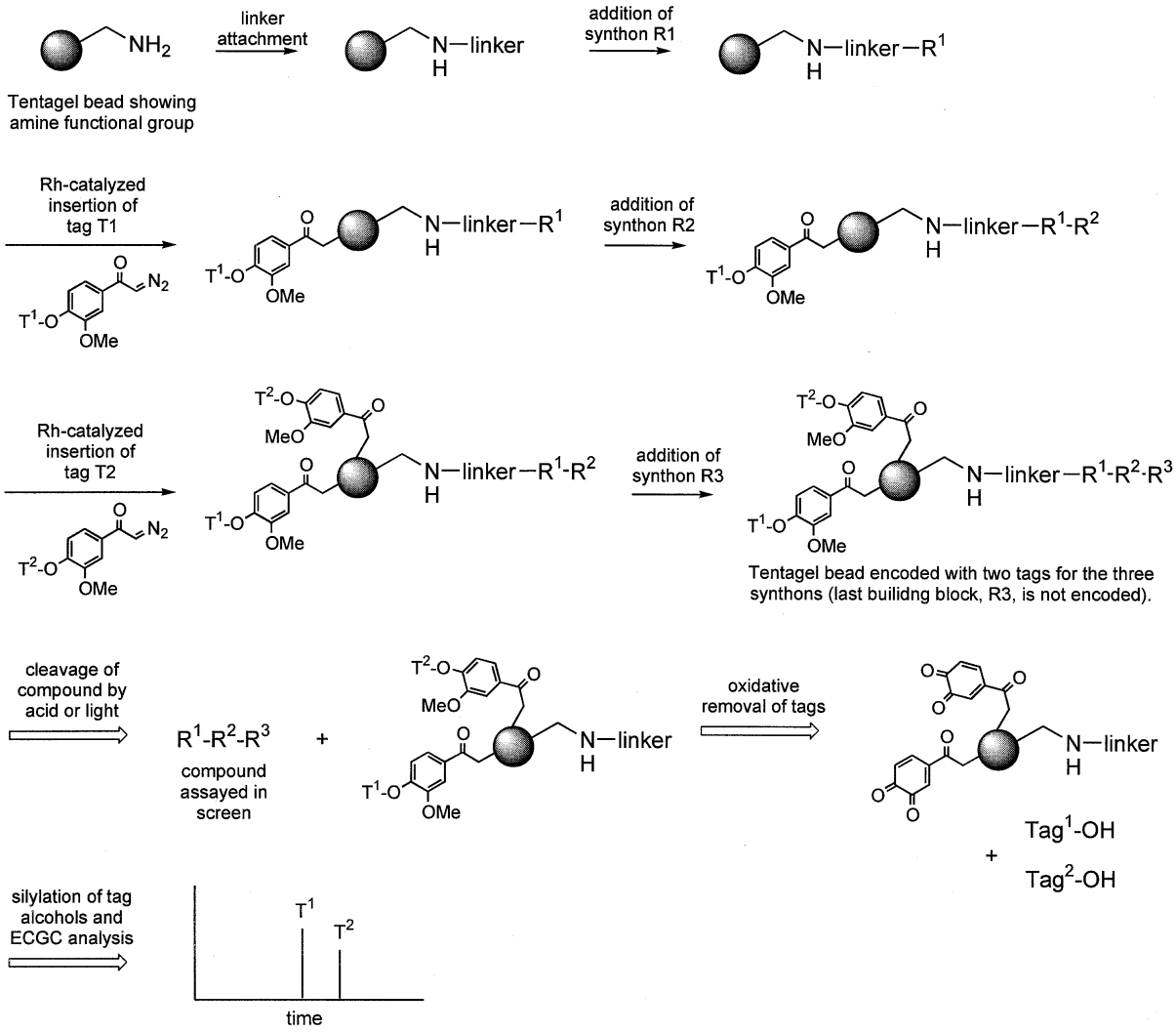


Figure 1. Anatomy of the electrophoric-based tagging technology.

compounds off the solid support (*library elution*), (ii) GC/ECD for tag analysis, and (iii) decode-assisted, single bead LC/MS statistical sampling methodology for broadly assessing library quality.

Library yield, purity and optimal elution method

There are four phases to library synthesis. Phase one is the library design and chemical feasibility phase where the chemist conceives of a library design and demonstrates its feasibility by carrying out a limited number of synthetic transformations. Phase two is solid-phase development in which many runs through the complete reaction sequence are performed, and the yield (gravimetric) and purity (HPLC) of many putative library members are measured. In this phase, synthon pairings are examined, defining their compatibility within the context of the proposed library. Phase three is the quality control (QC) sample submission phase; here, five to ten sets of on-bead and purified off-bead samples are submitted for detailed analysis. Rigorous identity, quantity, and purity for each QC sample is determined, simultaneously establishing the optimal elution solvent and elution time for the library. In phase four, library construction with molecular encoding is completed.

The QC samples, as submitted by the chemist, are designed to represent the range of clogPs present in the library and the significantly different chemistries

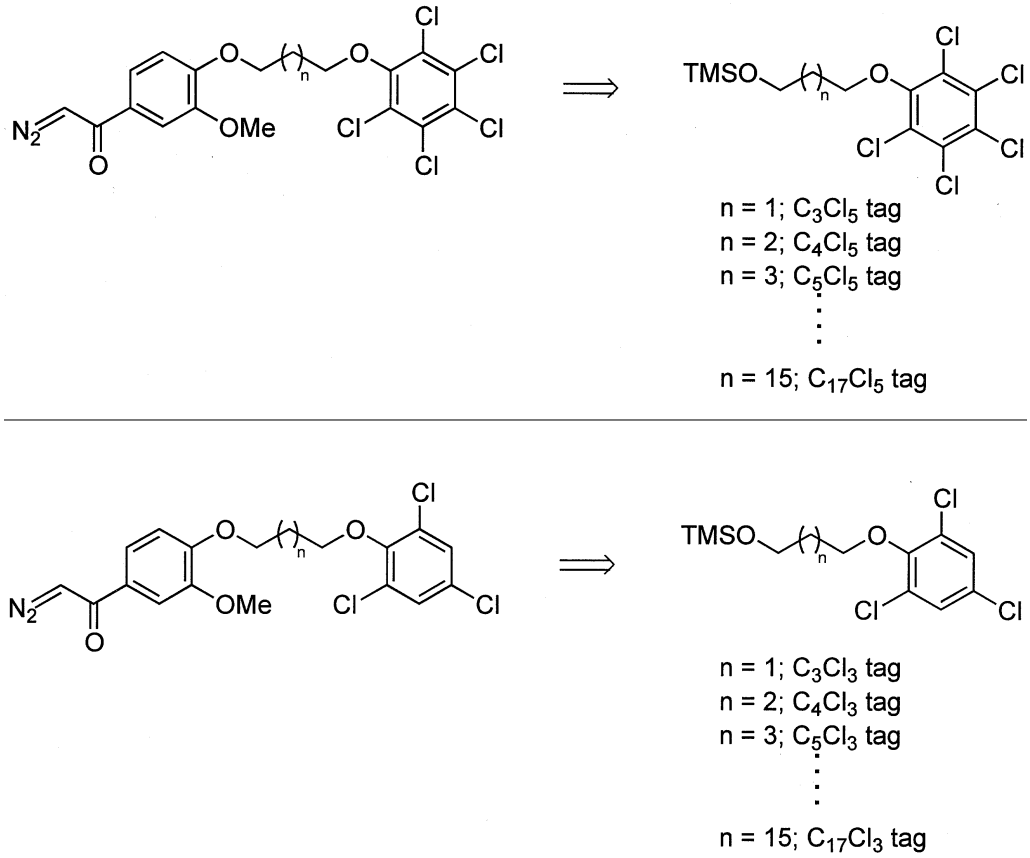
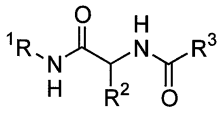


Figure 2. Structures of tagging molecules and corresponding trimethylsilyl ethers for GC/ECD analysis.

utilized in its synthesis. Because it is currently not practical to obtain quantitative information and elution conditions for each compound in the library, the QC compounds play a pivotal role in estimating the bead yield and purity for the entire library. The QC bead yield and purity information feeds back to both the chemist, who may decide that further optimization of the chemistry is needed before entering into the final phase of library construction, and the biologist, seeking to estimate screening concentration.

QC compound analysis and elution method development proceeds through two steps. The first step is developing an HPLC method and an external standard calibration curve for each of the purified off-bead QC samples. HPLC analysis is typically performed in tandem with on-line UV (or ELSD, depending on the presence or absence of a UV chromophore) and mass spectrometry. In general, a single HPLC method (indicated below) may be employed for the analysis of up to 90% of the QC compounds, while customized HPLC methods are developed for the remainder of the compounds. The second step involves eluting compound from the on-bead QC sample after photolysis (or chemical) cleavage and quantitating the yield against the corresponding external calibration curve. Alcohol-water (80-20, v/v) or alcohol-water-trifluoroacetic acid (80-20-3, v/v) and four elution time points (15, 30, 60, and 120 min) per elution solvent are routinely used in the initial elution method development. The selection of these two solvent systems stems from early work at Pharmacopeia in which a panel of elution solvents (e.g., water, acetonitrile, dimethylformamide, nitromethane, methylene chloride, and combinations thereof) were surveyed. The aqueous alcohol solutions were found compatible with all assay plate types, sufficiently non-volatile to conduct photo-elution at 50 °C, and to work well for most libraries. If satisfactory yield and purity are observed (closely matching the chemist's expected theoretical yield), then this elution method is validated and forwarded to the production department for library processing; if not, an alternative elution condition must

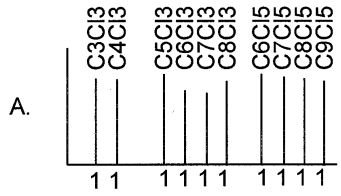


Hypothetical library of amino acid amides
12 (R1 amines) x10 (R2 amino acids) x 2 (R3 acid chlorides) = 240 members

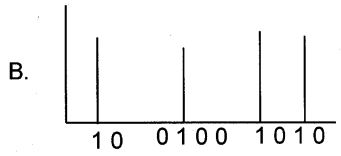
Amines (R1) (C3Cl3 parity tag, C5Cl3-C8Cl3 synthon tags)			Amino acids (R2) (C4Cl3 parity tag, C6Cl5-C9Cl5 synthon tags)			Acid chlorides (R3) (no tags, defines two sublibraries of 120 members each)
MeNH- 1 0001	EtNH- 1 0010	PrNH- 1 0100	-OC-Ala-NH- 1 0001	-OC-Phe-NH- 1 0010	-OC-Val-NH- 1 0100	-COMe sublibrary 1
BuNH- 1 1000	PenNH- 0 1111	HexNH- 0 0110	-OC-Leu-NH- 1 1000	-OC-Pro-NH- 0 0011	-OC-Asn-NH- 0 0110	-COPh sublibrary 2
HepNH- 0 1001	OctNH- 0 1100	NonNH- 1 0111	-OC-Asp-NH- 0 1001	-OC-Lys-NH- 0 1010		
DecNH- 1 1011	BnNH- 1 1101	PhNH- 0 1010	-OC-Arg-NH- 0 0111	-OC-Gly-NH- 0 1111		

Key

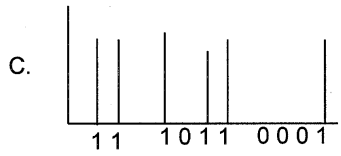
MeNH-
1 0001
- C8Cl3 present
- C7Cl3 absent
- C6Cl3 absent
- C5Cl3 absent } CxCl3 tag set for amine synthons
- C3Cl3 present } parity tag



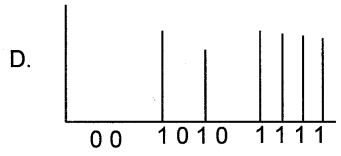
Chromatogram A: Decoded test bead containing all tags.
Chromatogram B: Decoded bead from sublibrary 1.
Chromatogram C: Decoded bead from sublibrary 1.
Chromatogram D: Decoded bead from sublibrary 2.



≡ PrNH- 1 0100 | -OC-Lys-NH 0 1010 | -COMe sublibrary 1 ⟹ PrNH-Lys-Ac



≡ DecNH- 1 1011 | -OC-Ala-NH 1 0001 | -COMe sublibrary 1 ⟹ DecNH-Ala-Ac



≡ PhNH- 0 1010 | -OC-Gly-NH 0 1111 | -COPh sublibrary 2 ⟹ PhNH-Gly-Bz

Figure 3. A hypothetical encoded library.

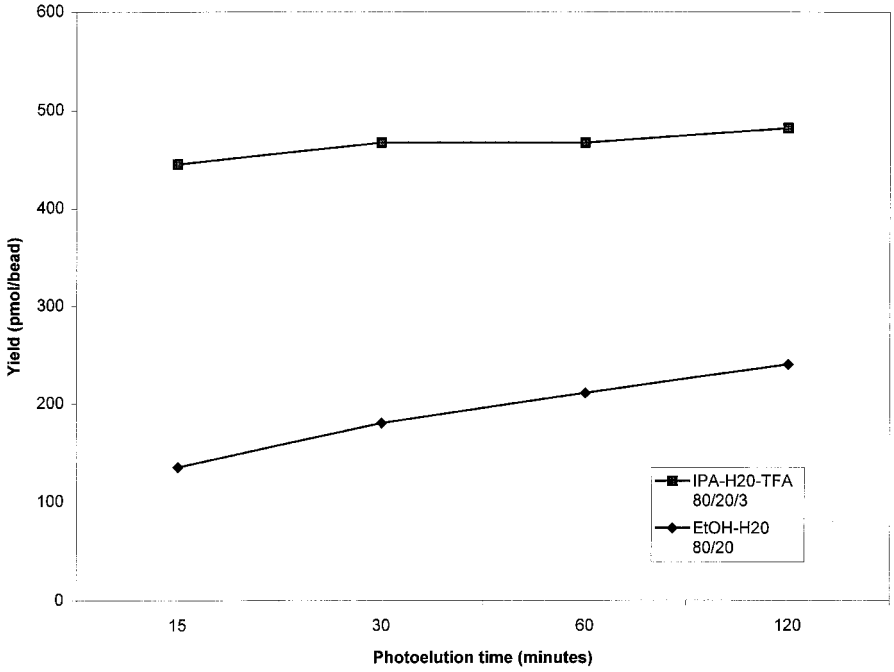


Figure 4. Bead yield determination for PS914204.

be developed, optimizing bead yield and purity.

It should be noted that at one point in Pharmacopeia's history, a partial elution strategy was used by the biologist to assay libraries [3]; however, this has been largely superceded by a full elution strategy. The preferred screening paradigm involves surveying libraries at 10–20 fully eluted compounds/well followed by full elution single compound plates of biologically active sublibraries. The survey plates are discarded, while beads from the single compound plates may be decoded as required. This full elution paradigm simplifies elution optimization and analysis, eliminating the routine need to perform detailed kinetic curves for each QC sample [1i].

Experimental for QC compound analysis (photo elution)

Equipment

(a) HPLC: Waters 2690 separation module or Waters 717 autosampler, Waters 600S solvent delivery system and Waters 616 pumps (Waters, Milford, MA); (b) HPLC detector: Waters 996 photodiode array detector; (c) Master plate: polystyrene, Millipore multiscreen-BVPP non-sterile 96 well filter plates with lids (Millipore, catalog number – MABBVN1250; (d) Derivative plate: polystyrene 96 well round bottom plates (Costar Corp. catalog number – 3794); Plastic bag: nylon/polyethylene film bag, LF5101 (Laminated Films and Packaging, Portsmouth, NH); (e) Centrifuge: Beckman GS-6 centrifuge; (f) UVHP light box: Pharmacopoeia's proprietary high-intensity, heated UV elution oven; (g) Convection oven: 380 FM forced air oven (VWR Scientific); (h) SpeedVac: Savant, AES 2000 system with assembly for microtiter plates (Savant SC210A).

Standard HPLC method

A Phenomenex Luna C_8 column, 3 cm $\times$ 3.0 mm, 3 μm particle size; mobile phase A is acetonitrile containing 0.05% TFA, mobile phase B is 0.05% TFA in water (v/v); 10% A and 90% B at 0 time, 90% A and 10% B at 2.5 min (linear gradient), hold at 90% A and 10% B for 0.5 min, return to 10% A and 90% B in 0.1 min, hold at 10% A and 90% B for 1.9 min. Flow rate is 1.0 mL/min. Total HPLC run time is 5 min.

Elution procedure

Library QC compound beads are arrayed into 96 well filter bottom plates by transferring a small quantity of beads (ca. 20) from a suspension of beads in isopropanol (IPA). The plates are dried (Speed-vac) and the number of beads per well is accurately recorded. Typically, the analysis is carried out using two elution

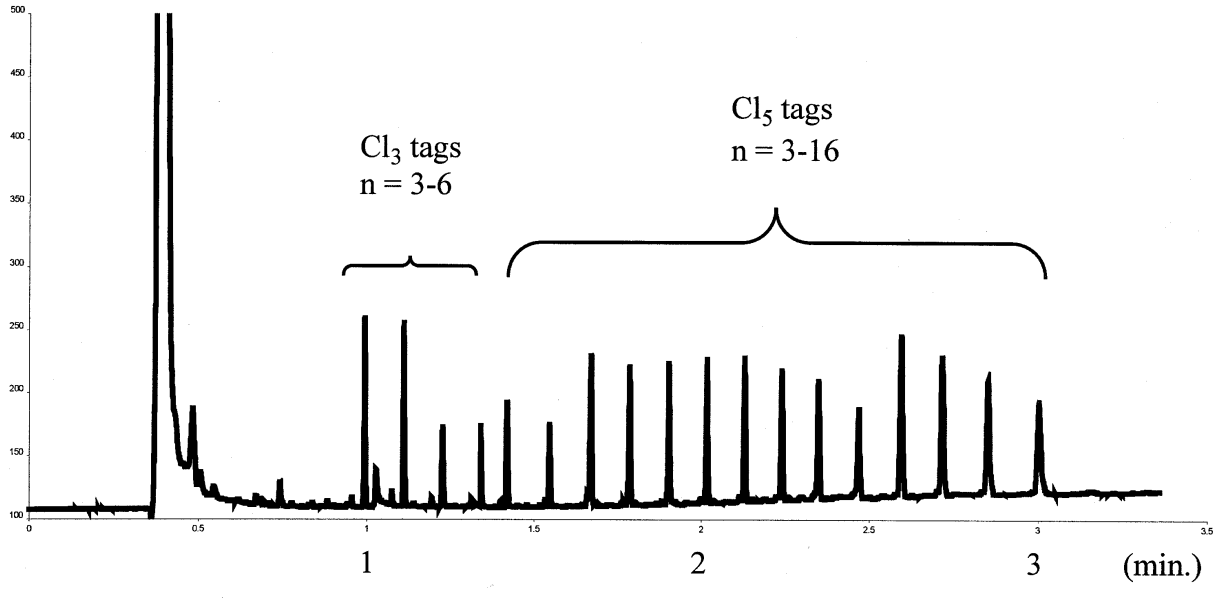


Figure 5. GC chromatogram of Cl_3 and Cl_5 tags on the HP 6890.

solvents, EtOH/H_2O, 80/20, v/v, and IPA/ H_2O/TFA, 80/20/3, v/v/v, with four elution time points (15, 30, 60, and 120 min) per solvent. All samples are arrayed in triplicate.

A stilt is attached to the bottom of the arrayed master plates. An elution solvent (150 μL) is added to each well including blank wells using a multichannel pipette. The plates are sealed, placed in a LF-5101 plastic bag, then the bag with plates are placed in a VWR 1380FM convection oven to pre-soak at 50 °C for 1 h. After 1 h, the bags are immediately transferred into the UVPH light box for photolysis. Plates are removed from the UVPH light box at pre-set time points and transferred to the VWR 1380FM convection oven for a 2 h post-soak at 50 °C. At the end of the post-soak, the plates are removed from the convection oven and taken out of the bags. The stilt is replaced with a pre-labeled derivative plate, and the elution solution is transferred from the master plate to the derivative plate by centrifuging at 1000 rpm for 1 min. The master plates are separated from the derivative plate and the derivative plate is dried overnight in the convection oven at 37 °C, or in SAVANT SpeedVac, SC 210A, for 1 h.

HPLC quantitation

Acetonitrile/water, 80/20, (100 μL, v/v) is added to the appropriate wells of the dried derivative plates. The plates are placed on a plate shaker and agitated for 60 s. The solution so obtained is transferred from the well into a pre-labeled 1.5-mL amber vial with glass insert. The eluted QC samples are quantitated by HPLC against an external standard curve generated from their respective, purified QC compounds. Bead yield is determined as given by the equation

$$\text{Bead yield (pmol/bead)} = \frac{A_{spl}}{A_{std}} \times \frac{C_{std\ (pmol/\mu L)} \times 100\ \mu L}{\text{number of beads/well}},$$

where

A_{spl} = HPLC area count of bead eluant peak.

A_{std} = HPLC area count of standard peak.

C_{std} = Concentration of the standard.

A typical analytical result is shown in Figure 4. In this example, IPA/water/TFA 80/20/3 (v/v/v) for 15 min photo-elution was determined as the preferred elution condition for PS914204.

Tag analysis using GC/ECD

The ease of introducing and removing the electrophoric tags from the beads is ideally suited for the binary encoding of libraries (Figures 1–3). Direct insertion of tags into the bead matrix circumvents the need for restrictive and cumbersome orthogonal protecting group strategies which are necessary with other

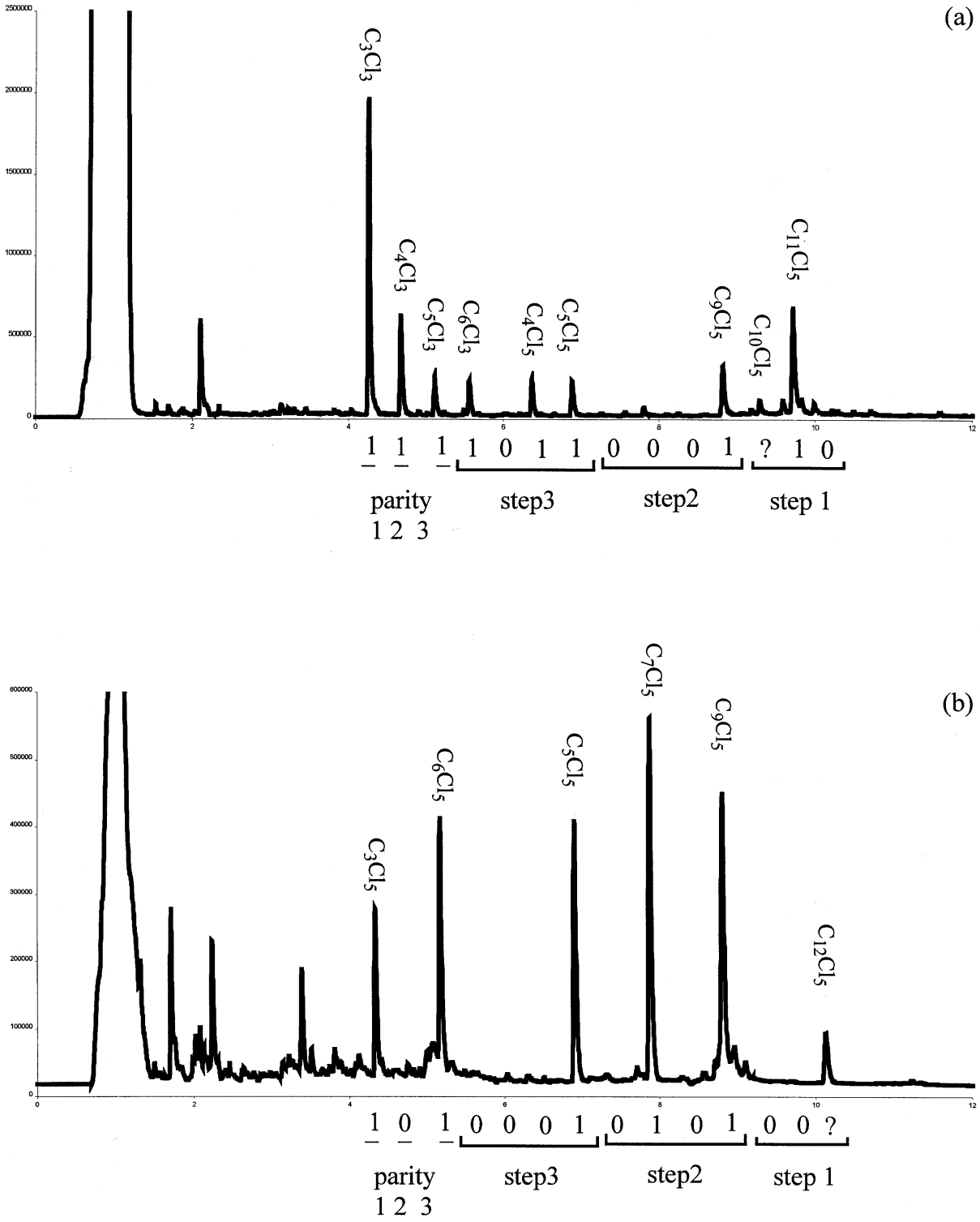


Figure 6. Rare events. When the GC/ECDC auto score is ambiguous, partity tags are used to correctly identify tag signals [6]. (a) GC/ECD auto score of $C_{10}Cl_5$ is incorrect. (b) GC/ECD auto score of $C_{12}Cl_5$ is correct.

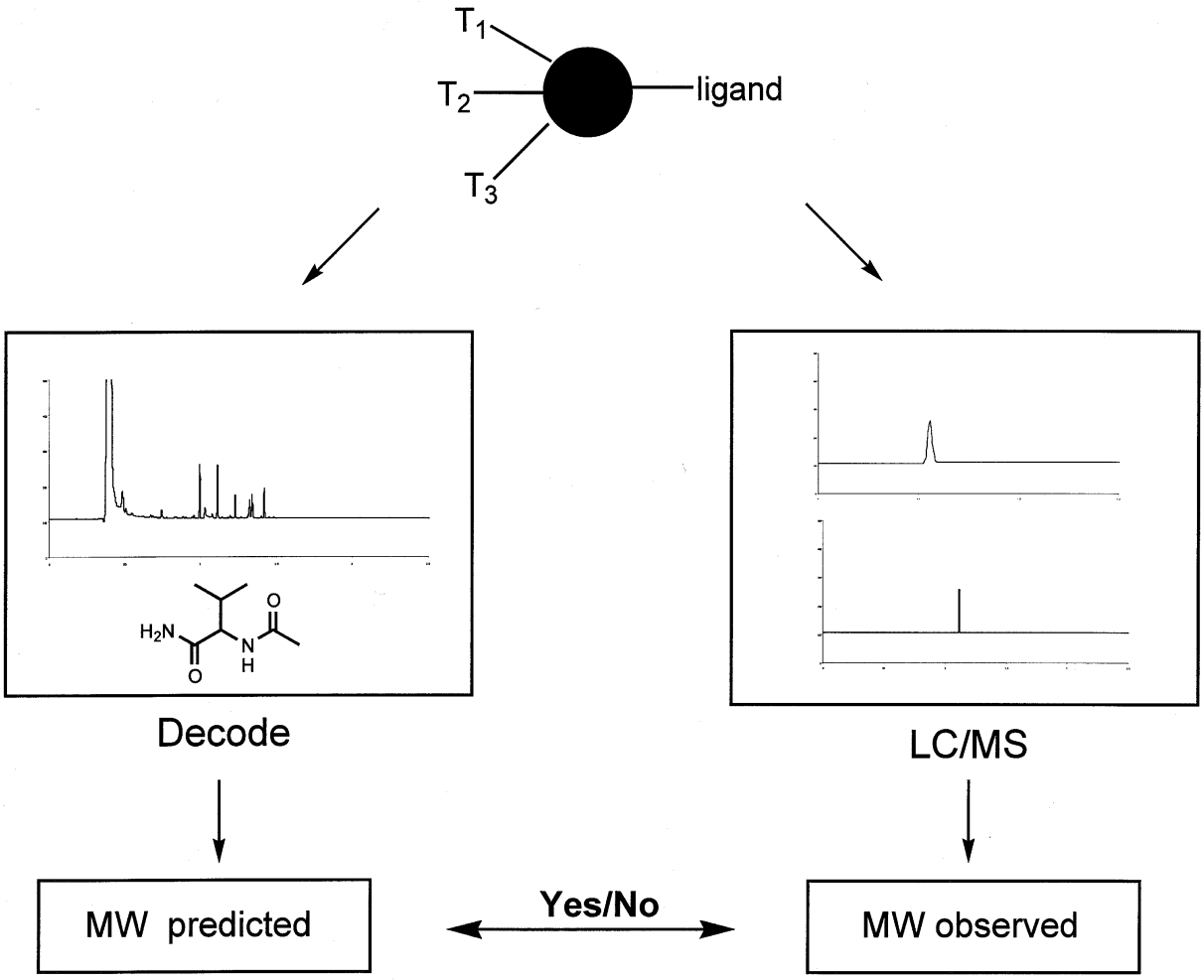


Figure 7. Schematic of decode-assisted, single bead LC/MS methodology.

Table 1. Example of high throughput data analysis for single bead LC/MS

Sample name	File name	Peak	Expected mass	Found intensity	Major peak	Retention time	MW of major peak
93	D06-40	1	651.4	2.66E + 06	No	8.5	540.2
		2		5.18E + 07	Yes	7.6	573.3
93	D07-40	1	746.3	6.18E + 07	Yes	7.0	746.4
93	D08-40	1	643.3	9.11E + 06	Yes	1.1	643.3
93	D09-40	1	590.3	1.66E + 07	Yes	1.2	590.3
93	D10-40	1	664.4	2.58E + 07	Yes	6.9	664.5
93	E01-41	1	676.4	1.75E + 07	No	6.8	676.4
93	E02-41	1	666.4	3.17E + 07	Yes	6.8	666.4
93	E03-41	1	683.3	3.67E + 07	Yes	6.8	683.4
93	E04-41	1	697.4	6.76E + 07	Yes	6.9	697.3
93	E05-41	1	744.4	1.79E + 07	Yes	6.8	744.4
93	E08-42	1	616.4	0.00E + 00	No	0.0	

HPLC Parameters for Single Bead LC/MS Analysis

Column:	Waters SYMMETRY, 3.5μ, C8, 2.1 x 30 mm
Mobile Phase:	A: H_2O, 1% acetic acid
	B: Acetonitrile, 1% acetic acid

Time	%B
0	10
2	90
9	90
9.1	10

Flow Rate:	200 μL/min
Detection:	UV @ 220, 260/Mass Spectrometer (MW 300-800)

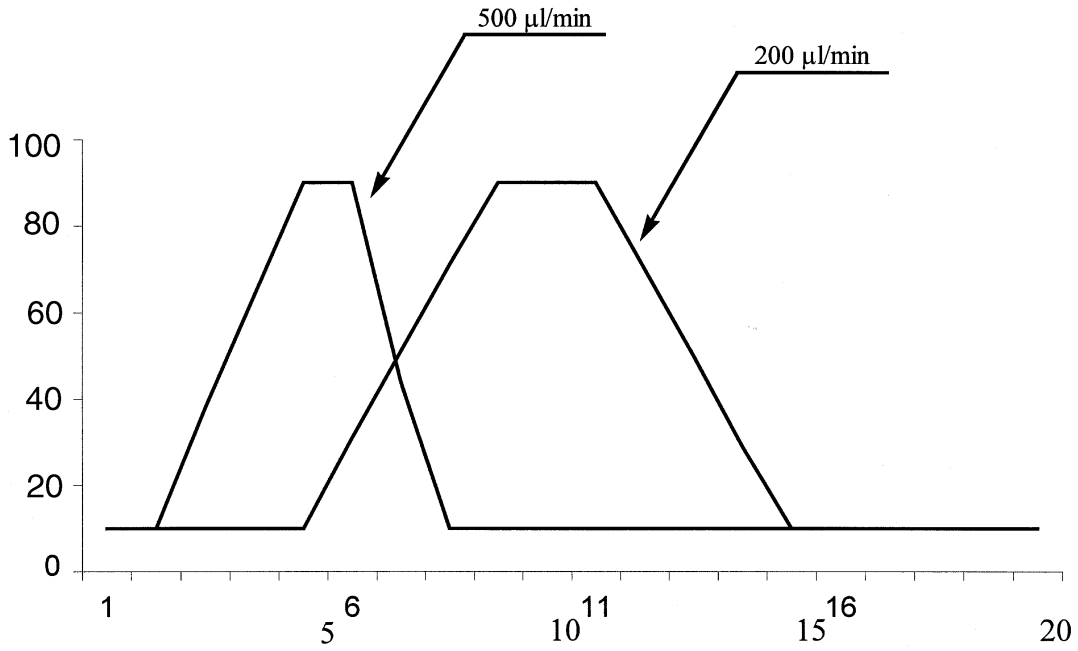


Figure 8. HPLC method for single bead LC/MS, HPLC system void volume study.

methods of molecular-based encoding. A set of 14 molecular tags is sufficient to encode a >16 million bits of information (2n). A hypothetical 360,000 member library, composed of a 30 × 30 × 15 × 25 synthon matrix, would require 17 tags for encoding. Five synthon tags and one parity tag each are required for the first and second sets of 30 synthons, and four synthon tags and one parity tag for the third set of 25 synthons. No tags are used for the final set of 25 synthons as these define separate sublibraries (Figures 1–3).

The diazoketone tags are halophenoxyether derivatives of vanillic acid (Figure 2). The process of encoding is simply the reaction of the carbene derived from diazoketones in the presence of a ruthenium catalyst with resin, resulting in the incorporation of the linker-tag molecules into the bead matrix. In the decoding process, a suspension of the tagged resin in octane is treated with a mild oxidant, aqueous ceric ammounium nitrate (CAN). This selectively, and cleanly liberates the halophenoxy aliphatic alcohols from the resin (vanillic linker remains attached to the resin) providing a solution of the lipophilic alcohols

Mass spectrometer parameters for single bead LC/MS analysis.

ESP voltage (v):	4500
Nebulizer gas (L/min)	8
Drying gas (L/min)	3
ESP source temp. (^{o}C)	350
Flow rate (μL/min):	200
Mass range (amu):	300-800
MS scan rate (second):	2.5 at 0.2 amu with dwell time 1 msec.

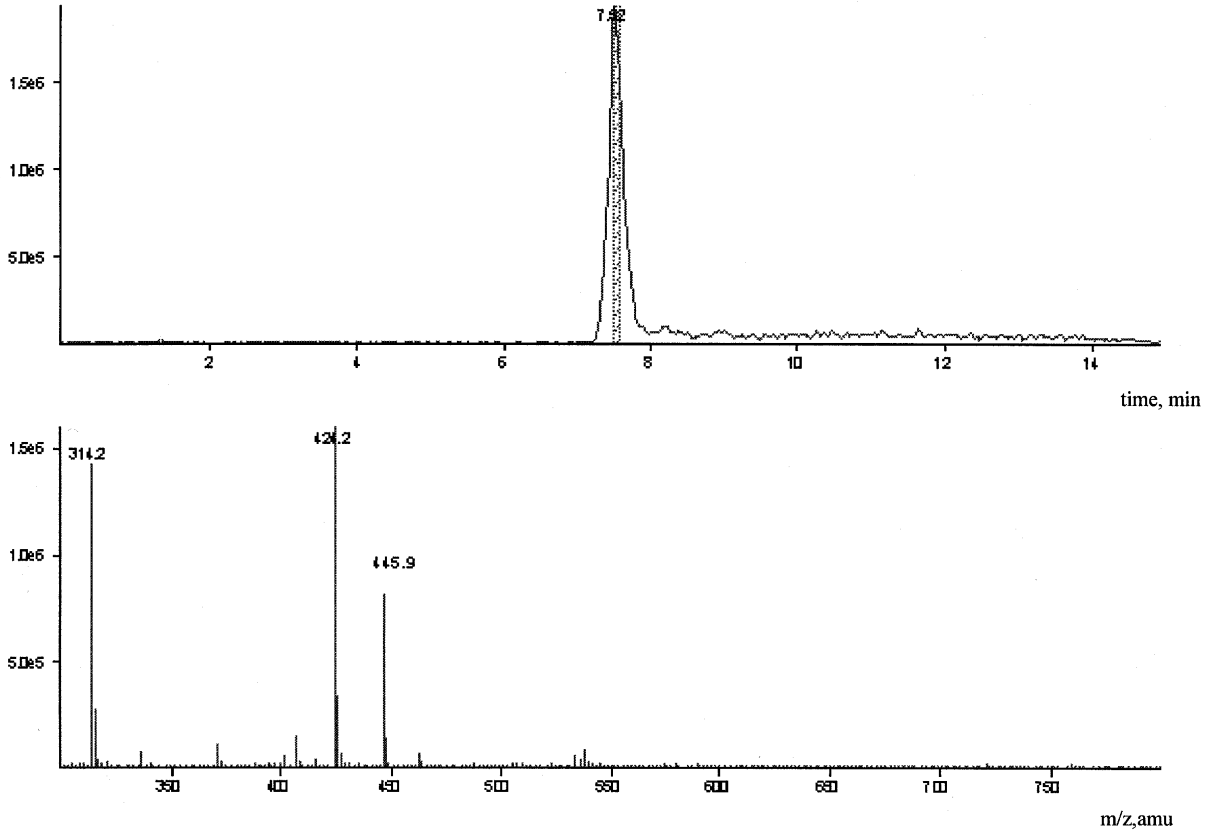


Figure 9. Single bead LC/MS method development.

in octane. The alcohols are then silylated with N-methyl-N-trimethylsilyltrifluoroacetamide (MSTFA) for GC/ECD analysis (Figure 1).

The GC/ECD analysis method has evolved over the past five years in an effort to achieve the highest throughput and efficiency as possible. The original method for extracting the tags began with sonicating the tagged bead in a melting point capillary containing 3 μL of hexane and 1 μL of 0.5 M aqueous CAN in 1:1 acetonitrile-water, followed by removal of the aqueous layer and silylation of the released tag alcohols [2]. This method was cumbersome and clearly unsuited for high throughput analysis. A modification to this method was developed in which single beads were arrayed into a glass insert of a GC vial. Oxidation was performed by adding 2 μL of aqueous

CAN (1:1 acetonitrile-water) and 10 μL of octane to the insert vial, and incubating the reaction for 3 h at 30 °C. After incubation, the octane layer containing the tag alcohols was transferred to another GC vial, and the alcohols silylated with 2 μL of MSTFA. In this method, all liquid transfers were handled by Packard 204 Multiprobe Robotic system and the analysis was carried out on the HP 5890 gas chromatograph. Although this method was utilized for several years at Pharmacopeia, it suffered from the *manual* arraying of beads into the insert vial (requiring a microscope and pipette), limiting the decode rate to about 4–5 beads per h per technician. In addition, the GC analysis time was 14 min on the HP 5890.

More recently, much greater efficiency has been achieved through the use of the HP 6890, which has an extremely rapid ramp-up of column temperature (ca. 150 °C min^{-1}) and is equipped with a μECD. The oxidation is now performed directly in the single bead arrayed master plate as received from Pharmacopeia's production department. The new instrumentation reduces the GC analysis time to 3 min. Figure 5 depicts actual a set of standard tag chromatograms obtained on the HP6890 GC instrument. Application of the HP 6890, coupled with other process changes, has increased output to 150 decodes per 7 h technician day.

The tags are identified by their retention time in the GC/ECD. The binary code number 1 indicates the presence of the tag, and 0 indicates the absence of the tag (Figure 3). A standard set of tags is used to calibrate the GC instrument daily. The retention time of each tag is recorded in the instrument. When the tags are decoded, the retention time of the individual tag is automatically compared with that of the standard and scored in the GC chromatogram. The GC method is highly reproducible, the auto decoding accuracy is >95%.

Although rare events, there are two situations where the auto tag scoring may be confused and parity tags assist in resolving potential ambiguities [6]. For example, occasionally an impurity from the tag extraction process is observed with a retention time close to that of an actual tag and is auto scored as a tag (Figure 6a). Since a parity tag for the first synthetic step (C_3Cl_3) is clearly evident, the incorrectly assigned $C_{10}Cl_5$ peak can be discounted. This is because the total number of tags must be and even number. The C_3Cl_3 parity tag plus the $C_{11}Cl_5$ synthon tag equals two, an even number, versus the C_3Cl_3, $C_{11}Cl_5$ plus $C_{10}Cl_5$ which equals three. Occasionally, the tag level of a true tag may be relatively low and close to the noise level of the instrument. This is illustrated in Figure 6b where there is a potential ambiguity in reading the $C_{12}Cl_5$ tag. Relying on the parity (C_3Cl_5) tag to always provide an even number of tags, the $C_{12}Cl_5$ tag is auto scored correctly.

The decoding process is performed for tag quality control, library quality control (statistical sampling discussed in the next section), and compound decoding following biological screening. Tag QC analysis is carried out after each tagging step to ensure that the designated synthon and parity tags are indeed present and in a minimum concentration. In theory, each bead should contain 0.5–1 pmole of tag, ca. 50 times higher than the μGC/ECD detection limit. The tag QC criteria requires a minimum of 0.25 pmole of each tag on a single bead to guarantee subsequent, error free decoding. Typically three beads from each reaction vessel (each tagged synthon) are sampled. A tag QC submission form is filled out by the chemist indicating which tags are to be evaluated. An external standard containing a mixture of three tags (C_3Cl_3, C_5Cl_5, $C_{12}Cl_5$) at a concentration of ca. 0.25 pmole/injection are analyzed in parallel with the tag QC samples. If the tag levels are at or above the standards, the tag QC is passed; if the tag levels are lower than the standards, the tagging step will be repeated. The tag QC process is carried out for each encoding step of every library.

For structure decode analysis, a decode request is submitted by the biologist when a putative active bead is identified. The decode data so obtained is recorded in an in-house database, which in turn processes the data and generates a decoded structure [1i]. The tags and corresponding synthons have already been registered in the database by the chemist. As a check on the system, the tags from each decoded bead are proof read by two technicians, one as the enterer, another as a verifier. The enterer and verifier enter their reading to the database independently, and upon consensus, a structure-decode report is formally issued.

Experimental for tag decoding

Equipment

(a) GC: Hewlett-Packard 5890 Series II Plus GC or 6890 Series Plus GC, Hewlett-Packard Automatic Liquid Sampler (G1897A, G1916A), Hewlett-Packard Sampler Controller Module. (b) GC detector: Hewlett-Packard Electron Capture Detector. (c) GC column: J&W DB-1 (0.25 μm), 15 m $\times$ 0.25 mm, J and W DB-1 (0.1 μm), 10 m $\times$ 0.1 mm. (d) Autosampler vials: clear, glass, 0.1 mL insert and 2 mL sample

vial. (e) Centrifuge: Hettich 30F benchtop centrifuge with swinging bucket rotor (1424A). (f) Vacuum oven: Fisher Scientific, Model 280. (g) Dry bath incubator: Fisher Scientific (11-718-8). (g) Data system: Hewlett-Packard ChemStation, Pharmacopeia PIE database. (h) Auto liquid handling system: Packard Multiprobe 204 DT.

Tag oxidation and silylation (in-vial oxidation)
To a vial with an insert containing a single bead is added 2 μL of 0.5 M aqueous cerium ammonium nitrate (CAN) solution (1:1 acetonitrile-water) and 10 μL of octane. The sample is incubated at 30 °C in a dry bath for 3 h. The octane layer (7 μL) is transferred to a GC autosampler vial. N-methyl-N-trimethylsilyltrifluoroacetamide (MSTFA; 2 μL) is added, then 1 μL of the solution is analyzed by GC/ECD.

GC/ECD analysis (HP 5890 – traditional method)

Column conditions

Column	15 m × 0.25 mm, 0.25 μm DB-1
Oven temperature	110 °C (1 min), 45 °C min^{-1} to 250 °C (2 min), 15 °C min^{-1} to 325 °C (10 min)
Flow rate	1 mL min^{-1} (helium constant flow rate)
Injection mode	Splitless
Inlet temperature	280 °C

ECD conditions

Makeup (N_2)	45 mL min^{-1}
Data rate	20 Hz

Single-bead LC/MS and statistical sampling

As discussed in the previous section, the encoding technology records the synthetic history of each individual library bead. Although library QC samples are useful in estimating the yield and purity of a library, it is highly desirable to have a much broader knowledge of the chemistry that took place on each bead during actual library construction. Considering the average yield on each bead is approximately 500 pmole/bead and a typical library consists of some 20 million beads due to high compound redundancy, LC/MS is currently the only option available to obtain structural information at the single bead level. This method provides the necessary sensitivity for analysis of single beads (detection limit at the sub-pM level), and at the very least, gives an accurate molecular weight of the compound of interest.

By combining LC/MS with tag decode analysis, the implied structure of the compound can be confirmed. Figure 7 illustrates this single bead LC/MS methodology. The compound from a single bead is eluted under conditions optimized for itþs particular library, and analyzed by LC/MS to provide the molecular weight of the compound. Subsequently, the tags from the same bead are oxidatively removed and analyzed, and a predicted structure and molecular weight are generated. Comparing the predicted molecular weight with the empirical value, yields a 'yes/no' answer, i.e., is the compound as predicted by the tags actually present or absent. The routine analysis of several hundred randomly selected beads from a library has proven to be a powerful statistical sampling tool to gather data on compound synthesis and the *combinatorial* success or failure of individual library synthons.

The generic requirements of an HPLC method for single bead LC/MS include: (1) a reversed-phase HPLC method with a gradient that guarantees all the compounds of a given library will be eluted from the column during the HPLC running time; (2) a low flow rate (200–500 μL min^{-1}) to optimize mass detector sensitivity of the PE Sciex 150 API (Turbo ion source); and (c) high-throughput as required by statistical sampling sizes of up to 1000 beads/library. To meet these requirements, a generic HPLC method was developed (Figure 8). A C8 column was chosen for its universal separation. Using the column in reversed-phase mode, most of the small molecules found in Pharmacopeia's libraries will be eluted from the column with satisfactory separation. The HPLC gradient may seem aggressive as it changes from 10–90% acetonitrile-water over a 2 min period. However, because of the void volume of the HPLC system, the actual gradient time is about 8.5 min. Figure 8 illustrates the actual HPLC gradient profile as monitored by UV at 215 nm. Since the system has ca. 1 mL of void volume, the actual gradient changes begin at 4.5 min, reaching 90% acetonitrile in 8.5 min, and holding at 90% for 2 min. It takes the system another 5 min to refresh itself (10% acetonitrile); therefore,

the current HPLC run time is ca. 15 min. Obviously, the run time can be shorted to 7–8 min by increasing the flow rate to 500 μL min^{-1}. Since the mass detector affords the best sensitivity at a flow rate of 200 μL min^{-1}, this flow rate is preferred for statistical sampling.

The parameters for the mass spectrometer are optimized after the LC method is established (Figure 9). Mass ranging from 300 to 800 amu covers 99% of compounds from the Pharmacopeia libraries. The scan rate is 2.5 s for the mass range, yielding ca. 12 data points for an HPLC peak width of 30 s, sufficient for quantitation. The voltages for the ion focus lenses are optimized according to the individual quality control (QC) compound from each library.

After all the parameters are chosen, the LC/MS experiment is performed for the QC compounds from the library. Figure 8 gives an example for a QC compound with molecular weight of 445.9 amu. The LC/MS chromatogram clearly shows a symmetrical peak and the correct molecular weight is assigned. Analysis of the QC compounds from the library and observation of good peak shapes for all the QC compounds, validates the LC/MS method. Lastly, the limit of detection for the LC/MS method needs to be established. The LC/MS experiments are carried out for the set of QC samples with a gradient concentration. The detection limit is established at the point where the signal to noise ratio is about 3 to 1. Typically the detection limit is 0.1 pmole. Some variability in the amount of compound on a bead is expected, due to differences in bead size (200$\pm$10 pmole). The detection limit at 0.1 pmol is sufficient for confirming the absence of compound when its molecular ion is not found as predicted from the decode.

The overall procedure for single bead LC/MS analysis proceeds through four steps. First, the elution of compound from a random library bead is carried out in parallel with a single QC bead. The QC bead provides information about the success of the elution condition. For each library to be studied, 5–10 beads from each sub library across the whole library are eluted. Second, tag decoding and analysis is completed. Third, LC/MS analysis is performed. The fourth step is the comparison of the molecular weight from both decode and LC/MS to ascertain the presence or absence of compound, i. e., a 'yes/no' answer [7,8].

An example of the single bead LC/MS analysis procedure is given by the analysis of library bead. After photoelution, the tags from the library bead are analyzed by GC/ECD. A predicted structure is generated as is the corresponding molecular weight, 542.2 amu. The LC/MS experiment is performed on the bead eluent. The total ion current (TIC) chromatogram shows two major peaks with retention times at 8.62 and 10.87 min respectively. Extraction of a selected ion chromatogram for the expected mass predicted by the tag decode (543.2$\pm$1 amu, 543 = M + H^+) yields the extracted ion current (XIC) chromatogram. The XIC chromatogram contains a major peak at 8.62 min. The mass spectrum of this peak reveals a major ion with molecular weight 542.2 as predicted by the decode analysis. Thus, the expected product is found and a 'yes' answer for that library bead is established. This type of comparison is carried out for all the library beads analyzed. A high throughput data analysis program (customized software) automatically performs the comparison and provides yes/no answers.

Upon completion of the LC/MS analysis, a report is also generated automatically. Table 1 describes the 'yes/no' answers using the intensity of the peak at the XIC chromatogram for the expected ion. For example, the instrumental noise level is $1 \times E + 05$. If the found intensity for a given ion is less than $3 \times E + 05$ (S/N < 3), the ion is not found. On the other hand, if the expected ion has an intensity higher than $3 \times E + 05$, it is treated as found. Thus, all the expected ions except the 11th entry (E08-42, MW 616.4) listed in Table 1 are found, indicating that the proposed synthetic chemistry on those beads actually took place. Table 1 also gives the answers for whether the major peak in the TIC is from the expected ion, providing information about impurities from the bead or the elution plates. For example, the first entry (D06–40) contains a major peak from an impurity as opposed to the expected ion. This impurity peak was traced to the polyethylene glycol on the bead, with a molecular weight of 573.3 amu and a retention time at 7.6 min.

Since the tags record the synthetic history on the bead, further comparison of the 'yes/no' answers with the synthetic steps on the bead can be done. This comparison affords an in depth knowledge of the overall library fidelity and the performance of individual synthons [7].

Summary

Assessing the quality of encoded libraries is a challenge due to the sheer size of the libraries (>50,000 members), high redundancy (200 fold, 10–15 million beads), small amounts of released compounds (ca.

500 pM/bead), and trace levels of tags. In our laboratories, we rely on the rigorous analysis of a set of representative QC samples to establish a library's average yield and purity. A statistical sampling protocol, based on the combined application of tag decoding and single-bead LC/MS, is employed to ascertain library fidelity and the performance of individual synthons. Application of the HP6890 has increased tag sensitivity and accelerated the speed at which decodes can be obtained. We are currently exploring the utility of ultrafine condensation particle counters, chemiluminescent nitrogen-specific detection and GC-based atomic emission detection to permit an even greater degree of library characterization.

References

1. (a) Baldwin, J.J. and Horlbeck, E.G., U.S. Patent 5 663 046, 1997.
   (b) Baldwin, J.J., Burbaum, J.J., Henderson, I. and Ohlmeyer, M.H.J., Synthesis of a small molecule combinatorial library encoded with molecular tags. J. Am. Chem. Soc., 117 (1995) 5588–5589.
   (c) Chabala, J.C., Baldwin, J.J., Burbaum, J.J., Chelsky, D., Dillard, L., Henderson, I., Li, G., Ohlmeyer, M.H.J. and Randle, T.L., Binary encoded small-molecule libraries in drug discovery and optimization. Perspect. Drug Discovery Des., 2 (1995) 305–318.
   (d) Baldwin, J.J., Burbaum, J.J., Chelsky, D., Dillard, L.W., Henderson, I., Li, G., Ohlmeyer, M.H.J., Randle, T.L. and Reader, J.C., Combinatorial libraries encoded with electrophoric tags. Eur. J. Med. Chem., 30 (1995) 349s–58s.
   (e) Burbaum, J.J., Ohlmeyer, M.H.J., Reader, J.C., Henderson, I., Dillard, Lawrence W., Li, G., Randle, T.L., Sigal, N.H., Chelsky, D. and Baldwin, J.J., A paradigm for drug discovery employing encoded combinatorial libraries. Proc. Natl. Acad. Sci. U.S.A., (1995) 6027–6031.
   (f) Chabala, J.C., Solid-phase combinatorial chemistry and novel tagging methods for identifying leads. Curr. Opin. Biotechnol., 6 (1995) 632–639.
   (g) Baldwin, J.J. and Henderson, I., Encoded split synthesis and solid-phase synthetic methodology. Med. Res. Rev., 16 (1996) 391–405.
   (h) Baldwin, J.J., Design, synthesis and use of binary encoded synthetic chemical libraries. Mol. Diversity, 2 (1996) 81–88.
   (i) Baldwin, J.J., Kirk, G.L., Dillard, L.W. and Connelly, J.A., Design, synthesis, and preparation of an encoded small molecular library for a wide variety of 96-well based bioassays. Methods Mol. Cellular Biol., 6 (1996) 74–88.
   (j) Baldwin, J.J. and Dolle, R.E., Deconvolution methods in solid-phase synthesis. Annu. Rep. Comb. Chem. Mol. Diversity, 1 (1997) 287–297.
   (k) Rokosz, L., Sigal, N.H., Burbaum, J.J., Chelsky, D., Dillard, L.W., Henderson, I., Li, G., Ohlmeyer, M.H.J., Randle, T.L. and Reader, J.C., Synthesis of binary encoded small molecule libraries and their application to drug discovery and optimization. *Innovation Perspect. Solid Phase Synth. Comb. Libr., Collect. Pap., Int. Symp., 4th Meeting Date 1995*, Epton, R. (Ed.) Mayflower Scientific: Birmingham, U.K., 1996, pp. 101–106.
   (l) Baldwin, J.J. and Henderson, I., Synthesis of encoded small molecule combinatorial libraries via ECLiPS. In *High Throughput Screening*, Gordon, E.M. and Kerwin Jr., J.F., (Eds.) Dekker, NY, 1997, pp. 167–190.
   (m) Baldwin, J.J., Small molecule libraries: Overview of issues and strategies in library design. In *Comb. Chem. Mol. Diversity Drug Discovery*, Gordon, E.M. and Kerwin Jr., J.F. (Eds.), Wiley-Liss, NY, 1998, pp. 181–188.
2. (a) Ohlmeyer. M.H.J., Swanson, R.N., Dillard, L.W., Reader, J.C., Asouline, G., Kobayashi, R., Wigler, M. and Still, W.C., Complex synthetic chemical libraries indexed with molecular tags. Proc. Natl. Acad. Sci. U.S.A., 90 (1993) 10922–10926.
   (b) Nestler, H.P., Bartlett, P.A. and Still, W. Clark, A General Method for Molecular Tagging of Encoded Combinatorial Chemistry Libraries. J. Org. Chem., 59 (1994) 4723–4724.
3. (a) Appell, K.C., Chung, T.D.Y., Ohlmeyer, M.J.H., Sigal, N.H., Baldwin, J.J. and Chelsky, D., Biological screening of a large combinatorial library. J. Biomol. Screening, 1 (1996) 27–31.
   (b) Appell, K.C., Chung, T.D.Y., Solly, K.J., Chelsky, D., Biological characterization of neurokinin antagonists discovered through screening of a combinatorial library. J. Biomol. Screening, 3 (1998) 19–27.
   (c) Carroll, C.D. and Orlowski, M., Screening aspartyl proteases with combinatorial libraries. Adv. Exp. Med. Biol., 436 (1998) 375–380.
   (d) Sigal, N.H. and Chelsky, D., Approaches and technologies for screening large combinatorial libraries. In *Comb. Chem. Mol. Diversity Drug Discovery*, Gordon, E.M. and Kerwin, J.F. (Eds.), Wiley-Liss, NY, 1998, pp. 181–188.
4. (a) Carroll, C.D., Patel, H., Johnson, T.O., Guo, T., Orlowski, M., He, Z.-M., Cavallaro, C.L., Guo, J., Oksman, A., Gluzman, I.Y., Connelly, J., Chelsky, D., Goldbert, D.E. and Dolle, R.E., Identification of potent inhibitors of plasmodium falciparum plasmepsin II from an encoded statine combinatorial library. Bioorg. Med. Chem. Lett., 8 (1998) 2315–2320.
   (b) Carroll, C.D., Johnson, T.O., Tao, Shewei, Lauri, G., Orlowski, M., Gluzman, I.Y., Goldbert, D.E. and Dolle, R.E., Evaluation of a structure-based statine cyclic diamino amide encoded combinatorial library against plasmepsin II and cathepsin D. Bioorg. Med. Chem. Lett., 8 (1998) 3203–3206.
5. (a) Tan, D.S., Foley, M.A., Shair, M.D. and Schreiber S.L., Stereoselective synthesis of over two million compounds having structural features both reminicient of natural products and compatible with minaturized cell-based assays. J. Am. Chem. Soc., 120 (1998) 8565–8566.
   (b) Kapoor, T.M., Andreotti, A.H. and Schreibert, S.L., Exploring the specificity of two homologous SH3 domains using structure-based, split-pool synthesis and affinity-based selection. J. Am. Chem. Soc., 120 (1998) 23–29.
   (c) Morken, J.P., Kapoor, T.M., Feng, S., Shirai, F. and Schreiber, S.L., Exploring the leucine-proline binding pocket of the Src SH3 domain using structure-based, split-pool synthesis and affinity-based selection. J. Am. Chem. Soc., 120 (1998) 30–36.
6. Parity tags are generally employed together with the synthon tag set to maintain the total tag count at an even number for any given set of encoded synthons. The parity tags assist in resolving the occasional tag ambiguity which arises from the GC auto scoring function.
7. The number of random beads that must be sampled from

a given library to give a statistically meaningful result is equivalent to 10× the largest combined synthon set. For an account of the statistical sampling methodology see: Dolle, R.E., Guo, J., O'Brien, L., Jin, Y., Piznik, M., Bowman, K.L., Li, W., Egan, W.J., Cavallaro, C.L., Roughton, A.L., Zhao, Q., Reader, J.C., Orlowski, M., Jacob-Samuel, B. and DiIanni Carroll, C., A statistical-based approach to assessing the fidelity of combinatorial libraries encoded with electrophoric molecular tags. Development and application of tag decode-assisted single bead LC/MS analysis. J. Combin. Chem., 2 (2000) 716–731.