JOURNAL OF combinatorial CHEMISTRY

Subscriber access provided by American Chemical Society

Article

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

Roland E. Dolle, Joan Guo, Linda O'Brien, Ying Jin, Mark Piznik, Kevin J. Bowman, Wenni Li, William J. Egan, Cullen L. Cavallaro, Andrew L. Roughton, Qian Zhao, John C. Reader, Marc Orlowski, Biji Jacob-Samuel, and Carolyn DiIanni Carroll

J. Comb. Chem., **2000**, 2 (6), 716-731• DOI: 10.1021/cc000052k • Publication Date (Web): 21 October 2000

Downloaded from http://pubs.acs.org on March 20, 2009

More About This Article

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

- Supporting Information
- 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

ACS Publications
High quality. High impact.

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

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

Roland E. Dolle,*[,†,‡] Joan Guo,[‡] Linda O'Brien,[‡] Ying Jin,[‡] Mark Piznik,[‡] Kevin J. Bowman,[‡] Wenni Li,[‡] William J. Egan,[§] Cullen L. Cavallaro,[†] Andrew L. Roughton,[†] Qian Zhao,[†] John C. Reader,[†] Marc Orlowski,[||] Biji Jacob-Samuel,[||] and Carolyn DiIanni Carroll[||]

Departments of Chemistry, Analytical Chemistry, Biology, and the Center For Informatics and Drug Discovery, Pharmacopeia, Inc., P.O. Box 5350, Princeton, New Jersey 08543

Received June 26, 2000

A statistical sampling protocol is described to assess the fidelity of libraries encoded with molecular tags. The methodology, termed library QA, is based on the combined application of tag decode analysis and single bead LC/MS. The physical existence of library compounds eluted from beads is established by comparing the molecular weight predicted by tag decode with empirical measurement. The goal of sampling is to provide information on overall library fidelity and an indication of the performance of individual library synthons. The minimal sampling size n for library QA is $10\times$ the largest synthon set. Data are reported as proportion (p) $\pm$ lower and upper boundary (*lb-ub*) computed at the 95% confidence level ($\alpha = 0.05$). As a practical demonstration, library QA was performed on a 25 200-member library of statine amides (size $= 40 \times 63 \times 10$). Sampling was conducted three times at $n \sim 630$ beads per run for a total of 1902 beads. The overall proportions found for the three runs were consistent with one another: $p = 84.4\%$, *lb-ub* $= 81.5-87.2\%$; $p = 83.1\%$, *lb-ub* $= 80.2-85.95$; and $p = 84.5\%$, *lb-ub* $= 81.8-87.3\%$, suggesting the true value of p is close to 84% compound confirmation. The performance p_i of individual synthons was also computed. Corroboration of QA data with biological screening results obtained from assaying the library against cathepsin D and plasmepsin II is discussed.

Encoded combinatorial libraries of small molecules are a valuable resource for the discovery of biologically active agents.[1] Originally conceived to simplify the process of deconvolution and compound identification in split-pool libraries,[2] a variety of strategies have now been described to encode libraries.[3] One strategy, successfully pioneered at Pharmacopeia, is a binary encoding protocol employing electrophoric molecular tags (ECLiPS technology).[3a−c,4] In this protocol, sets of synthons are serially combined through split-pool or direct divide[5] synthesis in tandem with the incorporation of binary sets of electrophoric tags on solid support. Each bead in an encoded library contains a compound, whose synthesis history is recorded by a unique set of attached tags. Because orthogonal reactive linkers are used in the construction of the library, compound and associated tags can be independently released allowing off-bead assays. The identity of any given library member is readily inferred through the process of decoding, i.e., electron capture detection/gas chromatography (GC/ECD) analysis of detached tags.

The construction of molecularly encoded combinatorial libraries demands that many runs through the complete reaction sequence be performed to optimize reaction conditions and to ensure synthon compatibility.[6] During this development phase, the gravimetric yield and purity (HPLC) of many putative library members are determined, and several library quality control (QC) compounds are prepared. Library QC compounds are rigorously analyzed and used to estimate the bead yield and purity of the completed library. Extensive synthon profiling, careful reaction optimization, and rigorous analysis of library QC compounds are necessary to complete a successful library synthesis. However, it is highly desirable to have a much broader knowledge of the chemistry that occurs during actual library construction and assurance that the compounds eluted from the beads are physically present in the wells of assay plates.[7] With this goal of assessing the overall quality of an encoded combinatorial library, a statistical sampling procedure was devised which combines the application of single bead LC/MS analysis with tag decode analysis to confirm the existence of putative library compounds.[8] The statistical-based, tag decode-assisted single bead LC/MS analysis is conveniently termed "library quality assurance" (library QA). The development and validation of library QA as a useful qualitative measure of library fidelity is demonstrated here in the evaluation of a 25 200-member library of statine amides.

Tag Decode-Assisted Single Bead LC/MS Analysis (Library QA)

Libraries prepared by parallel synthesis typically provide a manageable number of compounds (<5000) in sufficient quantity (>0.1 mg) to allow the structure identity, purity,

* Corresponding author. E-mail: roland@pharmacop.com.
† Department of Chemistry.
‡ Department of Analytical Chemistry.
|| Department of Biology.
§ Center For Informatics and Drug Discovery.

10.1021/cc000052k CCC: $19.00 © 2000 American Chemical Society
Published on Web 10/21/2000

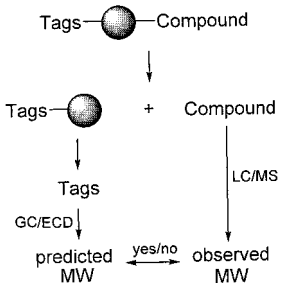


Figure 1. Tag decode-assisted single bead LC/MS.

and yield for all members (or a statistically relevant number) to be determined.[9] This may be accomplished by employing routine HPLC analysis coupled singularly or in combination with UV,[9,10] ELSD,[11] CLND,[12] and MS[13] detection, and more recently, high-throughput NMR analysis.[14] Automated HPLC may also be used to optionally purify every compound in a library prior to biological testing.[15] These approaches to library characterization and compound purification are generally not viable with encoded combinatorial libraries. Specifically, the sheer magnitude of the electrophorically encoded libraries, where 200−300 copies of a library are routinely prepared (e.g., 10−15 million beads per 50 000-member library), makes it impossible to classically characterize every library member. The average amount of compound present on the bead is approximately 200 pm/bead; hence, high-throughput compound analysis is in essence restricted to the more sensitive analytical technique of LC/MS.

One approach to characterizing the large encoded libraries is through library QA analysis. By combining LC/MS analysis with tag decode analysis, the physical existence of the structure of any given library compound can be confirmed. Figure 1 illustrates the library QA process. The compound from a single bead is eluted under conditions optimized for its particular library. The tags from the same bead are oxidatively removed and analyzed by ECD/GC, and a predicted molecular weight for the decoded compound generated. The compound (bead eluent) is then analyzed by LC/MS, applying a mass filter, to determine its molecular weight. Comparing the molecular weight predicted by the tag decode with the empirical value yields a "yes/no" answer establishing the presence or absence of the inferred compound. Statistical sampling of several hundred randomly selected beads from a library provides information on library fidelity and an indication of the performance of individual library synthons.

Statistical Considerations

The goal of sampling an encoded combinatorial library is both to determine the overall success of the solid-phase synthesis and to assess the performance of each synthon. Tag decodes reveal which compounds (synthons) are supposed to be on each bead, and LC/MS analysis reveals whether they are physically present. It is impractical to analyze all the compounds in a encoded library where the total bead count is in excess of one million; therefore, statistical sampling techniques are required.

Statistical sampling involves taking some portion of a population and estimating a property of that population (and the uncertainty in the estimate) using only the portion of the population analyzed.[16] For either an entire library or an individual synthon, the numerical result of the tag decode-assisted LC/MS analysis is given as the proportion p, where p is the number of compounds positively identified by LC/MS (x) divided by the sample size n. Lower and upper confidence intervals for p are also calculated. The confidence level indicates the probability of finding the true value of p in a given interval, e.g., $\alpha = 0.05$ represents the 95% confidence level and indicates that the true value of p is inside the specified interval with 95% probability. A narrower confidence interval therefore means that the estimate of p is more precise.

The calculation of confidence intervals for a proportion is an interesting task. In library QA analysis, each bead is analyzed only once (sampling without replacement) and the proportion p is calculated, which means that the sampling distribution follows a hypergeometric distribution.[16,17] The hypergeometric distribution may be well approximated with the more tractable binomial distribution, providing the population size N (library size) is much larger than n. The difference between the variance of the hypergeometric distribution and the binomial distribution is simply a multiplicative factor: $(N - n)/(N - 1)$, called the finite population correction factor. A review of Pharmacopeia libraries found that for both overall library analysis and individual synthon profiling, N was suffciently larger than n to make the finite population correction factor >0.989 on average; therefore, the approximation is considered valid for this problem.

Both the hypergeometric and binary distributions are discrete, i.e., if the sample size n is 20, x may only have the integer values 0, 1, 2, ..., 20. For a moderately large n, with p near $1/2$, the sampling distribution "smears" and is effectively no longer discrete, is symmetric, and thus can be well approximated by the normal distribution.[17] However, for moderately large n with p near 0 or 1, the binomial distribution is not symmetric and the Poisson distribution is a more accurate approximation, instead of the normal distribution.[17] Furthermore, confidence intervals for a proportion are the widest when p is near $1/2$ and the narrowest when p is near 0 or 1. Because the sampling of a combinatorial library for library QA analysis involves both large n (the entire sample of compounds) and small n (individual synthons), and because modern computational resources are sufficient, the binomial distribution was chosen to avoid further approximations.

To compute confidence intervals for the proportion p, the classical method of Clopper and Pearson[18] was modified. While this method does produce slightly conservative confidence intervals, i.e., slightly wider than they should be, it was believed that a conservative approach was most appropriate for this problem. The modifications were as follows: (1) for small n (<50), the binomial distribution was interpolated using a cubic-spline to determine the confidence interval endpoints and (2) for large n, the binomial distribution was used without interpolation to determine confidence interval endpoints. The binomial distribution is not symmetric (like the normal distribution), and the confidence intervals

are reported as a lower (*lb*) and upper (*ub*) bounds on the proportion *p*. Because the binomial distribution is discrete, some form of interpolation is necessary when computing the exact confidence intervals for a proportion derived from a small sample size. This is due to the fact that the distribution is not defined except at the discrete fractions of the sample size, while the confidence interval exists in a continuous sense because it is describing the population parameter (the proportion). For example, if $n = 10$ and $p = 0.5$, and a 95% confidence interval is desired, $\alpha = 0.05$, and one would want the *lb* associated with a cumulative probability (from the binomial *cdf*) of 0.025 and the *ub* associated with a cumulative probability of 0.975. However, for $n = 10$, only the discrete values 0, 1, 2, ..., 10 exist, with corresponding cumulative probabilities: 0.0010, 0.0107, 0.0547, 0.1719, 0.3770, 0.6230, 0.8281, 0.9453, 0.9893, 0.9990, 1.0000. No defined point on the distribution has cumulative probabilities of 0.025 or 0.975. Therefore, an exact 95% confidence interval cannot be computed, unless some form of interpolation is used. For a large sample size, there are enough discrete values defined that the $1 - \alpha$ criterion for the width of the confidence interval is usually easily met without interpolation.

The minimum sample size required to assess the fidelity of a combinatorial library depends more on the precision required to estimate *p* when assessing an individual synthon's performance than on the precision required for assessing overall library fidelity. For example, if $n = 300$ for an entire library, then the 95% confidence interval for $p = 0.5$ is 0.442 to 0.558, and if $p = 0.9$, the 95% confidence interval narrows to 0.863 to 0.934.

However, the sample size for the entire library (*n*) is not the sample size for an individual set of synthons. Consider library **1**, discussed in more detail below: there are three synthetic steps, and each step has a different sized synthon set ($40 \times 63 \times 10$). The notation $n_{i,j}$ to is used to identify the synthetic step (*i*) and the size of the synthon set (*j*) for the sample size of an individual set of synthons. For the final sublibrary of **1**, which contains 10 synthons, those 10 synthons will each be sampled *n*/10 times exactly. This occurs because the sample is drawn only from the final sublibrary; sampling is not performed at each step of the library synthesis, only at the end. So, if $n = 300$, the $n_{3,10} = 30$ for the 10 synthons in step 3. But for the second synthetic step, there are 63 synthons, and they are each sampled an *average* of only 4.76 times ($n_{2,63} \approx 4.76$). This "average effect" occurs because beads are pooled at the end of the first two synthetic steps, meaning that any random sample will not contain an exact and unchanging number of each of the 63 synthons used in the second synthetic step, but that the average sample size for those 63 synthons will be 4.76. This problem does not exist for the final 10 sublibraries because they are not pooled. Thus, the synthons in the largest sized synthon set will have the smallest $n_{i,j}$ and the number sampled of each of those synthons may be less than that required for a desired precision. Clearly, a minimum sample size must take into account how small $n_{i,j}$ can be and still provide a useful answer, while dealing with the fact that the overall *n* must be limited; otherwise the throughput of tag decode and LC/MS analysis will be overwhelmed.

Table 1. Computed Proportion (*p*) and Boundary Limits (*lb*-*ub*) for Representative Sample Sizes with High Positive Identification

sample size (*n*)	positive ID	*p* (%)	*lb* (%)	*ub* (%)
2	1	50.0	0.0	84.7
4	3	75.0	24.8	100
6	5	83.3	47.9	100
8	7	87.5	60.2	100
10	9	90.0	67.8	100
15	14	93.3	78.3	100
25	24	96.0	86.8	100
50	49	98.0	93.4	100
100	99	99.0	96.7	100
250	245	98.0	95.9	99.7

Table 2. Computed Boundary Limits (*lb*-*ub*) for Representative Sample Sizes When Proportion (*p*) of Positive Identification = 50.0%

sample size (*n*)	positive ID	*p* (%)	*lb* (%)	*ub* (%)
4	2	50.0	0.0	77.1
5	3	50.0	4.4	95.6
8	4	50.0	10.2	84.8
10	5	50.0	14.8	85.2
20	10	50.0	25.9	74.2
50	25	50.0	35.2	64.8
100	50	50.0	39.7	60.3
300	150	50.0	44.2	55.8
600	300	50.0	46.0	54.0
500	250	50.0	45.6	54.4

One reasonable solution to the problem of determining the proper balance between *n* and $n_{i,j}$ is to set the minimum average $n_{i,j}$ to 10× the number of synthons in the largest synthon set.[8b] Thus, in the case of library **1**, the largest synthon set has 63 synthons and $n_{2,63} \approx 10$, so the overall sample size of beads must equal 630. The other two synthon sets will then have larger sample sizes: $n_{1,40} \approx 16$ (an average, because this set was pooled for further reactions) and $n_{3,10} = 63$ (exact, because no pooling was performed). The imprecision of the estimate of *p* obtained in the worst case scenario ($n_{i,j} \approx 10$), determined by the width of the confidence intervals, is acceptable to yield a useful assessment of an individual synthon's performance if *p* is near 0 or 1 (see Table 1).

Table 1 gives the computed values for representative proportions, confidence intervals, and the corresponding observed confidence levels for various *n* (calculated at $\alpha = 0.05$). At large values of *n*, e.g., 500 (the typical composite QA sampling size), the confidence intervals are narrow. For example, if 500 decoded beads yielded 450 molecular weight confirmations, then $p = 0.900$ and the confidence interval is 0.872 to 0.926 with observed confidence of 95.6%. From this hypothetical QA result, it would be concluded, with a high degree of certainty, that true *p* is near 90% (in the range of 87.2−92.6%), and that the library fidelity is quite good.

When considering the smaller n_i values of sublibraries and individual synthons, the confidence boundaries are the widest, and thus the answers most uncertain, when *n* is small and *p* is near 0.5. For example, if $n_i = 10$, and $p = 0.5$ (50%), then the confidence interval is 0.148−0.852 at the 95% confidence level. In such a case, statistical analysis suggests

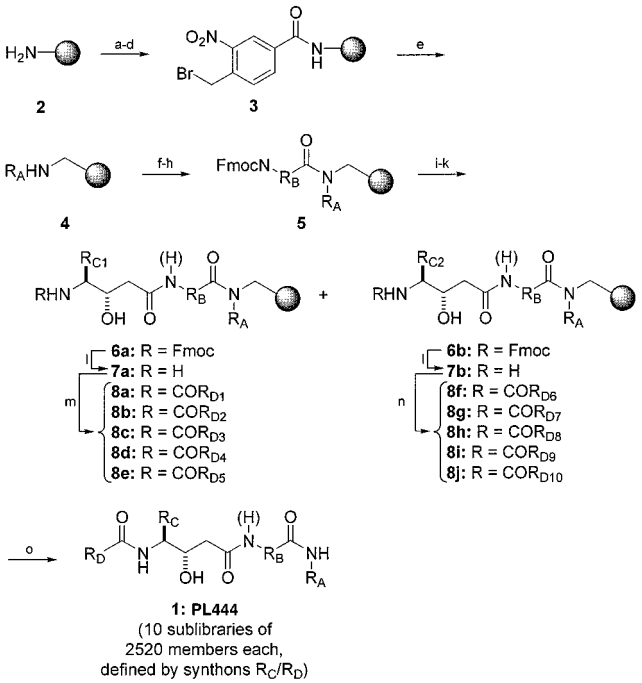
that it is difficult to reach any meaningful conclusion about the fidelity of a sublibrary or the performance of a synthon. As p approaches 0 or 1, the confidence interval narrows, and thus, the actual proportion of compounds in a sublibrary or performance of individual synthons may be estimated more precisely. Again referring to sample size $n = 10$, if $p = 0.90$, then the confidence interval is 0.678−1.00 at the 95% confidence level. Although still somewhat uncertain, the estimate that the true value of p in this instance lies approximately in the interval 70−100% and indicates that the fidelity of this particular synthon (or sublibrary) was reasonably good. Considering the extensive solid-phase optimization that proceeds library construction, p will often be near 0.9 and so the confidence intervals will be at their narrowest extreme.

Library QA provides a very good estimate of the quality of a combinatorial library on two levels. First, a highly precise measure of the proportion of compounds successfully synthesized in a library is calculated. The precision in the estimate of library fidelity is high because the sample size is on the order of several hundred. Library QA permits the chemist to identify and readily distinguish between poor, mediocre, and good quality combinatorial libraries. Second, statistical analysis of the sampling data can assist the chemist in ascertaining the performance of *individual* synthons in a library, with the proviso that the synthon sample size, $n_{i,j}$, must be sufficiently large (>10). Library QA provides no information regarding the *combinatorial* synthesis success of synthons or intermediates, as the required n would exceed a practical limit. Rather, the analysis provides information on average proportions of observed compounds and synthons in a library.

Synthesis of the Encoded Statine Amide Library (1)

The encoded solid-phase synthesis of the statine amide library **1**, is depicted in Scheme 1. Commercially available TentaGel S-NH_2 resin (Rapp Polymere) **2** was derivatized with bis-Fmoc lysine and distributed into 40 reaction vessels. The resin batches were encoded using a set of six electrophoric tags.[4a] The Fmoc protecting group was removed, and the resin batches were acylated with the photolabile linker, 4-bromomethyl-3-nitro-benzoic acid (Aldrich) to give the corresponding encoded resins **3**. The resins were then reacted with one of 40 $\mathbf{R_A}$ amines (10 equiv of $\mathbf{R_A}$ amine as a 0.5 M solution in THF; Figure 2) to furnish the amino resins **4**. Resin batches **4** were combined, mixed, and distributed into 63 reaction vessels. Each resin batch was acylated with one of 63 $\mathbf{R_B}$ Fmoc-protected amino acids (3 equiv of $\mathbf{R_B}$ amino acid, Figure 2) and subjected to binary encoding with a second set of six electrophoric tags. Encoded resins **5** so obtained were combined, mixed, and divided into two portions. Each portion of resin was first treated with 20% piperidine in DMF to remove the Fmoc-protecting group and then acylated with either Fmoc-statine **C1** (**5** to **6a**) or Fmoc-phenylalanine statine **C2** (**5** to **6b**). Fmoc deprotection in **6a**,**b** furnished the corresponding amino statine resins **7a**,**b**. Amine resins **7a**,**b** were not combined, but rather each resin batch was apportioned into five separate reaction vessels, ultimately creating 10 sublibraries. Resin batches derived

Scheme 1. Synthesis of the Statine Library **PL444**[a]



[a] Reagents and conditions: (a) TentaGel resin, 3 equiv each Bis-Fmoc-Lys, HOBT, 6 equiv DIC, CH_2Cl_2; (b) apportion into 40 reaction vessels and encode using six molecular tags; (c) 20% piperidine−DMF, 1 h; (d) 3 equiv each 4-bromomethyl-3-nitrobenzoic acid, HOBT, 6 equiv DIC, CH_2Cl_2; (e) one of 40 R_A amines (Figure 2): 10 equiv amine, THF, 12 h; (f) combine and split into 63 reaction vessels; (g) one of 63 R_B Fmoc amino acids (Figure 2): 3 equiv each Fmoc amino acid and HATU, 6 equiv iPr$_2$EtN, DMF, 6 h; (h) encode using seven molecular tags; (i) combine and split into two reaction vessels A and B; (j) 20% piperidine−DMF, 1.5 h; (k) 3 equiv Fmoc-statine (**C1**/reaction vessel A; **C2**/reaction vessel B), 3 equiv iPr$_2$EtN, DMF, 6 h; (l) split into five reaction vessels, then 20% piperidine−DMF, 1.5 h; (m) one of five R_D carboxylic acids (**D1**−**D5**; Figure 2), 3 equiv each RCOOH and HATU, 6 equiv iPr$_2$EtN, DMF, 6 h; (n) one of five R_D carboxylic acids (**D6**−**D10**; Figure 2), 3 equiv each RCOOH and HATU, 6 equiv iPr$_2$EtN, DMF, 6 h; (o) 365 nm, EtOH−water, 50 °C, 30 min.

from **7a** were acylated with one of the carboxylic acid subunits $\mathbf{R_{D1-5}}$, while resins **7b** were acylated with one of the carboxylic acid subunits $\mathbf{R_{D6-10}}$. Finally, the resins were exposed to a cocktail of TFA-phenol-EDT-water (1:1:1) to remove the acid-labile protecting groups in the amino acid subunits $\mathbf{R_B}$, generating sublibraries **8a**−**j**. Irradiation of the fully encoded library at 365 nm furnished library **1**. Library **1**, prepared in 200-fold redundancy, was thus comprised of 10 sublibraries defined by $\mathbf{R_C}$/$\mathbf{R_D}$, for a total of 25 200 unique compounds.

Library QA Protocol

Statine library **1**, synthesized via two combinatorial steps and a two-step N-derivatization sequence, is composed of 10 sublibraries: $40(\mathbf{R_A}) \times 63(\mathbf{R_B}) \times 10(\mathbf{R_C}/\mathbf{R_D}) = 2520$ unique compounds per sublibrary. The target sample size n for the QA analysis is thus 630 beads, corresponding to 10× the largest synthon set—$\mathbf{R_B}$. In practice, 60−65 beads per sublibrary were arrayed as single beads in 96-well filter bottom plates. The beads were suspended in a solution of aqueous ethanol and irradiated at 365 nm for 30 min according to a previously determined elution protocol.[6] Bead eluents, containing the detached compounds, were filtered into derivative plates and dried in vacuo. The beads were

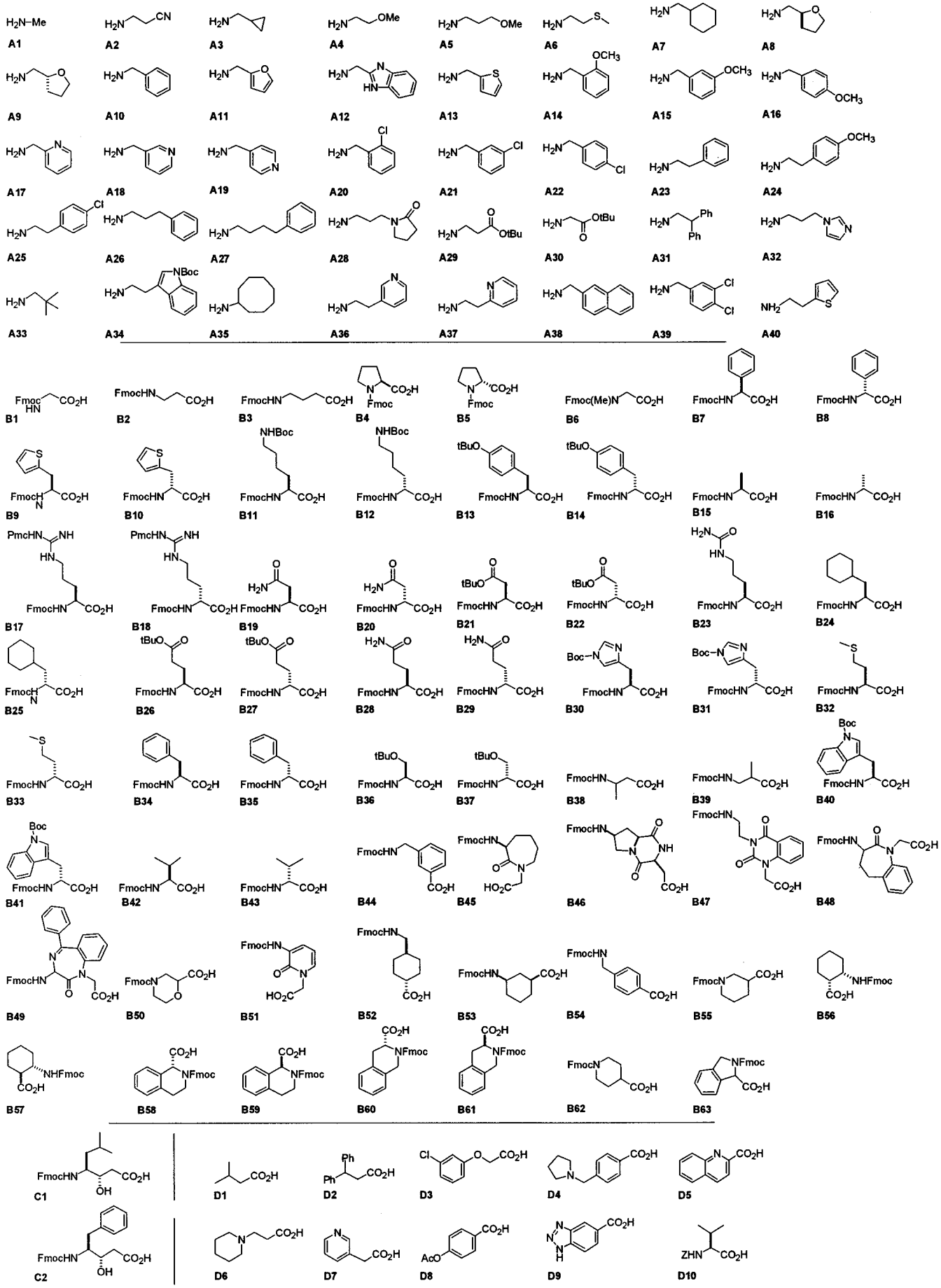


Figure 2. Basis set for library **1**. (Synthons **B38**, **B39**, **B48**−**53**, **B55**−**57**, and **B63** are racemic. All other synthons possessing one or more chiral centers are optically active with stereochemistry as shown.)

decoded, generating a structure and a predicted molecular weight for each compound. The dried compounds in the derivative plates were redissolved in a small volume of CH_3-CN−water (4:1) and analyzed by LC/MS.

Mass spectral analysis was carried out in the positive ion mode only.[19] The total ion current (TIC) chromatogram for each sample was first obtained. A selected ion chromatogram was then extracted for the expected mass (predicted by the tag decode), yielding an extracted ion current (XIC) chromatogram.[20] If the XIC chromatogram revealed the expected mass $(M+H)^+$, then a "found (F)" answer for that library bead was secured. If the expected mass ion was not detected, then a "not found (NF)" was recorded.[21] A custom data analysis program automatically performed the molecular weight comparisons.

Biological Screening

Statines are a well-known class of transition-state isosteres displaying inhibitory activity against aspartic acid proteases.[22] Library **1** was screened against two aspartic acids proteases, human cathepsin D (cat D)[23] and malarial plasmepsin II (plm II).[24−26] Solution-based assays for each enzyme were carried out using fluorescent energy transfer substrates. Screening was conducted via a two-part protocol. First, a survey screen was conducted in which 1 library equivalent (ca. 25 000 beads) were screened at ca. 30 beads per well to identify the most active sublibrary. This was followed by screening 2 library equivalents of the most active sublibrary at the single bead level (ca. 5000 beads).

Results and Discussion

Library QA Results. Three QA analyses were performed on library **1**.[27] Statistical data is presented for all three composite analyses (Figure 3A), one representative sublibrary, A, and B synthon analysis (QA run-3; Figures 3B, 4, and 5). Comparative answers are color coded: found (F; green) and not found (NF; red). The proportion p of confirmed compounds was computed by dividing the number of compounds found by the total number of beads analyzed (F/n; $\alpha = 0.05$). Values for p and their lower-upper boundaries (*lb-ub*) are listed in Tables 3−5.

The sample size n for the first run (QA run-1) was 610 beads (Figure 3A, Table 3). The total number of found answers F ($(M+H)^+$) was 515; $p = 84.4\%$. The lower boundary (*lb*) for p was 81.5%, and the upper boundary (*ub*) was 87.2%. The large value of n and narrow boundary for p suggests the true value of p lies between ca. 81−87%. Thus, one would expect values for p to be found in this range 95 out of 100 times the QA analysis was performed (95% confidence interval). In support of this assumption, two additional QA analyses were conducted. The sample sizes n for each run were as follows: QA run-2, $n = 645$ beads and QA run-3, $n = 647$ beads. The found rates for these two runs were 536 beads and 547 beads, respectively, from which the computed proportions are $p = 80.2\%$, *lb-ub* = 80.2−85.9% and $p = 84.5\%$, *lb-ub* = 81.8−87.3%. The overlapping boundaries for p obtained for the three QA analyses are in excellent agreement with one another, consistent with statistical expectations.

A total of 1902 beads were decoded and analyzed by single bead LC/MS, representing a library sampling of 7.6%. Given that the library was prepared in 200-fold redundancy, a small number of replicate structures is expected during decoding. Experimentally, 65 replicate structures (64 duplicates and 1 triplicate) were found. The F versus NF assignments matched for 61 of the replicates or 94% (data not shown). The high reproducibility of the comparative assignments observed for the replicate structures, together with the consistent composite p values of the three QA analyses, provide a satisfactory level of validation for this statistical-based approach to library quality assessment.[28]

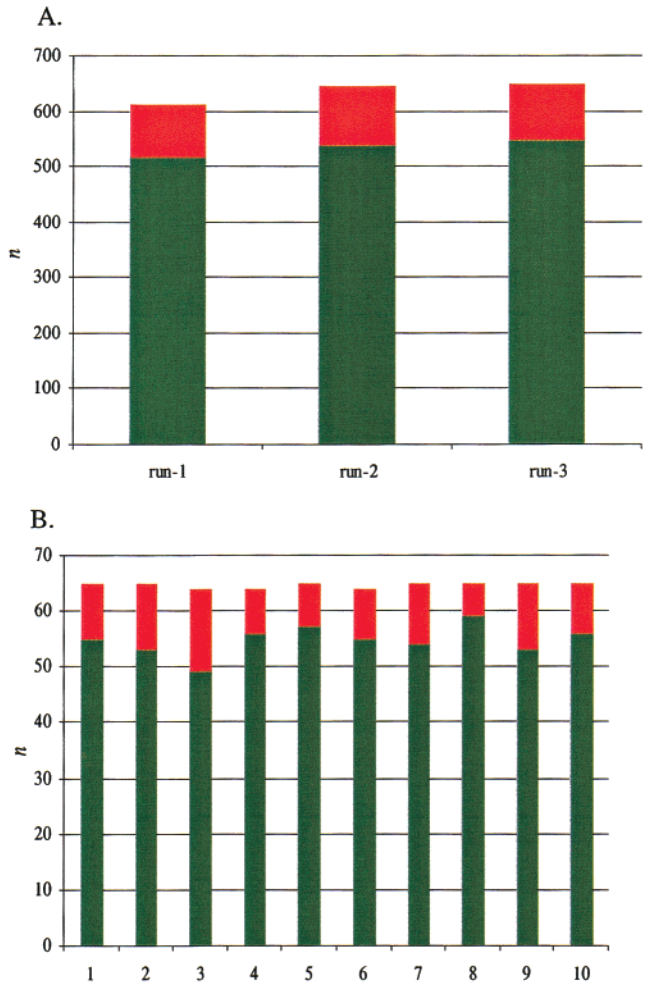


Figure 3. Composite library **1** QA analysis (A) and sublibrary (**C**/**D** synthon) analysis for QA run-3 (B) (F = green, NF = red).

Examination of the statistical results obtained for each sublibrary reveals that the F and NF answers are fairly evenly distributed (Table 3). The data from QA run-3 is representative of the sublibrary analyses and is graphically depicted in Figure 3B. The proportion of positively identified compounds in QA run-3 ranged from 76.6% (sublibrary 3) up to 91.2% (sublibraries 2 and 9). The sample size n for each sublibrary is 64−65 beads, and the typical *lb-ub* associated with p is ca. ±12%, compared with ±5% for the composite p values where n = ca. 650. The overlapping ranges for p across all sublibraries are again in excellent internal agreement for the three QA analyses. The data implies that there were no gross synthetic failures in the two-step derivatization sequence in which the **C**/**D** synthons were coupled to resin intermediate **5** (Scheme 1).

Table 3. Composite and Sublibrary QA Analysis for Library **1**

	run 1				run 2				run 3			
QA	*n*	F	*p* (%)	*lb-ub* (%)	*n*	F	*p* (%)	*lb-ub* (%)	*n*	F	*p* (%)	*lb-ub* (%)
composite	610	515	84.4	81.5−87.2	645	536	83.1	80.2−85.9	647	547	84.5	81.8−87.3
sublibrary												
8a	61	53	86.9	77.1−95.5	65	54	83.1	72.8−92.4	65	55	84.6	74.6−93.6
8b	63	56	88.9	79.8−96.8	63	53	84.1	73.9−93.3	65	53	91.2	81.5−95.0
8c	61	52	85.2	75.1−94.3	65	53	81.5	70.9−91.2	64	49	76.6	65.1−87.3
8d	62	49	79.0	67.7−89.5	65	51	78.5	67.3−88.7	64	56	87.5	78.1−95.7
8e	62	56	90.3	81.6−97.8	65	52	80.0	69.1−90.0	65	57	87.7	78.5−95.8
8f	62	47	75.8	64.8−86.8	64	50	78.1	66.9−80.5	64	55	85.9	76.2−94.6
8g	63	48	76.2	64.5−87.0	65	53	81.5	70.9−91.2	65	54	83.1	72.8−92.4
8h	57	50	87.7	77.8−96.4	65	55	84.6	74.6−93.6	65	59	90.8	82.5−97.9
8i	60	51	85.0	74.7−94.2	64	56	87.5	78.1−95.7	65	53	91.2	81.5−95.0
8j	59	53	89.8	80.7−97.6	64	59	92.2	84.3−98.8	65	56	86.2	76.5−94.7

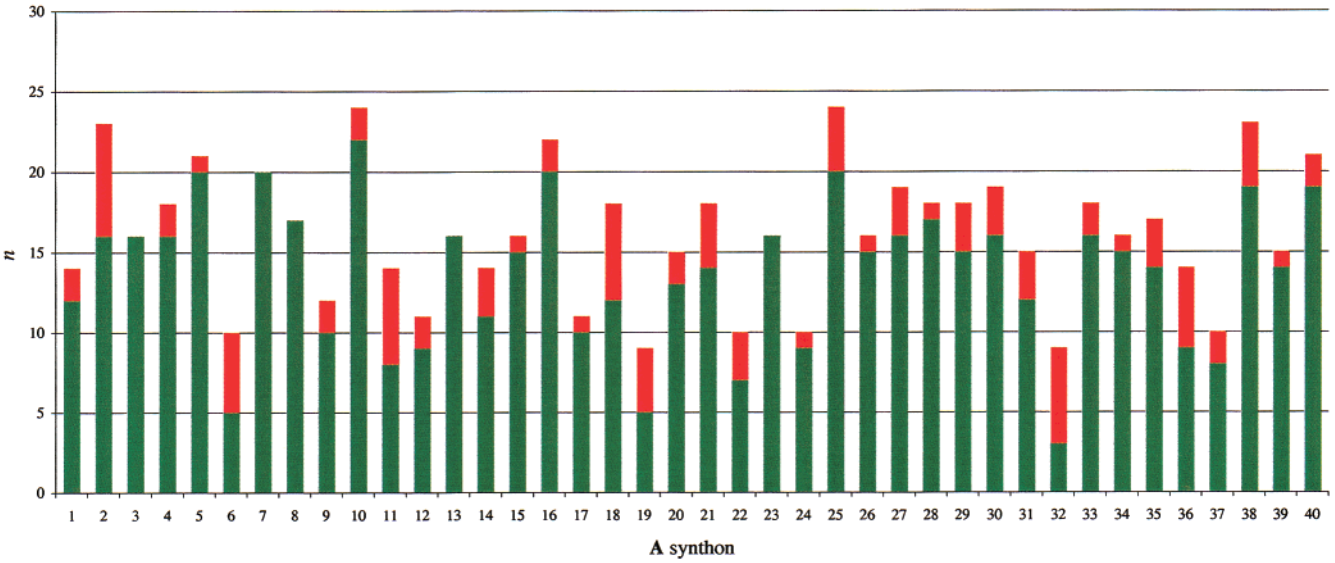


Figure 4. Synthon set **A** analysis (QA run-3) (F = green; NF = red).

Table 4. Proportion (*p*) and Boundaries (*lb-ub*) for Selected **A** Synthons

	run 1				run 2				run 3			
A	*n*	F	*p* (%)	*lb-ub* (%)	*n*	F	*p* (%)	*lb-ub* (%)	*n*	F	*p* (%)	*lb-ub* (%)
1	16	14	87.5	69.8−100	13	9	69.2	39.4−95.9	14	12	85.7	65.7−100
3	17	16	94.1	80.8−100	13	10	76.9	53.0−100	16	16	100	89.1−100
4	15	15	100	81.9−100	10	9	90.0	67.8−100	18	16	88.9	73.0−100
5	18	17	94.4	81.8−100	6	6	100	60.7−100	21	20	95.2	84.3−100
6	9	5	55.6	17.9−91.6	12	5	41.7	10.9−73.8	10	5	50.0	14.8−85.2
8	15	13	86.7	67.9−100	22	20	90.9	77.7−100	17	17	100	83.8−100
9	17	16	94.1	80.8−100	6	4	66.7	25.9−100	12	10	83.3	60.3−100
11	22	17	77.3	55.6−95.5	8	3	37.5	6.0−77.7	14	8	57.1	27.6−85.7
12	16	12	75.0	49.6−97.1	16	8	50.0	22.7−77.3	11	9	81.8	56.9−100
13	11	11	100	76.2−100	17	14	82.4	63.4−100	16	13	81.3	61.2−100
15	11	11	100	76.2−100	15	12	80.8	58.5−100	16	15	93.8	79.6−100
17	12	9	75.0	49.4−100	16	15	93.8	79.6−100	11	10	90.9	70.6−100
18	16	12	75.0	49.6−97.1	13	9	69.2	39.4−95.9	18	12	66.7	41.4−90.0
19	17	11	64.7	38.6−89.1	13	6	46.2	16.0−76.9	9	5	55.6	17.9−91.9
22	11	11	100	76.2−100	15	14	93.3	78.3−100	10	7	70.0	40.4−100
23	10	9	90.0	67.8−100	7	7	100	65.2−100	16	16	100	82.9−100
29	18	18	100	84.7−100	9	7	77.8	48.2−100	18	15	83.3	65.3−100
30	10	8	80.0	53.0−100	6	5	83.3	47.9−100	19	16	84.2	67.0−100
32	11	5	45.5	12.6−79.2	15	8	53.3	24.9−81.3	9	3	33.3	0.0−54.3
33	15	13	86.7	67.9−100	13	13	100	79.4−100	18	16	88.9	73.0−100

Statistical information regarding the performance of the **A** synthons may also be obtained from the library QA analysis (Figure 4 and Table 4). The sample size n_i for the **A** synthons varies from 10 to 24 with an average $n_i = 16$ (645 beads/40 **A** synthons). Recorded p_i values range from a low of 33.3% (**A32**, QA run-3) to 100% (e.g., **A3**; QA run-3) with the majority (ca. 35 out of 40) of the **A** synthons having p_i's >75%. The computed confidence boundaries are

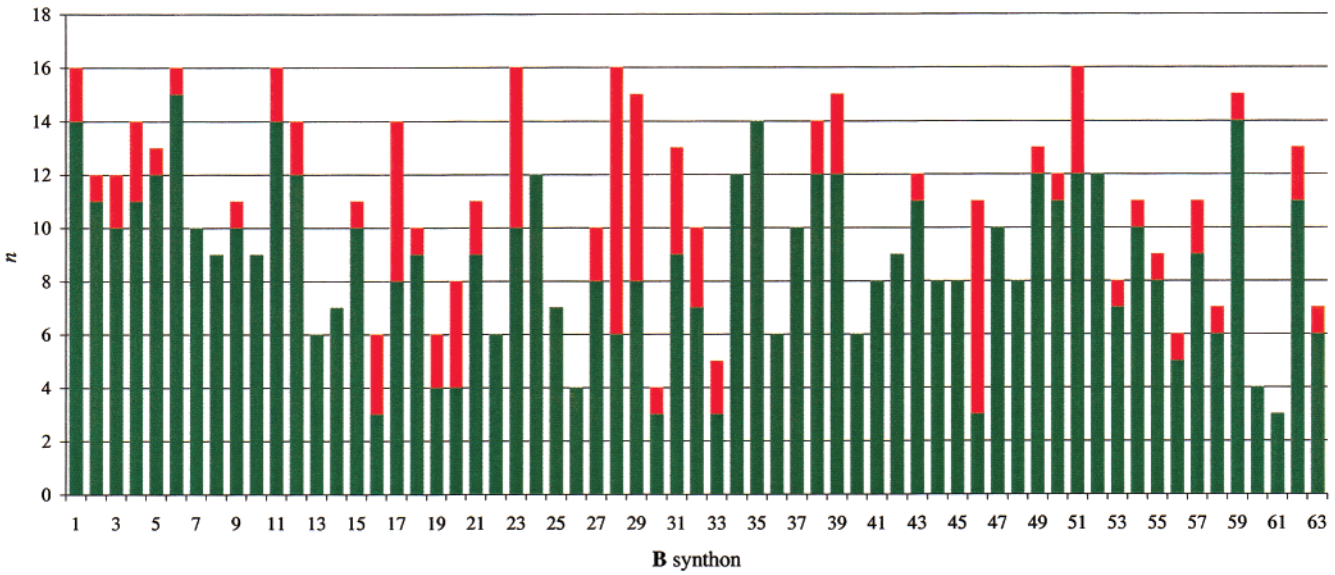


Figure 5. Synthon set **B** analysis (QA run-3) (F = green; NF = red).

much wider with the smaller sample sizes, particularly as p_i approaches 50%. The confidence limits for any given proportion overlaps in each QA run. As a group, the hydrophobic synthons, including the aliphatic amines (e.g., **A1**, **A3**, **A7**), arylalkylamines (e.g., **A10**, **A15**, **A20**), and the amino ethers (e.g., **A4**, **A5**, **A8**, **A9**), were strong performers, suggesting that they were successfully incorporated into the library. Noteworthy are the relatively high average proportions (p_i >80%) found for the hindered amines, neopentylmethylamine (**A33**) and cyclooctylamine (**A35**).

One exception to this trend is the neutral, hydrophobic synthon, methylthioethylamine (**A6**).[29] The proportion $p_{\mathbf{A6}}$ = 50%, *lb-ub* = 14.8−85.2% in QA run-3, and comparable $p_{\mathbf{A6}}$ values were found in the other two QA runs: $p_{\mathbf{A6}}$ = 55.6%, *lb-ub* = 17.9−91.6% (QA run-1), and $p_{\mathbf{A6}}$ = 41.7%, *lb-ub* = 10.9−73.8% (QA run-2). The true value of $p_{\mathbf{A6}}$ cannot be estimated with any certainty, although implicit in the data is that the synthon's performance is mediocre at best. The poor showing for **A6** is likely a reflection of its propensity to undergo photooxidation upon its release from the bead. It is reasonable to assume that the addition of amine **A6** to resin (chemical step; **3** to **4**, Scheme 1) occurred in high yield, analogous to other structurally related, successfully incorporated amines, e.g., synthon **A4** ($p_{\mathbf{A4}}$ 88.9%).[30] It can be argued too, that the poor performance of furfurylamine (**A11**; $p_{\mathbf{A11}}$ = 57.1%, *lb-ub* = 27.6−85.7%) may also be due to the known photooxidation of this heterocycle.

In contrast to the performance of the neutral, hydrophobic synthons, the hydrophilic synthons possessing positive or negative charged atoms or functional groups are more variable. For example, the proportions determined for the pyridinylalkylamines (**A17**, **A18**, **A19**, **A36**, and **A37**) are $p \sim 50\%$. The 3-aminopropylimidazole (**A32**) synthon is an especially poor performer: $p_{\mathbf{A32}}$ = 33.3%, *lb-ub* = 0.0−54.3%. On the basis of the QA results, it is likely that many of the putative compounds in the library possessing these synthons are absent. Library QA is performed on cleaved compounds, and it is unclear whether the divergence between the performance of hydrophobic versus hydrophilic **A** synthons, is a result of chemical failure, the poor diffusion of these compounds out of the bead matrix, photochemical sensitivity, or a combination of these or other factors.[31,32]

Figure 5 displays the histogram for the F and NF answers for a set of **B** synthons (QA run-3), and Table 5 provides computed p_i's for selected **B** synthons. Sample sizes $n_{\mathbf{B}}$ range from 3 to up to 16 with the average $n_{\mathbf{B}}$ = 10 beads (647 beads/63 **B** synthons). The experimentally determined proportions will invariably have the largest boundary versus the other synthon sets at n = 10, which is regarded as a minimal sampling size for meaningful interpretation. The **B** synthons have their widest boundary as p approaches 50%, e.g., synthon **B29**: $p_{\mathbf{B32}}$ = 53.3%, *lb-ub* = 24.9−81.3% (QA run-3). A sufficient level of confidence in $p_{\mathbf{B}}$ may be realized at the high and low positive identifications, and there are a number of synthons that meet this criteria. Glycine (**B1**), sarcosine (**B6**), the acyclic β- and γ-amino acids (**B2**, **B3**, **B38**, **B39**), the hydrophobic α-amino acids including L- and D-prolines (**B4**, **B5**) and L- and D-phenylglycines (**B7**, **B8**), tetrahydroisoquinoline carboxylic acids (**B58**−**B61**), and many of the unnatural, peptidomimetic-type amino acids (**B45**, **B47**−**B57**, **B62**, **B63**) may all be regarded as well performing synthons. The proportions observed for the aspargines (**B19**, **B20**), glutamines (**B28**, **B29**), L-citrulline (**B23**), and the methionines (**B32**, **B33**) are near p = 50%; hence, the true p for these synthons cannot be accurately estimated. Other amino acids with hydrophilic side chains (aspartic acids (**B21**, **B22**), glutamic acids (**B26**, **B27**), histamine (**B30**, **B31**), arginines (**B17**, **B18**), and the lysines (**B11**, **B12**) faired reasonably well, with p values close to 70%, although boundaries are on the order of ±30%. One synthon, **B46**, was clearly a problem in the library. Proportion $p_{\mathbf{B46}}$ in QA run-3 = 27%, *lb-ub* = 0.0−45.6%, and for QA run-2, $p_{\mathbf{B46}}$ = 30.8%, *lb-ub* = 4.1−60.6%. This synthon was not found at all in QA run-1: $p_{\mathbf{B46}}$ = 0%, *lb-ub* = 0.0−34.8%. The poor performance synthon **B46** is believed to be due to a chemical failure in the library (*vida infra*).[33,34]

Enzyme Assay Results. Initially, 1 equiv of library **1** was screened against cat D and plm II at ca. 30 compounds

Table 5. Proportion (*p*) and Boundaries (*lb-ub*) for Selected **B** Synthons

	run 1				run 2				run 3			
B	*n*	F	*p* (%)	*lb-ub* (%)	*n*	F	*p* (%)	*lb-ub* (%)	*n*	F	*p* (%)	*lb-ub* (%)
3	6	5	83.3	47.9−100	10	8	80.0	53.0−100	12	10	83.3	60.3−100
7	7	7	100	65.2−100	6	5	83.3	47.5−100	10	10	100	74.1−100
9	10	10	100	74.1−100	10	9	50.0	67.8−100	11	10	90.9	70.6−100
13	11	11	100	76.2−100	11	11	100	76.2−100	11	11	100	76.2−100
15	8	7	87.5	60.2−100	10	6	60.0	24.4−93.4	11	10	90.9	70.6−100
16	15	13	86.7	67.9−100	13	12	92.3	75.0−100	6	3	50.0	4.4−95.6
18	9	8	88.9	64.4−100	7	5	71.4	35.1−100	10	9	90.0	67.8−100
19	15	11	73.3	46.6−96.8	10	3	30.0	0.0−49.6	6	4	66.7	25.9−100
20	10	7	70.0	40.4−100	9	3	33.3	0.0−54.3	8	4	50.0	10.2−89.8
23	10	8	80.0	53.0−100	11	5	45.5	12.6−79.2	16	10	62.5	35.3−88.2
26	12	10	83.3	60.3−100	9	4	44.4	1.0−82.1	4	4	100	47.3−100
28	11	8	72.7	45.3−100	9	8	88.9	64.4−100	16	6	37.5	11.8−64.7
29	12	8	66.7	35.1−95.3	14	10	71.4	43.2−96.4	15	8	53.3	24.9−81.3
30	10	7	70.0	40.4−100	7	2	28.6	0.0−50.6	4	3	75.0	24.8−100
32	3	3	100	36.8−100	7	4	57.1	13.6−97.9	10	7	70.0	40.4−100
46	7	0	0.0	0.0−34.8	13	4	30.8	4.1−60.6	11	3	27.3	0.0−45.6
58	4	4	100	47.3−100	8	7	87.5	60.2−100	7	6	85.7	54.9−100
59	9	9	100	71.7−100	14	13	92.9	76.8−100	15	14	93.3	78.3−100
60	8	8	100	68.8−100	11	11	100	76.2−100	4	4	100	47.3−100
61	9	8	88.9	64.4−100	9	8	88.9	64.4−100	3	3	100	36.8−100

per well. The estimated screening concentration was 1 μM, based on the bead yields of a set of QC compounds (data not shown). The highest activity was observed in sublibrary **8j**, defined by the synthon pair **C2/D10**, followed by sublibrary **8a**, defined by the synthon pair **C1/D1**. Subsequently, two additional follow-up copies of sublibrary **8j** were screened at a single compound per well (5040 beads, 87% library coverage). Because of the exceptionally high potency observed in this sublibrary during the survey screen, the estimated screening concentration was reduced to ca. 0.05 μM. Overall, sublibrary **8j** appeared more active and selective against cat D than plm II.

Only those wells (single beads) that showed activity equal to or less than 30% control activity remaining were decoded. This follow-up evaluation resulted in a total of 55 decoded structures for plm II, including 13 replicate structures (three compounds seen 4×, two compounds seen 3×, and nine compounds seen 2×) and 19 unique structures. For cat D, the follow-up evaluation resulted in 76 decoded structures, including 21 replicate structures (two compounds seen 4×, six compounds seen 3×, and 14 compounds seen 2×) and 22 unique structures. The primary screening activity (% control remaining) obtained for the decoded, replicate plm II structures indicated marginal selectivity (<10-fold versus cat D). In contrast, the decoded replicate cat D structures showed a range of selectivities from 1:1 to >20:1.

Figure 6 provides a graphic representation of the $\mathbf{R_A}$ and $\mathbf{R_B}$ synthon frequencies for the decoded structures of sublibrary **8j** against the two enzymes. Both enzymes preferred a rather broad range of $\mathbf{R_A}$ amines, providing these were neutral and hydrophobic. Plm II showed the largest range for its 32 decoded structures: 21 out of 55 **A** synthons were observed, including alkyl, cycloalkyl, and arylalkylamines. For cat D, which was somewhat more selective, 15 out of 40 **A** synthons were observed for the 76 decoded structures; phenethylamine (**A23**; observed 12 times) had the highest frequency. Interestingly, the amines containing positive or negative charge groups, e.g., the pyridinylmethylamines (**A17**−**19**, **A36**, and **A37**), glycine and the β-aminoalanine synthons (**A29**−**30**), and imidazolylethylamine (**A32**) were not observed in the decoded structures, suggesting inhibitors with such substituents have estimated K_i values of >50 nM.

In contrast to the broad structural variation at position $\mathbf{R_A}$, the enzymes were more stringent in their preferences for the $\mathbf{R_B}$ amino acids. Hydrophobic residues were again preferred over hydrophilic residues. L-Cyclohexylalanine (**B24**) and L-thienylalanine (**B9**) were found in 23 out of 32 decodes for plm II, along with L-tyrosine (**B13**), L-phenylalanine (**B34**), and L-valine (**B42**). Although cat D shared many of the same amino acids with plm II, including L-thienylalanine, L-tyrosine, L-phenylalanine, and L-valine, L-cyclohexylalanine **B24** was the most frequently observed. In addition, there were several **B** synthons that were unique to cat D, which appeared to impart a high degree of selectivity: D-tetrahydroisoquinoline-1-carboxylic acid (D-Tiq; **B58**), 3-aminopropionic acid (**B3**), 3-amino-cyclohexane carboxylic acid (**B53**), and the aminocaprolactam, (**B45**).

Five replicate structures were chosen for resynthesis. These included the two most potent plm II inhibitors (**9**: **A14-B24-C2-D10** and **10**: **A8-B24-C2-D10**, Table 6), and three of the most potent and selective cat D inhibitors (**11**: **A21-B45-C2-D10**; **12**: **A23-B53-C2-D10**; and **13**: **A11-B24-C2-D10**) as identified from the primary screening data. Inhibitors **9** and **10** possess K_i values of 29 nM and 16 nM against plm II and show modestly selective (1.5−2×) versus cat D: **9**: K_i = 44 nM, **10**: K_i = 32 nM. The selected resynthesized compounds for cat D also confirmed the activity and selectivity as observed in the primary screen. Inhibitor **11**, containing the D-Tiq residue, possesses a K_i = 3 nM, while compounds **12** and **13** possess K_i's = 4 nM and 18 nM, respectively. Statines **11**−**13** are ca. 5 to 70-fold selective versus plm II, and are structurally unique among previously reported cathepsin D inhibitors.[35]

Corroborating Screening SAR Data with Library QA Analysis. When an encoded library is screened and found

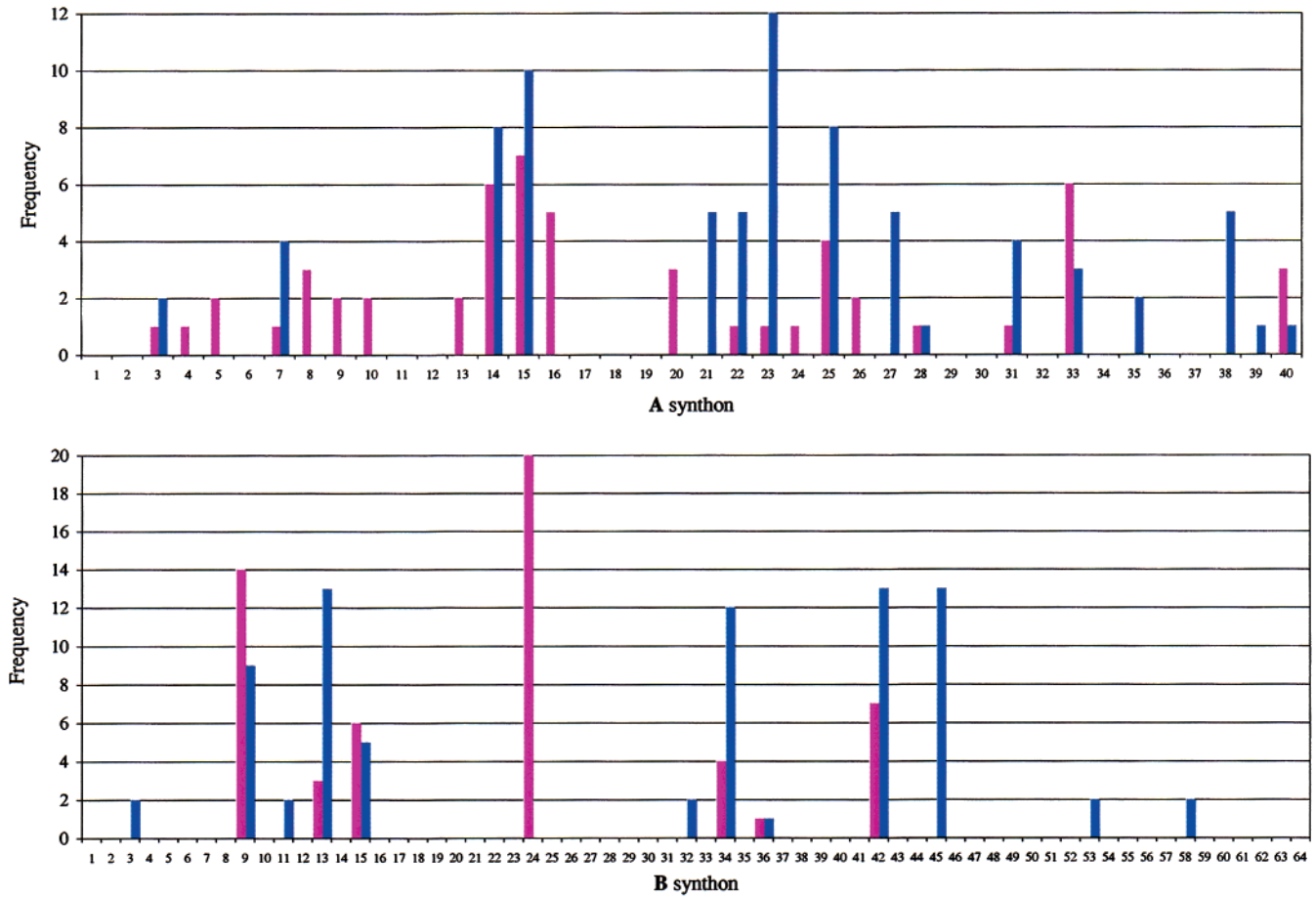


Figure 6. **A** and **B** synthon frequencies for decoded structures from sublibrary **8j** (plm II: pink, cat D: blue).

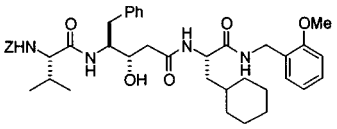
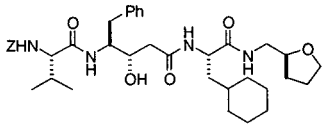
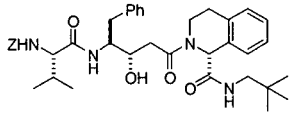
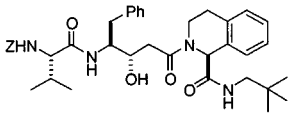
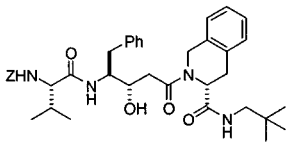
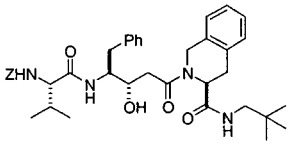
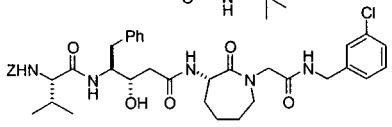
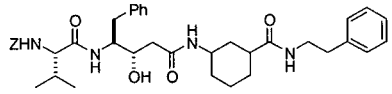
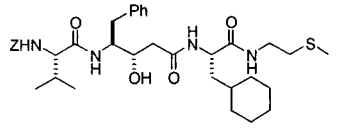
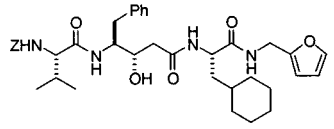
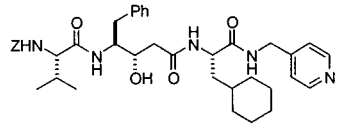
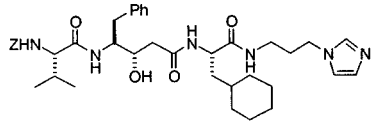
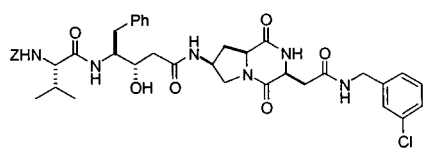
biologically active, a large number of beads are decoded revealing the structures of potentially active compounds. Confirmation of biological activity comes from compound resynthesis and evaluation, and typically, only 5−10 compounds are resynthesized for this purpose. Compounds selected for resynthesis are those found multiple times (replicate structures) and are the most potent as per the primary screen. (It is this strategy that was followed in the screening of library **1** described above.) Upon confirmation of biological activity, a decision may be made to resynthesize additional compounds from the library, synthesize individual analogues, or prepare a follow-up optimization library. In any event, the decision regarding the design and synthesis of new compounds, will, in part, rely on the nascent SAR obtained from the original screen. It is therefore important that such SAR information be reliable and of superior quality. This is the benefit of library QA analysis. By rendering information on the overall success of library synthesis and the performance of individual synthons, library QA can be utilized to corroborate and elevate the level of confidence in the screening SAR data.

For example, SAR information obtained from decoded structures indicates that plm II and cat D have strong preferences for hydrophobic **A** synthons. Comparison of the preferred **A** synthons against the performance of the same synthons in library QA (Figures 4 and 6) reveals that all were strong performers. However, one synthon conspicuously absent from the screening data is the methylthioethylamine synthon **A6**. Given the fact that this synthon is neutral and hydrophobic,[29] it might be expected to be among the decoded structures; however, this was not the case. The QA data for **A6** showed a rather dubious performance (average *p* averaged for the three QA runs = 48.4%, *lb-ub* = 29.4−67.5%). This uncertainty suggests, at the very least, that compounds containing this synthon are likely to be underrepresented in the follow-up single bead screen. In light of the QA data, statine **14** was synthesized. Statine **14**, a putative library member, is an analogue of **9**, in which the *o*-methoxybenzylamine (**A14**) is replaced with methylthioethylamine (**A6**). Statine **14** possesses a K_i = 12 nM against plasmepsin II and a K_i = 44 nM against cathepsin D, and is thus similar in potency and selectivity to inhibitor **9** (Table 6).

Similarly, furfurylamine **A11** was not found in decoded structures,[29] although the structurally analogous *R*- and *S*-tetrahydrofuranylmethylamines (**A8** and **A9**) were both present. Library QA analysis clearly shows uncertainty regarding the presence of this synthon in cleaved library compounds. As was the case for the synthon **A6**, putative library members containing **A11** in combination with preferred $\mathbf{R_B}$ amino acids, would be expected to be highly active. To test this hypothesis, statine **15** was prepared. This analogue of **9** (**A14** in **9** exchanged for **A11** in **15**) was indeed found to be a potent aspartyl protease inhibitor: **15**: K_i = 26 nM, plm II; 46 nM, cat D. In the absence of QA data, it may have been erroneously concluded that synthons **A6** and **A11** were "inferior" **A** synthons.

As a final comparative example for the **A** synthons, positive charged, hydrophilic amines such as the pyridinyl-

Table 6. Inhibition Constants (K_i) for Resynthesized Compounds

No.	Structure	K_i (nM)[e] plm II	cat D
9[a]		29	44
10[b]		16	32
11[a]		210	3
18[c]		44,000	280
19[c]		18,000	290
20[c]		28,000	2,900
12[a]		29	4
13[a,d]		72	18
14[c]		12	44
15[c]		26	46
16[c]		15	140
17[c]		7.0	530
21[c]		62	56

[a] Duplicate decode. [b] Triplicate decode. [c] Not a screening decode. [d] Mixture of diastereomers, see Figure 2. [e] Average of >2 determinations, std error, <15%.

methylamines (**A17**−**19**, **A36**, and **A37**) and imidazolylpropylamine **A32** were not found in any of the decodes from the screen.[36] This suggests that compounds containing these synthons would be much weaker inhibitors (>30% control remaining) relative to those inhibitors containing the neutral, hydrophobic **A** amines. Inspection of the QA data for these synthons reveals uncertainty in their performance. In particular, **A19** (p = 55.6%, *lb-ub* = 17.9−91.9%; QA run 3) and **A32** (p = 33.3%, *lb-ub* = 0.0−54.3%, QA run 3) may be considered poor to mediocre synthons. Although the screening SAR strongly argues that compounds containing these synthons are unlikely to be as potent as their hydrophobic congeners, this information cannot be reliably established based on library screening. To obtain a more accurate portrait of the SAR, putative library members, statines **16** and **17** containing the 4-pyridinylmethyl (**A19**) and imidazoylylpropylamine (**A32**) synthons, were synthesized. Remarkably, compounds **16** (K_i = 15 nM, plm II; K_i = 140 nM, cat D) and **17** (K_i = 7 nM, plm II; K_i = 530 nM, cat D) were found to be the most potent and selective inhibitors of plm II identified in the library (Table 6). The ability of plm II to tolerate protonated **A** synthons versus cat D, leading to enhanced selectivity, is a salient SAR feature that would have been completely lost relying solely on screening decode data.

One novel and unexpected discovery observed in screening, and verified by resynthesis, was the identification of D-Tiq (**B58**) as a $P_{2'}$ amino acid residue selective for cathepsin D. This is exemplified by inhibitor **11**: K_i = 3 nM, cat D; K_i = 210 nM, plm II. Although all four isomeric tetrahydroquinoline carboxylic acids **B58**−**61** were included in the library, only **B58** appeared in the decodes. Because it is uncommon to find D-amino acid residues located at the $P_{2'}$ position in protease substrates and inhibitors,[35] the reliability of the screening SAR may be suspect. Examination of the library QA data for synthons **B58**−**61** reveals that they were a successful group of synthons (p > 85%). In this instance, the QA data reinforces the confidence in the SAR screening data in that D-Tiq is in fact the preferred synthon. This was substantiated with the synthesis and evaluation of inhibitors **18**−**20** (Table 6), the diastereomers of **11** where the **B58** D-Tiq residue was replaced with L-Tiq (**B59**), D- and L-Tic (**B50** and **B61**). Inhibitor **18** containing L-Tiq has a K_i = 280 nM against cat D, the D- and L-Tic containing inhibitors **19** and **20** possess inhibition constants of 290 nM and 2900 nM, respectively, against cat D. These data are clearly consistent with the activity observed in primary screening and corroborated by library QA.

Last, library QA indicates a virtual failure of synthon **B46**. In the absence of these data, it would have been concluded from screening that inhibitors containing **B46** would have potencies estimated to be >50 nM. In reality, library members containing this synthon were probably never present in the screen. Hence, no conclusion may be drawn regarding the activity of library compounds containing the **B46** residue. Individual compounds containing this synthon must be synthesized separately in order to determine their biological profile. One compound (**21**) of this class was prepared, an analogue of inhibitor **12**, substituting synthon **B46** for synthon **B45** (Table 6). Statine **12**, identified as a duplicate

from the screen, is a potent, selective inhibitor of cat D (4 nM, 7-fold selective); compound **21** containing **B46** was found 10-fold less active and nonselective: K_i for **21** = 56 nM, cat D; K_i = 62 nM, plm II. Several other synthons including L- and D-asparagine (**B19** and **B20**), L- and D-glutamine (**B28** and **B29**), and L- and D-methionine (**B32** and **B33**) would likewise have to be synthesized as discrete compounds to more accurately establish the enzymes' affinity toward these amino acids.

Summary

Library QA is a powerful analytical protocol to assess the quality of large, encoded combinatorial libraries. Based on the combined application of tag decoding and single bead LC/MS, QA analysis would be virtually impossible to carry out without an encoding strategy due to redundant masses produced during split synthesis. The statistical theory and its application to library assessment are analogous to the accepted statistical sampling practices used in industry to ascertain the quality of mass-produced material, e.g., tablet quality control in the pharmaceutical industry. The simplicity of tag reading and rapid acquisition of SAR information is arguably the most significant advantage of encoding technology versus other deconvolution techniques.[2] Library QA serves to substantiate and enhance the value of nascent SAR obtainment from library screening.

The statistical sampling protocol may be coupled with UV analysis, evaporative light scattering detection (ELSD), or chemiluminescent nitrogen detection (CLND), providing an avenue to assess not only identity, but the purity and semiquantitation of library members. This is especially true for libraries with bead yields in excess of 1 nM/bead. The use of on-line hyphenated LC/UV/ELSD(CLND)/MS instrumentation in library QA analysis will be the subject of future reports from our laboratories.

Experimental Section

Library Synthesis. Preparation of Encoded Resin 2. Bis-Fmoc lysine (11.2 mmol, 6.8 g) and HOBt (11.2 mmol, 1.5 g) were added to a suspension of Rapp TentaGel resin (S-NH_2, 12 g, 0.32 mmol/g, 3.84 mmol, 180−220 μm) in DMF (60 mL). The reaction mixture was shaken for 30 min to dissolve the reagents, and then DIC (22.4 mmol, 3.6 mL) was added in one portion. The reaction mixture was shaken overnight (Burrell wrist-action shaker), drained, and washed with DMF (3 × 50 mL), MeOH (3 × 50 mL), and DCM (3 × 50 mL). Resin loading was determined from Fmoc titration to be 0.6 mmol/g. The bis-Fmoc coupled resin was apportioned into 40 reaction vessels (0.35 g resin per vessel) and encoded with one or more of the $C_{12}Cl_5$-, $C_{11}Cl_5$-, $C_{10}Cl_5$-, C_9Cl_5-, C_8Cl_5-, and C_7Cl_5-diazoketone tags to produce the appropriate binary code.[4a] The encoded resins were treated with a solution of piperidine−DMF (1:4; 10 mL), shaken for 1.5 h, drained, and washed with DMF (3 × 10 mL), MeOH (3 × 10 mL), and DCM (3 × 10 mL). The resins were then treated with a preincubated (30 min) solution of 4-bromomethyl-3-nitrobenzoic acid (0.63 mmol, 163 mg), HOBt (0.63 mmol, 85 mg), and DIC (1.26 mmol, 0.2 mL) in DMF (5 mL). The resulting suspensions were shaken overnight, drained, and washed with DMF (3 × 10 mL), methanol (3 × 10 mL), and DCM (3 × 10 mL), and then dried in vacuo to give resins **3**.

Addition of the Amine (A) Synthons. Encoded resins **3** (0.35 g, 0.6 mmol/g, 0.21 mmol) were treated with one of 40 amines **A** (ca. 10 equiv of amine as a 0.5 M solution in THF) and shaken overnight. The resins were drained and washed with DMF (2 × 10 mL), MeOH (2 × 10 mL), MeOH−TFA (10:1; 2 × 10 mL), MeOH (2 × 10 mL), DMF (2 × 10 mL), and DCM (2 × 10 mL) and dried in vacuo. The amine resins **4** were combined, mixed, and distributed into 63 reaction vessels.

Addition of the Fmoc-Amino Acid (B) Synthons. Resins **4** (0.22 g, 0.6 mmol/g, 0.13 mmol) were treated with a 0.2 M solution of one of 63 amino acids **B** (3 equiv) in DMF followed by the addition of HATU (0.40 mmol, 152 mg). Diisopropylethylamine (DIEA; 0.8 mmol, 0.15 mL). The reaction suspensions were shaken for 6 h, drained, washed with DMF (2 × 10 mL), MeOH (2 × 10 mL), DMF (2 × 10 mL), and DCM (2 × 10 mL) and then dried. The Fmoc-resins **5** were encoded with one or more of the C_6Cl_5-, C_5Cl_5-, C_4Cl_5-, C_3Cl_5-, C_6Cl_3-, and C_5Cl_3-diazoketone tags to produce the appropriate binary code.[4a] The encoded resins were combined, mixed, and distributed into two reaction vessels.

Addition of the Fmoc-Statine (C) Synthons. Resins **5** (7.2 g, 0.6 mmol/g, 4.32 mmol) were suspended in piperidine−DMF (1:4; 50 mL) and shaken for 1.5 h. The resins were drained and washed with DMF (3 × 50 mL), MeOH (3 × 50 mL), and DCM (3 × 50 mL). The corresponding amine resins were treated with a 0.2 M solution of one of two statines **C** (3 equiv) in DMF to which was added HATU (13.1 mmol, 5 g) and DIEA (26.5 mmol, 4.8 mL). The reaction mixtures were shaken for 6 h and then drained. The resins were washed with DMF (3 × 50 mL), MeOH (3 × 50 mL), and DCM (3 × 50 mL) and dried in vacuo to give resin **6a** (**A(1−40)-B(1−63)-C1-Fmoc**) and resin **6b** (**A(1−40)-B(1−63)-C2-Fmoc**).

Addition of Carboxylic Acid (D) Synthons. Resins **6a** and **6b** (ca. 7.2 g, 0.6 mmol/g, 4.3 mmol each) were suspended in piperidine−DMF (1:4; 50 mL) and shaken for 1.5 h, then drained, and washed with DMF (3 × 50 mL), MeOH (3 × 50 mL), and DCM (3 × 50 mL). Deprotected amine resin **7a** was distributed equally into five reaction vessels (1.40 g, 0.6 mmol/g, 0.84 mmol). Each reaction vessel was charged with a 0.2 M solution of one of five acids **D** (**D1−5**; 2.5 mmol; 3 equiv) in DMF followed by the addition of HATU (2.5 mmol, 950 mg) and DIEA (5.1 mmol, 0.95 mL). The reaction mixtures were shaken for 6 h and washed with DMF (3 × 20 mL), MeOH (3 × 20 mL), and DCM (3 × 20 mL). The acid-labile protecting groups in the amino acid **B** synthons were removed upon exposure of the resins to a solution of TFA−phenol−EDT−water (80:5:5:3.5; 15 mL) for 1.5 h. In the case of subunit **D5**, the resin was first treated with a 20% solution of hydrazine in DMF for 3 h to remove the phenolic-acetyl protecting group. Following the deprotection protocol, the resins were washed with THF−water (1:1; 3 × 20 mL), DMF (3 × 20 mL), MeOH (3 × 20 mL), and DCM (3 × 20 mL) and dried in vacuo to give

sublibraries **8a**−**e**. In analogous fashion, amine resin batches **7b** were derivatized with acids **D6**−**10** affording sublibraries **8f**−**j**.

Photolytic Cleavage. Sublibraries **8a**−**j** were arrayed in to 96-well filter-bottom plates (30 beads per well for initial survey screening, or single beads per well for follow up plates of sublibrary **8j** and library QA analysis) using a custom bead arraying apparatus. The dried beads in each well were suspended in 150 μL of EtOH−water (80:20) and irradiated at 365 nm for 30 min at 50 °C employing a custom UV light chamber. The 30 min UV exposure was followed by a 2 h post soak in the dark at 50 °C. The bead eluent was collected into a 96-well assay plate, and the elution solvent was removed in a Genevac (0.1 Torr) for 2 h at 40 °C. Assay plates containing the dried compounds were used for the enzyme assays and library QA analysis.

Compound Resynthesis. Pure compounds were resynthesized using the solid-phase methodology outlined above or analogous solution-phase synthesis. After cleavage from the resin, the compounds were purified by preparative HPLC (Gilson 215). ^{1}H NMR were obtained on a Varian 300 MHz spectrophotometer, and chemical shifts (δ) are relative to TMS. Low-resolution mass spectra (LRMS) were obtained in-house on a Finnegan LCQ, while high-resolution mass spectra (HRMS: FAB+) were obtained courtesy of Dr. George Dubay: Paul M. Gross Chemical Laboratories, Duke University, Durham, NC 27708.

A14-B24-C2-D10 (9). R_f = 0.19 (EtOAc). ^{1}H NMR (CD_3-OD): δ 7.72 (d, 1H, J = 7.8 Hz), 7.40−7.10 (m, 14H), 6.90 (m, 2H), 6.60 (bs, 1H), 5.25 (d, 1H, J = 7.8 Hz), 5.11 (s, 3H), 4.40 (m, 4H), 4.05 (bs, 1H), 3.82 (s, 3H), 3.70 (m, 2H), 3.00 (m, 5H), 2.55 (m, 1H), 2.10 (m, 1H), 1.95 (m, 1H), 1.60 (m, 6H), 1.10 (m, 2H), 0.88 and 0.83 (doublets, 3H each, J = 6.6 Hz). LRMS *m*/*z* 715 $(M+H)^+$. HRMS (FAB+) $[M+H]^+$ calcd for $C_{41}H_{54}N_4O_7$: 715.8438, found 715.8441.

A8-B24-C2-D10 (10). R_f = 0.35 (5% MeOH−EtOAc). ^{1}H NMR ($CDCl_3$): δ 7.64 (d, 1H, J = 7.8 Hz). 7.40−7.10 (m, 10H), 6.58 (bs, 1H), 5.28 (d, 1H, J = 7.0 Hz), 5.17 and 5.15 (doublets, 1H each, J = 12.0 Hz), 4.38 (dd, 1H, J = 7.0 and 12.0 Hz), 3.72 (m, 4H), 3.38 (m, 1H), 3.17 (m, 3H), 2.60 (m, 6H), 2.10 (m, 1H), 2.0−1.8 (m, 5H), 1.75−1.50 (m, 8H), 1.20 (m, 4H), 0.88 and 0.85 (doublets, 3H each, J = 6.6 Hz). LRMS *m*/*z* 679 $(M+H)^+$. HRMS (FAB+) $[M+H]^+$ calcd for $C_{38}H_{54}N_4O_7$: 679.3679, found 679.3680.

A33-B58-C2-D10 (11). R_f = 0.24 (EtOAc). ^{1}H NMR ($CDCl_3$): δ 7.40−7.10 (m, 14H), 6.60 (bt, 1H), 6.25 (bs, 1H), 5.20−4.90 (m, 5H), 4.50−4.40 (m, 2H), 4.20−3.95 (m, 3H), 3.35 (m, 1H) 3.00−2.40 (m, 5H), 2.21−2.10 (m, 3H), 0.96 and 0.81 (d, 3H each, J = 6.6 Hz), 0.63 (s, 9H). LRMS *m*/*z* 671 $(M+H)^+$. HRMS (FAB+) $[M+H]^+$ calcd for $C_{39}H_{50}N_4O_6$: 671.3806, found 671.3804.

A21-B45-C2-D10 (12). R_f = 0.18 (EtOAc). ^{1}H NMR ($CDCl_3$): δ 7.45 (d, 1H, J = 6.3 Hz), 7.40−7.10 (m, 14H), 6.85 (bs, 1H), 5.30 (d, 1H, J = 7.0 Hz), 5.10 (bs, 2H), 4.65 (t, 1H), J = 7.5 Hz), 4.38 (d, 2H, J = 7.5 Hz), 4.20−3.60 (m, 4H), 3.40 (d, 1H, J = 14.7 Hz), 3.00 (bs, 2H), 2.60 (s. 7H), 2.43 and 2.19 (m, 1H each), 2.00−1.55 (m, 4H), 0.86 and 0.83 (d, 3H each, J = 6.3 Hz). LRMS *m*/*z* 734 $(M+H)^+$. HRMS (FAB+) $[M+H]^+$ calcd for $C_{39}H_{48}ClN_5O_7$: 734.3345, found 734.3340.

A23-B53-C2-D10 (13). R_f = 0.15 (EtOAc). ^{1}H NMR (CD_3OD): δ 7.40−7.20 (m, 15H). 5.20 and 5.17 (doublets, 1H each, J = 7.2 Hz), 4.11 (m, 3H), 3.84 (d, 1H, J = 7.0 Hz), 3.60 (m, 1H), 3.39 (1H, m), 3.25 (bs, 3H), 2.99−2.80 (m, 4H), 2.21 (m, 3H), 2.00−1.80 (m, 6H), 7.39−1.18 (m, 5H), 0.87 and 0.85 (doublets, 3H each, J = 6.6 Hz). LRMS *m*/*z* 671 $(M+H)^+$. HRMS (FAB+) $[M+H]^+$ calcd for $C_{39}H_{50}N_4O_6$: 671.3806, found 671.3810.

A6-B24-C2-D10 (14). R_f = 0.22 (EtOAc). ^{1}H NMR ($CDCl_3$): d 7.58 (d, 1H, J = 7.8 Hz), 7.40−710 (m, 10H), 6.98 (bs, 1H), 5.45 (d, 1H, J = 6.9 Hz), 5.12 (s, 2H), 4.35 and 4.32 (doublets, 1H each, J = 7.5 Hz), 4.07 (bs, 1H), 3.78 (m, 2H), 3.40 (m, 2H), 3.00 (d, 2H, J = 6.6 Hz), 2.50 (m, 4H), 2.15 (m, 1H), 2.10 (s, 3H), 1.90 (m, 1H), 1.60 (m, 8H), 1.40 (bs, 1H), 1.15 (m, 4H), 0.88 and 0.82 (doublets, 3H each, J = 6.9 Hz). LRMS *m*/*z* 669 $(M+H)^+$. HRMS (FAB+) $[M+H]^+$ calcd for $C_{36}H_{52}N_4O_6S$: 669.3693, found 669.3694.

A11-B24-C2-D10 (15). R_f = 0.35 (5% MeOH−EtOAc). ^{1}H NMR ($CDCl_3$) δ 7.49 (m, 1H), 7.40−7.00 (m, 12H), 6.71 (bs, 1H), 6.28 and 6.20 (singlets, 1H each), 5.21 (d, 1H, J = 3.7 Hz), 5.05 (s, 2H), 4.38 (m, 3H), 4.05 (bs, 1H), 3.76 (m, 2H), 3.01 (m, 3H), 2.50 (m, 1H), 2.05−1.95 (m, 2H), 1.63 (m, 6H), 1.40−0.95 (m, 5H), 0.90 and 0.86 (doublets, 3H each, J = 6.7 Hz). LRMS *m*/*z* 675$(M+H)^+$; (FAB+) $[M+H]^+$ calcd for $C_{38}H_{50}N_4O_7$: 675.3781, found 675.3801.

A19-B24-C2-D10 (16). R_f = 0.11(5% MeOH-EtOAc). ^{1}H NMR ($CDCl_3$): δ 8.60 (d, 1H, J = 5.7 Hz), 8.18 (m, 1H), 7.68 (d,1H, J = 5.7 Hz), 7.35 and 7.18 (singlets, 10H total) 7.02 and 6.87 (m, 2H each), 5.15 and 5.11 (d, 1H each, J = 12.3 Hz), 4.61 (m, 2H), 4.27 (m, 1H), 4.11 (m, H), 3.85 (m, 1H), 3.65 (m, 1H), 3.20−2.80 (m, 7H), 2.34 (m, 1H), 2.01-(m, 1H), 1.71 (m, 6H), 1.23 (m, 4H), 0.95 (m, 3H), 0.89 and 0.81 (doublets, 3H each, J = 6.9 Hz). LRMS *m*/*z* 686 $(M+H)^+$. HRMS (FAB+) $[M+H]^+$ calcd for $C_{39}H_{51}N_5O_6$: 686.3918, found 686.3912.

A32-B24-C2-D10 (17). R_f 0.10 (15% MeOH−EtOAc). ^{1}H NMR ($CDCl_3$): δ 8.78 (s, 1H), 7.89 (s, 1H), 7.71 (d, 1H, J = 9.3 Hz), 7.40−7.00 (m, 13H), 5.94 (d, 1H, J = 6.9 Hz), 5.10 and 5.07 (doublets, 1H each, J = 11.7 Hz), 4.11 (m, 4H), 3.78 (t, 1H, J = 7.5 Hz), 3.39 and 3.18 (m, 1H each), 3.00−2.75 (m, 3H), 2.40 and 2.20 (m, 1H each), 2.00 (m, 4H), 1.60 (m, 6H), 1.20−0.95 (m, 6H), 0.90 and 0.80 (doublets, 3H each, J = 6.9 Hz). LRMS *m*/*z* 703 $(M+H)^+$. HRMS (FAB+) $[M+H]^+$ calcd for $C_{39}H_{54}N_6O_6$,; 703.4046, found 703.4039.

A33-B59-C2-D10 (18). R_f = 0.24 (EtOAc). ^{1}H NMR ($CDCl_3$): δ 7.40−7.05 (m, 14H), 6.45 (bs, 2H), 5.20 (m, 3H), 4.55 (bs, 2H), 4.30 and 4.20 (m, 1H each), 3.95 (t, 1H, J = 6.6 Hz), 3.60 (bm, 2H), 3.40 (dd, 1H, J = 4.2 and 15.4 Hz), 3.05−2.40 (m, 7H), 2.05 (m, 1H), 0.87 and 0.75 (d, 3H each, J = 6.6 Hz, 0.63 (s, 9H). LRMS *m*/*z* 671 $(M+H)^+$. HRMS (FAB+) $[M+H]^+$ calcd for $C_{39}H_{50}N_4O_6$: 671.3806, found 671.3826.

A33-B60-C2-D10 (19). R_f = 0.21 (EtOAc). ^{1}H NMR ($CDCl_3$): δ 7.40−7.05 (m, 14H), 6.55 (d, 1H, J = 7.8 Hz), 6.18 (bs, 1H), 5.85 (s, 1H), 5.10 (m, 1H), 5.05 (s, 2H), 5.00−

4.70 (m, 2H), 4.20−4.00 (m, 3H), 3.95 (m, 1H), 3.75 (m, 1H), 3.61 (m, 1H), 3.00−2.80 (m, 4H), 2.60 (m, 1H), 2.40 (d, 1H, $J = 15.0$ Hz), 2.05 (m, 1H), 0.93 and 0.84 (doublets, 3H each, $J = 6.6$ Hz) 0.75 (s, 9H). LRMS *m*/*z* 671 $(M+H)^+$. HRMS (FAB+) $[M+H]^+$ calcd for $C_{39}H_{50}N_4O_6$: 671.3806, found: 671.3801.

A33-B61-C2-D10 (20). $R_f = 0.21$ (EtOAc). ^{1}H NMR ($CDCl_3$): δ 7.20−7.05 (m, 14H), 6.45 (d, 1H, $J = 7.8$ Hz), 6.38 (bs, 1H), 5.85 (s, 1H), 5.10−5.05 (m, 3H), 4.08 (m, 2H), 3.95 (dt, 1H, $J = 4.8$ and 7.5 Hz), 3.65 and 3.58 (m, 1H each), 3.25 (m, 1H), 2.90 (m, 5H), 2.45 (m, 2H), 2.05 (m, 1H), 0.89 and 0.78 (doublets, 3H each, $J = 6.6$ Hz), 0.78 (s, 9H). LRMS *m*/*z* 671 $(M+H)^+$. HRMS (FAB+) $[M+H]^+$ calcd for $C_{39}H_{50}N_4O_6$: 671.3806, found 671.3825.

A21-B46-C2-D10 (21). $R_f = 0.15$ (2% MeOH−EtOAc). ^{1}H NMR ($CDCl_3$): δ 7.20−7.05 (m, 13H), 6.85 (bs, 1H), 5.07 (bs, 2H), 44.76 (m 1H), 4.46 (m 1H), 4.37 (m 2H), 4.34 (m, 1H), 4.10−3.79 (m, 4H), 3.57 (m, 1H), 3.11 (m, 2H), 3.07 (m, 2H), 2.61−2.40 (m, 4H), 2.31−2.205 (m, 3H), 0.89 and 0.81 (doublets, 3H each, $J = 6.5$ Hz). LRMS *m*/*z* 775 $(M+H)^+$. HRMS (FAB+) $[M+H]^+$ calcd for $C_{40}H_{47}ClN_6O_8$: 775.3102, found 775.3109.

Library QA Analysis. Decoding. Beads (60−63 per sublibrary) were arrayed into 96-well filter-bottom plates as single-bead/well, and the compounds were photoeluted as described above. Tag cleavage was accomplished by treating the beads directly with 10 μL of an aqueous ceric ammonium nitrate solution (freshly prepared 0.3 M solution) and 50 μL of octane to cleave the tags. After 1 h at 25 °C, the octane extracts of tag alcohols (35 μL) were transferred into GC vials, and *N*,*O*-bis(trimethylsilyl)-acetamide (MSTFA; 5 μL) was added. The electron capture detection GC/ECD analysis of the tag trimethylsilyl ethers (1 μL injection) was carried out on a HP 6890 plus gas chromatograph (DB-1 column (J&W Scientific) 5 m × 0.1 mm i.d., 0.1 μm film). A temperature ramp of 125−325 °C was performed in 3 min, then held at 325 °C for 0.5 min. The EC detector was maintained at 350 °C, and the auxiliary gas was set at 60 psi. Automatic data processing of the GC/ECD trace against the registered encoded synthons generated the predicted structures, and molecular weights of each compound for the individual decoded beads. A total 1902 decodes were generated for the QA analysis.

LC/MS Analysis. The corresponding compounds in the dried assay plates were redissolved in 50 μL of acetonitrile−water (80:20) in preparation for LC/MS analysis. Mass spectrometry was performed using a PE SCIEX API 150EX single stage quadrupole mass spectrometer equipped with a turbo ESI source. Liquid chromatography was performed using a Hewlett-Packard model 1100 pump. A sample loop of 25 μL was injected via a PE 200 auto sampler. Separation was carried out on a Phenomenex Luna C18 column (30 × 3 mm i.d., 3 μm). The mobile phase consisted of water−acetonitrile (containing 0.02% trifluoroacetic acid) and was programmed as follows: 88% water to 10% water in 2.5 min, hold for 0.5 min, then 88% water for 2.0 min (reequilibration). The HPLC eluent (1 mL/min) was split 1:4 at the outlet of the column by means of a zero-dead-volume tee splitter resulting in the introduction of 20% of the HPLC eluent in the ESI interface. The electrospray voltage was set to 4.5 kV, and the capillary temperature was 350 °C. The sheath gas (nitrogen) pressure was adjusted to 40 psi and an auxiliary (turbo) gas (nitrogen, 6 L/min) was added in order to achieve nebulization. Custom software permitted rapid data processing and generation of found (F) and not found (NF) assignments.

Enzyme Assays. Plasmepsin II Assay.[37a] An assay mix was prepared containing 50 mM sodium acetate (pH 5.0), 1 mg/mL BSA, 0.01% Tween 20, 12.5% glycerol, 18% DMSO, and 12 μM plam II substrate.[37b] Twenty-five μL of the assay mix was added to each well of the 96-well microtiter plate containing dried down bead eluate or empty control wells. The plates were then sonicated to solublize the compounds. The reaction was initiated with the addition of 25 μL of 8 nM plm II which was in 50 mM sodium acetate (pH 5.0), 1 mg/mL BSA, 0.01% Tween 20, and 12.5% glycerol. The final concentrations were 4 nM plm II, 6 μM plm II substrate, 9% DMSO, 50 mM sodium acetate (pH 5.0), 1 mg/mL BSA, 0.01% Tween 20, and 12.5% glycerol. The reaction was incubated for 10 min at 25 °C and then quenched by the addition of 25 μL of 1 M Tris (pH 8.5) and 50% DMSO. The EDANS fluorescence was measured using the Tecan, SLT FluoStar fluorescence plate reader with an excitation filter of 350 nm and an emission filter of 510 nm. To determine kinetic parameters, identical assays were used except the substrate and inhibitor concentrations were varied and the reactions were monitored over time. Values for K_i were determined from slope replots of the Lineweaver−Burke analysis.

Cathepsin D Assay.[37c] The assay for cat D was performed in a manner similar to the plm II assay except that the final concentrations were 0.8 nM cat D, 6 μM cat D substrate, 6% DMSO, 25 mM sodium formate (pH 3.5), and 1 mg/mL BSA.

References and Notes

(1) (a) Dolle, R. E. *J. Combin. Chem.* **2000**, *2*, 383−433. (b) Dolle, R. E.; Nelson, K. H., Jr. *J. Comb. Chem.* **1999**, *1*, 235−282. (c) Dolle, R. E. *Mol. Diversity* **1998**, *3*, 199−233. (c) Dolle, R. E. *Mol. Diversity* **1998**, *4*, 233−256.

(2) For reviews and descriptions of deconvolution methods, see: (a) Schneider, S. E.; Anslyn, E. V. *Adv. Supramol. Chem.* **1999** *5*, 55−120. (b) Furka, A.; Bennett, W. D. *Comb. Chem. High Throughput Screen.* **1999**, *2*, 105−22. (c) Baldwin, J. J.; Dolle, R. E. *Annu. Rep. Comb. Chem. Mol. Diversity* **1997**, *1*, 287−297.

(3) (a) Baldwin, J. J. *Mol Diversity* **1996**, *2*, 81−88. (b) Ohlmeyer, M. H. J.; Swanson, R. N.; Dillard, L. W.; Reader, J. C.; Asouline, G.; Kobayashi, R.; Wigler, M.; Still, W. C. *Proc. Natl. Acad. Sci., U.S.A.* **1993**, *90*, 10922. (c) Borchardt, A.; Still, W. C. *J. Am. Chem. Soc.* **1994**, *116*, 373−374. (d) Seneci, P. In *Comb. Chem. Technol.*; Miertus, S., Fassina, G., Eds.; Decker: New York, 1999; pp 127−167. (e) Shchepinov, M. S.; Chalk, R.; Southern, E. M. *Tetrahedron* **2000**, *56*, 2713−2724. (f) Scott, R. H.; Barnes, C.; Gerhard, U.; Balasubramanian, S. *Chem. Commun.* **1999**, 1331−1332. (g) Barnes, C.; Scott, R. H.; Balasubramanian, S. *Recent Res. Dev. Org. Chem.* **1998**, *2*, 367−379. (h) Hochlowski, J. E.; Whittern, D. N.; Sowin, T. J. *J. Comb. Chem.* **1999**, *1*, 291−293. (i) Hughes, I. *J. Med. Chem.* **1998**, *41*, 3804−3811. (j) Fitch, W. L.; Baer, T. A.; Chen, W.; Holden, F.; Holmes, C. P.; Maclean, D.; Shah, N.; Sullivan, E.; Tang, M.; Waybourn, P.; Fischer, S. M.; Miller, C. A.; Snyder, L. R.

J. Comb. Chem. **1999**, *1*, 188−194. (k) Wagner, D. S.; Markworth, C. J.; Wagner, C. D.; Schoenen, F. J.; Rewarts, C. E.; Kay, B. K.; Geysen, H. M. *Comb. Chem. High Throughput Screening* **1998**, *1*, 143−153. (l) Xiao, X.-U.; Nova, M. P. In *Combinatorial Chemistry*; Wilson, S. R., Czarnik, A. W., Eds.; Wiley: New York, 1997; pp 135−152. (m) Rahman, S. S.; Busby, D. J.; Lee, D. C. *J. Org. Chem.* **1998**, *63*, 6196−6199. (n) Guiles, J. W.; Lanter, C. L.; Rivero, R. A. *Angew. Chem. Int. Ed.* **1998**, *37*, 926−928. (o) Battersby, B. J.; Bryant, D.; Meutermans, W.; Matthews, D.; Smythe, M. L.; Trau, M. *J. Am. Chem. Soc.* **2000**, *122*, 2138−2139. (p) Lane, S.; Pipe, A. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 798−814. (q) Silen, J. L.; Lu, A. T.; Solas, D. W.; Gore, M. A.; Maclean, D.; Shah, N. H.; Coffin, J. M.; Bhinderwala, N. S.; Wang, Y.; Tsutsui, K. T.; Look, G. C.; Campbell, D. A.; Hale, R. L.; Navre, M.; DeLuca-Flaherty, C. R. *Antimicrob. Agents Chemother.* **1998**, *42*, 1447−1453. (r) Schwabcher, A. W.; Shen, Y.; Johnson, C. C. *J. Am. Chem. Soc.* **1999**, *121*, 8669−6671. (s) Furka, A.; Christensen, J. W.; Healy, E.; Tanner, H. R.; Saneii, H. *J. Comb. Chem.* **2000**, *2*, 220−223.

(4) (a) Baldwin, J. J.; Burbaum, J. J.; Henderson, I.; Ohlmeyer, M. H. J. *J. Am. Chem. Soc*. **1995**, *117*, 5588−5589. (b) McMillan, K.; Adler, M.; Auld, D. S.; Baldwin, J. J.; Blasko, E.; Browne, L. J.; Chelsky, D.; Davey, D.; Dolle, R. E.; Eagen, K. A.; Erickson, S.; Fledman, R. I.; Glaser, C. B.; Mallari, C.; Morrissey, M. M.; Ohlmeyer, M. H. J.; Pan, G.; Parkinson, J. F.; Phillips, G. B.; Polokoff, M. A.; Sigal, N. H.; Vergona, R.; Whitlow, M.; Young, T. A.; Devlin, J. J. *Proc. Natl. Acad. Sci., U.S.A.* **2000**, *97*, 1506−1529. (c) Horlick, R. A.; Ohlmeyer, M. H.; Stroke, I. L.; Strohl, B.; Pan, G.; Schilling, A.; Paradkar, V.; Quintero, J. G.; You, M.; Riviello, C.; Thorn, M. B.; Damaj, B.; Fitzpartick, V. D.; Dolle, R. E.; Webb, M. L.; Baldwin, J. J.; Sigal, N. H.. *Immunopharm.* **1999**, *43*, 169−177.

(5) (a) Baldwin, J. J.; Horlbeck, E. G. U.S. Patent 5 663 046, 1997.

(6) For a description of Pharmacopeia's library quality control procedures see: Dolle, R. E. In *Optimization of Solid Phase Combinatorial Synthesis*, Yan, B., Czarnik, A. W., Eds.; Marcel Dekker: New York, in press.

(7) For approaches to the analytical characterization of large mixture libraries see: (a) Srebalus, C. A.; Li, J. *Am. Soc. Mass Spectrom.* **2000**, *11*, 352−355. (b) Loo, J. A. *Eur. J. Mass Spectrom.* **1997**, *3*, 93−104. (c) Berlin, K.; Jain, R. J.; Tetzlaff, C.; Steinbeck, C.; Richert, C. *Chem. Bio.* **1997**, *4*, 63−77. (d) Nawrocki, J. P.; Wigger, M.; Watson, C. H.; Hays, T. W.; Senko, M. W.; Benner, S. A.; Eyler, J. R. *Rapid. Commun. Mass Spectrom*. **1996**, *10*, 1860−1864. (e) Dunyevskiy, Y. M.; Vouros, P.; Wintner, E. A.; Shipps, G. W.; Carell, T.; Rebek, J., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6152−6157. (f) Promerantz, S. C.; McCloskey, J. A.; Tarasow, T. M.; Eaton, B. E. *J. Am. Chem. Soc*. **1997**, *19*, 3861−3867.

(8) (a) The use of tag decode-assisted single bead LC/MS for library quality control was described by researchers at Glaxo Wellcome and the Affymax Research Institute: Lewis, K. C.; Fitch, W. L.; Maclean, D. *LCGC* **1998**, *16*, 644−649. (b) In this account by Lewis, the quality control of a 42 875-member thiazolidinone library was described via tag decode-assisted single bead LC/MS analysis. The library consisted of 3 building block sets with 35 members in each set (35^3). The number of beads sampled was n = three times the largest synthon set, or 105 beads/pool (3×35). This value of n was thought to be the "best compromise between analysis time [25 min per sample (although decode analysis speed has been improved, see ref 3j)] and gathering the information for library evaluation." As per our statistical computation, this sample size is sufficient for rendering judgment at the level of the library pools, even when $p = 50\%$, *lb-ub* = 40−60% (see $n = 100$ in Tables 1 and 2). However, at the level of individual synthons (p_i) an average of $n_i = 3$ leads to a high degree of uncertainty in data interpretation. When $n_i = 3$, and 3 out of 3 beads confirm, then $p_i = 100\%$, *lb-ub* = 36.8−100% ($\alpha = 0.05$); when 2 out of 3 beads confirm, then $p_i = 66.7\%$, *lb-ub* = 7.9−100% ($\alpha = 0.05$).

(9) (a) Egner, B. J.; Bradley, M. *Drug Discovery Today* **1997**, *2*, 102−109. (b) Gallop, M. A.; Fitch, W. L. *Curr. Opin. Chem. Bio.* **1997**, *1*, 94−100. (c) Fitch, W. L. *Mol. Diversity* **1999**, *4*, 39−45. (d) Snyder, L. R.; Kirkland, J. J.; Glajch, J. L. *Practical HPLC Method Development*, 2nd ed.; John Wiley & Sons: New York, 1997.

(10) (a) Kyranos, J. N.; Hogan, J. C., Jr. *Modern Drug Discovery* **1999**, 73−81. (a) Kyranos J. N.; Hogan, J. C., Jr. *Anal. Chem. News Features* **1998**, 389A−395A. (b) Patek, M.; Drake, B.; Labl, M. *Tetrahedron Lett.* **1995**, 36, 2227−2230. (c) Ostresh, J. M.; Husar, G. M.; Blondelle, S. E.; Dorner, B.; Weber, P. A.; Houghten, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11138−11142.

(11) (a) Kibbey, C. E. *Mol. Diversity* **1995**, 1, 247−258. (b) Hsu, B. H.; Orton, E.; Tang, S.-Y.; Carlton, R. A. *J. Chromatogr. B* **1999**, *725*, 103−112. (c) Lebl. M.; Krchnak, V.; Ibrahim, G.; Pires, J.; Burger, C.; Ni, Y.; Chen, Y.; Podue, D.; Mudra, P.; Pokorny, V.; Poncar, P.; Zenisek, K. *Synthesis* **1999**, 1971−1978. (d) Fang, L.; Wan, M.; Pennacchio, M.; Pan, J. *J. Comb. Chem.* **2000**, *2*, 254−257.

(12) (a) Borney, J.-F. A.; Homen, M. E. *LCGC* **2000**, *18*, S14−S19. (b) Taylor, E. W.; Qian, M. G.; Dollinger, G. D. *Anal. Chem.* **1998**, *70*, 3339−3347.

(13) (a) Sage, A. B.; Little, D.; Giles, K. *LCGC* **2000**, *18*, S20−S29. (b) Sussmuth, R. D.; Jung, G. *J. Chromatogr. B* **1999**, *725*, 49−65. (c) Marshall. P. S. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 778−781. (d) Dulery, B. D.; Verne-Mismer, J.; Wolf, E.; Kugel, C.; Van Hijfte, L. *J. Chromatogr. B* **1999**, *725*, 39−47. (e) Fitzgerald, M. C.; Harris, K.; Shevlin, C. G.; Siuzdak, G. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 979−982. (f) Richmond, R.; Gorlach, E.; Seifert, J.-M. *J. Chromatogr. A* **1999**, *835*, 29−39. (g) Richmond, R.; Gorlach, E. *Anal. Chim. Acta* **1999**, *390*, 175−183. (h) Richmond, R.; Gorlach, E. *Anal. Chim. Acta* **1999**, *394*, 33−42.

(14) Keifer, P. A.; Smallcombe, S. H.; Williams, E. H.; Salomon, K. E.; Mendez, G.; Belletire, J. L.; Moore, C. D. *J. Comb. Chem.* **2000**, *2*, 151−171 and references therein.

(15) (a) Schultz, L.; Garr, C. D.; Cameron, L. M.; Bukowski, J. *Bioorg. Med. Chem. Lett*. **1998**, *8*, 2409−2414. (b) Zeng, L.; Burton, L.; Yung, K.; Shushan, B.; Kassel, D. B. *J. Chromatogr. A* **1998**, *794*, 3−13.

(16) Thompson, S. K. In *Sampling*; John Wiley & Sons: New York, 1992.

(17) Berry, D. A.; Lindgren, B. W. *Statistics: Theory and Methods*; Pacific Grove: Brooks/Cole Publishing Company, 1990; Chapter 4.

(18) Clopper, C. J.; Pearson, E. S. *Biometrika* **1934**, *26*, 404−413.

(19) The detection limit (2.5 pmol total injection) and the detector linear range (2.5∼175 pmol) used in the comparative QA analysis were established using the set of quality control compounds from library **1** (data not shown). For an example of using a set of structurally related compounds for calibration, see: Lai, C.-C.; Tsai, P.-L.; Yu, C.; Her, G.-R. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 486-475 and references therein. Further details regarding the use of library QC compounds as the calibration set for the mass spectral analysis will be reported elsewhere.

(20) The small quantities of cleaved compounds obtained from the beads prohibited the direct use of TIC chromatograms (ref 8) to confirm putative library compounds. This is due to the interference of bead matrix present in the reconstituted samples. The use of XIC chromatograms is standard practice in drug metabolism where biological matrix may obscure trace compounds of interest.

(21) (a) The criterion for positive identification (F) was the peak height in the XIC chromatogram must be >4.0E+04 (for this library) and the observed molecular weight equal to $(M+H)^+$. If the signal-to-noise (S/N) ratio of the peak defined by the molecular weight in the XIC chromatogram was less than 3:1, the targeted molecular ion was not observed and recorded as not found (NF). (b) For ca. 2−3% of the F answers, the S/N of the peak in the XIC chromatogram was <3:1 and the peak height was in the lower end of the detector linear range (<4.0E+04), or the peak was a minor component of the XIC chromatogram.

(22) (a) Abdel-Meguid, S. S. *Med. Res. Rev.* **1993**, *13*, 731−778. (b) Rich, D. H.; Francesco, G.; Salituro, G.; Holladay, M. W. In *Conformationally Directed Drug Design*; Vida, J. A., Gordon, M., Eds.; ACS Symposium Series 251; American Chemical Society: Washington, DC, 1984; Chapter 10. (c) Barrett, A. J. In *Proteinases in Mammalian Cells and Tissues*; Barrett, A. J., Ed.; Elsevier/North-Holland Biomedical Press: New York, 1977; pp 209−248.

(23) (a) Takahashi, T.; Tang, J. *Methods Enzymol.* **1981**, *80*, 565−581. (b) Baldwin, E. T.; Baht, T. N.; Gulnik, S.; Hosur, M. V.; Sowder, R. C., Jr.; Cachau, R. E.; Collins, J.; Silva, A. M.; Erickson, J. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6796−6800.

(24) (a) Silva, A. M.; Lee, A. Y.; Gulnik, S. V.; Majer, P.; Collins, J.; Bhat, T. N.; Collins, P. J.; Cachau, R. E.; Luker, K. E.; Gluzman, I. Y.; Francis, S. E.; Oksman, A.; Goldberg, D. E.; Erickson, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10034−10039. (b) Moon, R. P.; Tyas, L.; Certa, U.; Rupp, K.; Bur, D.; Jacquet, C.; Matile, H.; Loetscher, H.; Grueninger-Leitch, F.; Kay, J.; Dunn, B. D.; Berry, C.; Ridley, R. G. *Eur. J. Biochem.* **1997**, *244*, 552−560.

(25) Library **1** was designed to explore the $P_{2'}$ specificity of aspartic acid proteases, namely cat D and plm II. This library complements previous encoded libraries of statine amides defining the P_1 and P_2 specificity for these enzymes: (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.; Dolle, R. E. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2315−2320. (b) Carroll, C. D.; Johnson, T. O.; Tao, Shewei; Lauri, G.; Orlowski, M.; Gluzman, I. Y.; Goldbert, D. E.; Dolle, R. E. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 3203−3206.

(26) Plasmepsin II has been the target for other libraries containing aspartic acid transition-state isosteres: Haque, T. S.; Skillman, G.; Lee, C. E.; Habashita, H.; Gluzman, I. Y.; Ewing, T. J. A.; Goldberg, D. E.; Kuntz, I. D.; Ellman, J. A. *J. Med. Chem.* **1999**, *42*, 1428−1440.

(27) Two different analysts carried out the library QA analyses over a four month period. Each QA (ca. 650 beads) required approximately 32 h total GC time (3 min per decode) and approximately 55 h total of LC/MS time (5 min per sample).

(28) Between the 1902 decoded beads in the three QA runs and the 131 decodes obtained from screening, there were 16 overlapping structures identified. In each case, there was a 100% correlation between these biologically active compounds and a found (F) designation in the QA analysis adding an additional level of validation for the library QA analysis (data not shown).

(29) Calculated logP (clogP) for selected synthons: 0.1, **A6**; −0.1, **A8**; 0.27, **A11**; 1.0, **A14**; −0.4, **A19**; 1.8, **A21**; 1.4, **A23**; −0.5, **A32**; 1.2, **A33**.

(30) In support of the successful attachment of **A6** to resin, in all cases where the **A6**-containing compounds were NF in the QA analysis, the corresponding sulfoxide $(M+OH)^+$ was found as a major product in the respective XIC chromatograms.

(31) Factors influencing F and NF comparative answers well appreciated. For example, the ability of ESI to produce polycharged ions (e.g., $M+2H)^+$ is well known in peptides having two or more basic residues in acidic solution (see refs 13). Thus library compounds with multiple basic synthons, if double protonated, would be recorded as NF. Cationization, solvent-adduct formation, ESI-induced oxidation, and ESI-discrimination effects (i.e., differences in ionization efficiencies for competing compounds) can also lead to NF answers.

(32) The tendency to see higher overall performance of the hydrophobic synthons and a higher confirmation rate of lipophilic compounds per se may in part be an artifact of the LC/MS analysis. For example, LC is conducted employing a generic 5 min gradient on a C18 column designed to capture 95+% of the library compounds (see Experimental Section). Compounds having multiple charged groups (e.g., the putative compound **A8-B14-C1-D4**) may be not be retained on the column and hence register as NF.

(33) A statistical sampling QA analysis (609 beads) was carried out on the tagged Fmoc-intermediate **5**: $p = 96.1\%$. UV data was also obtained for many of the intermediates, which showed >85% purity (unpublished observation). Specifically, synthon **B46** gave $p = 100\%$ ($n = 5$), indicating chemical failure post coupling.

(34) Chemical yield of resynthesized **21** was 10% versus the average 50% yields obtained for the other resynthesized compounds in Table 6.

(35) A series of -X-Pro-Phe-Phe-Val-D-Leu- synthetic peptides where -Phe-Phe- represents the P_1 and $P_{1'}$ residues, respectively, are reported as inhibitors of cathepsin D: Lin, T.-Y.; Williams, H. R. *J. Biol. Chem.* **1979**, *254*, 11875−11883. For other studies mapping out subsite specficity for statine-containing inhibitors of cathepsin D, see: (a) Majer, P.; Collins, J. R.; Gulnik, S. V.; Erickson, J. W. *Protein Sci.* **1997**, *6*, 1458−1466. (b) Agarwal, N. S.; Rich, D. H. *J. Med. Chem.* **1986**, *29*, 2519−2524. (c) Greenlee, W. *Med. Res. Rev.* **1990**, *10*, 173−236. (d) Jupp, R. A.; Dunn, B. M.; Jacobs, J. W.; Vlasuk, G.; Arcuri, K. E.; Veber, D. F.; Perlow, D. S.; Payne, L. S.; Poger, J.; de Laszlo, S.; Chakravarty, P. K.; tenBroeke, J.; Hangauer, D. G.; Ondeyka, D.; Greenlee, W. J.; Kay, J. *Biochem. J.* **1990**, *265*, 871−878. (e) Rao, C. M.; Scarborough, P. E.; Kay, J.; Batley, B.; Rapundalo, S.; Klutchko, S.; Taylor, M. D.; Lunney, E. A.; Humblet, C. C.; Dunn, B. M. *J. Med. Chem.* **1993**, *36*, 2614−2620. (f) Scarborough, P. E.; Guruprasad, K.; Topham, C.; Richo, G. R.; Conner, G. E.; Blundell, T. L.; Dunn, B. M. *Protein Sci.* **1993**, *2*, 264−276.

(36) These amines are fully protonated in the assay buffer, pH = 3.5, cathepsin D assay; pH = 5.0, plm II.

(37) (a) Luker, K. E.; Francis, S. E.; Gluzman, I. Y.; Goldberg, D. E. *Mol. Biochem. Parasitol.* **1996**, *79*, 71−78. Plm II was a gift from Dr. Daniel E. Goldberg, Howard Hughes Medical Institute, Washington University School of Medicine, St. Louis, MO, 63110. (b) The plasmepsin II substrate is DABCYL-γ-aminobutyric acid-Glu-Arg-Met-Phe-Leu-Ser-Phe-Pro-EDANS, a custom synthesis from AnaSpec, Inc., 2149 O'Toole Avenue, Suite F, San Jose, CA 95131. (c) Cat D from human liver was purchased from ART Biochemicals, Athens Research Technology, P.O. Box 5494, Athens, GA 30604. The cat D substrate, Ac-Glu-Glu(EDANS)-Lys-Pro-Ile-Met-Phe-Phe-Arg-Leu-Gly-Lys(DABCYL)-Glu-NH_2, was purchased from AnaSpec, Inc.

CC000052K