Anal. Chem. **2004,** *76,* 7278−7287

Streamlined System for Purifying and Quantifying a Diverse Library of Compounds and the Effect of Compound Concentration Measurements on the Accurate Interpretation of Biological Assay Results

Ioana G. Popa-Burke,*,† Olga Issakova,‡ James D. Arroway,† Paul Bernasconi,† Min Chen,† Louis Coudurier,† Scott Galasinski,† Ajit P. Jadhav,† William P. Janzen,† Dennis Lagasca,‡ Darren Liu,‡ Roderic S. Lewis,† Robert P. Mohney,† Nikolai Sepetov,‡ Darren A. Sparkman,† and C. Nicholas Hodge†

Amphora Discovery Corporation, P.O. Box 12169, Research Triangle Park, North Carolina 27709, and Nanosyn, Inc. 3760 Haven Avenue, Menlo Park, California 94025

As part of an overall systems approach to generating highly accurate screening data across large numbers of compounds and biological targets, we have developed and implemented streamlined methods for purifying and quantitating compounds at various stages of the screening process, coupled with automated "traditional" storage methods (DMSO, −20 °C). Specifically, all of the compounds in our druglike library are purified by LC/MS/UV and are then controlled for identity and concentration in their respective DMSO stock solutions by chemiluminescent nitrogen detection (CLND)/evaporative light scattering detection (ELSD) and MS/UV. In addition, the compound−buffer solutions used in the various biological assays are quantitated by LC/UV/CLND to determine the concentration of compound actually present during screening. Our results show that LC/UV/CLND/ELSD/MS is a widely applicable method that can be used to purify, quantitate, and identify most small organic molecules from compound libraries. The LC/UV/CLND technique is a simple and sensitive method that can be easily and cost-effectively employed to rapidly determine the concentrations of even small amounts of any N-containing compound in aqueous solution. We present data to establish error limits for concentration determination that are well within the overall variability of the screening process. This study demonstrates that there is a significant difference between the predicted amount of soluble compound from stock DMSO solutions following dilution into assay buffer and the actual amount present in assay buffer solutions, even at the low concentrations employed for the assays. We also demonstrate that knowledge of the concentrations of compounds to which the biological target is exposed is critical for accurate potency determinations. Accurate potency values are in turn particularly important for drug discovery, for understanding structure−activity relationships, and for building useful empirical models of protein−ligand interactions. Our new understanding of relative solubility demonstrates that most, if not all, decisions that are made in early discovery are based upon missing or inaccurate information. Finally, we demonstrate that careful control of compound handling and concentration, coupled with accurate assay methods, allows the use of both positive and negative data in analyzing screening data sets for structure−activity relationships that determine potency and selectivity.

With the advent of combinatorial and parallel chemistry, high-throughput screening (HTS) has become the major tool employed in the discovery of biologically active compounds—drugs, chemicals, pesticides, etc. Significant progress has been made in improving HTS methods over the past decade − assays have become more reliable and measurement variability has been greatly reduced.

Another important factor which impacts the quality of the results from any HTS assay is the collection or library of compounds. The quality of the compound collection screened is paramount and has been discussed in many published reports.[1−10] From an analytical perspective, the purity of a compound collection

* Corresponding author. Phone: (919) 287-6209. Fax: (919) 806-3477. E-mail: Ioana.Popa-Burke@amphoracorp.com.

† Amphora Discovery Corp.

‡ Nanosyn, Inc.

(1) Ajay, W. W. P.; Murcko, M. A. *J. Med. Chem.* **1998**, *41*, 3314−3324.
(2) Nilakantan, R.; Immermann, F.; Haraki, K. *Comb. Chem. High Throughput Screening* **2001**, *5* (2), 105−110.
(3) Jenkins, J. L.; Kao, R. Y. T.; Shapiro, R. *Proteins: Struct., Funct., Genet.* **2003**, *50*, 81−93.
(4) Miller, M. A. *Nat. Rev. Drug Discovery* **2002**, *1*, 220−227.
(5) Lipinski, C. A. *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 235−249.
(6) Lipinski, C. A. *Drug Discovery Today* **2003**, *8* (1), 12−16.
(7) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Delivery Rev.* **1997**, *23*, 3−25.
(8) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Delivery Rev.* **2001**, *46*, 3−26.
(9) Janzen, W.; Bernasconi, P.; Cheatham, L.; Mansky, P.; Popa-Burke, I.; Williams, K.; Worley, J.; Hodge, N. *Optimizing the Chemical Genomics Process*; Marcel Dekker: New York, 2004.

10.1021/ac0491859 CCC: $27.50
Published on Web 11/10/2004

is now being addressed by many laboratories, typically by MS-based methods either for the characterization or purification of compounds.[11−16] In our laboratories, all of the library compounds are purified using LC/MS, where fractionation is based on the molecular ion of the target compound.[17,18] Compounds are then controlled for identity and quantitated in their respective DMSO stock solutions by flow injection analysis (FIA) MS/UV/chemiluminescent nitrogen detection (CLND)/evaporative light scattering detection (ELSD).

Most biological assays tolerate 1% or less of DMSO in the final aqueous medium used. In contrast to the DMSO storage solutions, the aqueous assay solutions are short-lived (compounds are analyzed and the solutions are then disposed of). While longer term stability may not be a problem, initial solubility and stability can be. While several reports have addressed DMSO solubility and stability, it is surprising that there is almost no information available on the fate of library compounds in aqueous solutions. This probably results from a combination of factors, including the complexity and cost of the analysis, but the fact that library purity and stability are problems that have been relatively recently addressed must lie at the core since this analysis is impossible until that baseline is established. Clearly there is no shortage of analyses of small numbers of druglike compounds in aqueous medium. However, to be relevant to the HTS paradigm, many sample measurements of small volumes stored over long time periods must be performed to achieve statistical significance. Perhaps the most important factor enabling our analysis has been the recent development of stable, sensitive detectors suitable for use in an industrialized setting. These methods, particularly CLND, allow direct determination of concentration within acceptable error limits (as described herein). Earlier attempts using HPLC and MS or UV detection did not provide the needed combination of speed, throughput, and linearity over the low concentrations present in aqueous solution.

To address the problems of DMSO solubility, DMSO stability, storage, aqueous solubility, and aqueous stability, attempts can be made to correct for each issue individually. Many improvements have been made in each of these steps[19] but there is no such thing as an "ideal solution". We propose another approach, which starts with known concentrations of pure compounds, uses existing (and relatively standard) compound storage and handling solutions, and measures the actual amount of compound screened. This approach can be applied to anything from single compounds to large screening libraries.

This ongoing study measures the concentrations in DMSO and buffer for purified compounds from a large compound collection. A comparison of the concentrations measured in the assay plates diluted with buffer with the initial concentrations of the compounds in the stock DMSO plates is shown. These data highlight the significant difference between DMSO and aqueous solubility of compounds, even at the low concentrations used for screening.

The solubility and stability of compounds were measured concurrent with screening. This allowed for correction of the IC_{50} and K_i values to their true concentration in buffer. Here we present some of the implications these data have on biological results, emphasizing the effects on HTS screening interpretation for drug discovery programs.

EXPERIMENTAL SECTION

Materials. Standards used for method development and proof-of-principle CLND analysis (propranolol, caffeine, verapamil, terfenadine, staurosporine), as well as trifluoroacetic acid (TFA) and sodium mono- and diphosphate, were purchased from Sigma-Aldrich (St Louis, MO) with a stated purity of >99%. HPLC grade methanol (MeOH), isopropyl alcohol, and high-purity DMSO were purchased from Burdick & Jackson (Muskegon, MI). (Mark Homan, personal communication: B&J HPLC grade MeOH has been tested and found to have suitably low background levels of nitrogen for use with the chemiluminescent nitrogen detector.) All water was purified through a Milli-Q purification system (Millipore Corp., Bedford, MA).

The library employed in the present study grew to a final size of ~130 000 unique molecules. We required a library that would both provide multiple active molecules against a broad range of biological targets and permit rapid optimization of these "hits" into bona fide drug leads. To accomplish this end, we used known and internally developed methods of diversity analysis, functional group matching with known drug libraries, and "hole-filling" techniques. Description of the library building methods will be published elsewhere, but the overall charasteristics can be summarized briefly in terms of accepted "rule of five" properties[7] (see also Supporting Information).

The calculated log P values of the final library display a remarkably symmetric normal distribution around the center maximum of 3.4−3.6, using bin sizes of 0.2 unit. In this case, the maximum bin contained 6182 molecules. The molecular weights show a slightly skewed distribution toward higher values, since we excluded compounds of MW < 180 except for some known small biologically active molecules. Using bins of 20 Da, the curve shows a flattened maximum between MW 340−380 and contains ~23 000 molecules. Total polar surface area also shows a slightly skewed distribution to the higher values, with a maximum between 56 and 64 units (8-unit bins). Number of rotatable bonds and number of hydrogen bond acceptors per molecule both have roughly normal distribution centered at eight and three, respectively. The number of hydrogen bond donors per molecule is heavily weighted toward values of zero, one (maximum) and two, and drops off sharply at higher values. Graphical plots of the histograms for each of these parameters are included in the Supporting Information.

(10) Walters, W. P.; Murcko, M. A. *Adv. Drug Delivery Rev.* **2002**, *54*, 255−271.
(11) Taylor, E. W.; Qian, M. G.; Dollinger, G. D. *Anal. Chem.* **1998**, *70*, 3339−3347.
(12) Yan, B.; Fang, L. L.; Irving, M.; Zhang, S.; Boldi, A. M.; Woolard, F.; Johnson, C. R.; Kshirsagar, T.; Figliozzi, G. M.; Krueger, C. A.; Collins, N. *J. Comb. Chem.* **2003**, *5* (5), 547−559.
(13) Lambert, P. H.; Bertin, S.; Fauchere, J. L.; Volland, J. P. *Comb. Chem. High Throughput Screening* **2001**, *4* (4), 317−332.
(14) Chan, E. C. Y.; Wang, A. W.; Jia, W. P.; Dollinger, G. D. *J. Pharm. Biomed. Anal.* **2002**, *29* (1−2), 139−146.
(15) Bronstrup, M. *Mod. Mass Spectrom. Top. Curr. Chem.* **2003**, *225*, 283−302.
(16) Felder, E. R.; Martina, K.; Scarpella, S.; Tato, M. *Chimia* **2003**, *57* (5), 229−236.
(17) Zeng, L.; Burton, L.; Yung, K.; Shushan, B.; Kassel, D. B. *J. Chromatogr., A* **1998**, 794, 3−13.
(18) Kassel, D. B. *Chem. Rev.* **2001**, 101, 255−267.
(19) Janzen, W. P., Ed. *High Throughput Screening: Methods and Protocols*; Methods in Molecular Biology 190; Humana Press: Totowa, NJ, 2002.

Prepurification Flow Injection Mass Spectrometry Analysis (FIA/MS). All Prepurification FIA/MS work was performed in high-throughput mode on a system comprising an SCL-10A system controller (Shimadzu, Tokyo, Japan), two LC-10ADvp HPLC pumps (Shimadzu), a LEAP HTS Pal autosampler (CTC Analytics, Zwingen, Switzerland), an SPD-10ADvp UV detector (Shimadzu), an electrospray single quadrupole mass spectrometer (API-100, API 150, or API 165, AB/MDL Sciex, Foster City, CA), a Gilson 204 or Gilson 215 (Gilson Inc.) fraction collector, a Gilson 215 liquid handler (Gilson Inc.), a Macintosh G4 or G3 processor, and an evaporative light scattering detector ELSD SEDEX 75 (Sedere).

Samples for injection were prepared using a Gilson 215 liquid handler by adding appropriate volumes of DMSO directly to the plates received from a vendor to achieve 15 mM. Since DMSO easily absorbs water from the air and can cause compounds to precipitate, a proprietary nitrogen-filled enclosure for the Gilson 215 liquid handler was used. For the FIA, an injection volume of 2 μL was used. Injections were performed directly from stock plates. The mobile phases (phase A, 100% water, 0.1% TFA; phase B, 100% acetonitrile, 0.12% TFA) were delivered to the injector by the HPLC pumps running in isocratic mode: 70% B for 40 s, cycle time ~1 min/sample. The flow rate was 200 μL/min. Acquired data were processed by Nanosyn's proprietary software, which checked for a compound identity and amount of targeted ion and automatically reported the samples that were not suitable for LC/MS purification.

Purification. All purification work utilized automated reversed-phase high-performance liquid chromatography/mass spectrometry (HPLC/MS). Each system comprised an SCL-10A system controller, two LC-10ADvp HPLC pumps, a LEAP HTS Pal autosampler, an SPD-10ADvp UV detector, an electrospray single quadrupole mass spectrometer (API-100, API 150, or API 165), a Gilson 204 or Gilson 215 fraction collector, and a Macintosh G4 or G3 processor.

All hardware components were used as received from the vendors. Compounds that passed the prepurification step were rearrayed in 96-well plates by transferring 65 μL of the 15 mM solutions, using the Gilson 215 liquid handler. This step was performed in the nitrogen-filled enclosure, and plates were immediately heat-sealed to ensure no water absorption by DMSO. An injection volume of 60 μL was used. Approximately ~900 nmol of each compound was injected. Fraction collection was triggered by the molecular ion of the targeted compound. All purification work was performed in analytical mode utilizing a Chromolith SpeedRod RP-18e C18 analytical column (4.6 mm × 50 mm) from Phenomenex. To increase throughput, multiple LC/MS systems (up to four at a time) were used for purification.

The binary gradient mobile phases (phase A, 100% water, 0.1% TFA; phase B, 100% acetonitrile, 0.1% TFA) were delivered to the injector by HPLC pumps controlled by a Shimadzu controller running the following gradient conditions: 0% B for 0.2 min, 0% B to 100% B in 2.5 min, a hold at 100% B for 1 min, and reequilibration for 0.3 min. The flow rate was 4 mL/min. The column effluent was split into a detection stream and a collection stream in a ratio of 1:40. The detector stream passed through UV detectors detecting at 215 and 254 nm and then were directed to the ion source of mass spectrometer. The collection stream was sent either to a fraction collector or to waste being controlled by the vendor-provided Applescript FC script 2.0. The Applescript triggered the start of collection when the extracted ion chromatogram corresponding to the expected mass of the sample reached the user-defined threshold. Nanosyn's proprietary process of assigning a threshold to each compound was used, which includes prepurification analysis by flow injection mass spectrometry, followed by automated processing of the data, resulting in the appropriate threshold for each compound. The maximum collection time was calculated based on the time required to fill 80% of the collection vessel.

A test sample of 15 mM fluorescein (Sigma-Aldrich, St. Louis, MO) was injected for purification after every 48 samples analyzed. A reproducible signal from the test sample both on UV and MS was considered as evidence of high performance of the analytical instrumentation.

DMSO Quantitation. Upon completion of the purification sample queue, fractions were evaporated using a Genevac HT-12 evaporator (Genevac Technology) or a Virtis Genesis 25EL freeze-drier (Vertis, Gardiner, NY) for heat-sensitive samples. Dried compounds were then redissolved using a Beckman Multimek 96 automated 96-channel pipettor (Beckman Instruments, Fullerton, CA) in 300 μL of DMSO, shaken for ~6 h, and centrifuged. Following redissolution, compounds were analyzed by flow injection MS/ELSD/CLND analysis. The system for analysis comprised an SCL-10A system controller, an LC-10ADvp HPLC pump, a Gilson 215 autosampler, an electrospray single quadrupole mass spectrometer (API 150), a chemiluminescent nitrogen detector CLND 8060-M (Antek Instruments, Houston, TX), and an evaporative light scattering detector ELSD SEDEX 75 (Sedere).

The settings for the CLND were as follows: photomultiplier tube (PMT) voltage −750 V, sensitivity high × 10, argon 100 cm^3/min, oxygen 200 cm^3/min, ozone 25 cm^3/min, and oven temperature 1050 °C. Since DMSO easily absorbs water from the air and can cause compounds to precipitate, a proprietary nitrogen-filled enclosure for the autosampler was used. A single mobile phase (premixed 50% MeOH, 50% H_2O, 0.1% TFA) was delivered to the injector at a flow rate of 400 μL/min by the HPLC pump controlled by the Shimadzu controller. Following the injection, the stream was split in a 1:1:2 ratio going to CLND, ELSD, and MS, respectively. A 3-μL injection volume was used. Calibration check samples of 2 mM Fmoc-His-Boc were run every 48 samples and blanks every 96 samples. Immediately prior to postpurification analysis, CLND calibration curves were generated using the Fmoc-His-Boc as a calibration standard. Data from the CLND and ELSD were collected through an analog-to-digital converter and processed using Nanosyn's proprietary software. The software utilizes available information about a compound at every step of the process (prepurification, purification, postpurification) and extracts purity and quantity data for that same compound. All DMSO quantitation determinations took into account the molecular weight of each individual compound, and the results yielded were expressed as molar concentrations.

The MS/ELSD/CLND system was dedicated solely to quantitation of purified compounds to avoid system contamination by compounds from reaction mixtures. Within a period of 18 months ~200 000 compounds were analyzed on the CLND/ELSD/MS system. We found that this high load on the instrument could be

maintained with routine maintenance procedures, which included washing of the CLND pyrotube with deionized water every 4−6 weeks and washing the membrane drier with MilliQ water every week. We also found that, after the system was shut down, it was important to run 10−20 test samples to allow the system to stabilize.

Aqueous Quantitation. An automated HP1100 HPLC system (Agilent, Palo Alto, CA) was used for aqueous quantitation. It comprised a vacuum degasser, a binary pump, a thermostated multiwell plate autosampler, a thermostated column compartment, and a diode array detector with a standard flow cell. The PEEK tubing used throughout the system had a 0.12-mm internal diameter. Reversed-phase HPLC was carried out on a Zorbax-SBAq (Agilent) column (2.1 × 50 mm, 5 μm). The binary gradient mobile phases (phase A, water, 0.05% TFA; phase B, MeOH, 0.05% TFA) were delivered to the injector by the Agilent HPLC pumps, running the following gradient conditions: 0% B to 100% B over 2.45 min, a hold at 100% for 2.05 min, steep gradient to 0% B over 0.1 min, and reequilibration at 0% B for 2 min. Following UV detection, the eluent flow was connected directly to a CLND 8060R (Antek Instruments, Houston, TX). The solvent flow rate through the CLND detector was 0.3 mL/min. The settings for the CLND were as follows: PMT voltage −750 V, sensitivity high × 5, argon 50 cm^3/min, oxygen 250 cm^3/min, ozone 25 cm^3/min, and oven temperature 1050 °C. These HPLC/UV/CLND systems were dedicated exclusively to aqueous quantitation work for two reasons: (1) they could not be used with acetonitrile or any other nitrogen-containing solvent, and (2) they have to be kept in optimal condition because many of the compounds quantitated were very close to the limit of detection (LOD, 5:1 signal to noise, 5 μM nitrogen).

Several routine steps were employed to maintain the performance of the CLND. Antek Intruments has recently introduced a new nebulizer tip (sapphire) for the CLND. This nebulizer significantly improved the sensitivity of our detector (with the stainless steel tip a 5:1 signal to noise ratio was obtained with a 20 μM propranolol sample, whereas the same result was obtained with the 5 μM propranolol sample). Even with a new nebulizer, we observed loss of signal sensitivity in time, which was attributed to the membrane drier. To maintain a limit of detection of 5 μM propranolol, a new membrane drier was installed every 2−3 weeks of continuous operation. To reduce the cost of operations, a washing procedure was employed. Twice a week the membrane drier was washed with Milli-Q water and then dried using purified air (Mark Homan, personal communication.) This simple procedure extended the lifetime of the membrane driers (in terms of loss of sensitivity) by a factor of 3.

Standard curves and solvent blanks were run with every plate, interspersed between samples. Data from the CLND was collected through an analog-to-digital converter and analyzed together with the UV data using a combination of an Amphora proprietary Agilent ChemStation software macro, and an Amphora proprietary automation template. After automated blank subtraction, integration, and data transfer, each chromatogram was visually inspected as a check for integration. Due to the variation in solubilities of compounds on any one plate, the automatic integration parameters were set to generally fit most of the samples in the set, with the outliers being manually integrated by an operator. The peak areas and peak heights from automatic integration were imported to an Excel template for further data analysis. Last, all quantitation data were uploaded to an Oracle database. Chromatograms were output as .gif files during automatic integration and can be viewed through an Amphora proprietary application. Data transfer was automated to minimize errors. In addition, the QC process involved comparing the new run with any previous analyses of the same compound present in the database.

Long-Term DMSO Storage and Liquid Handling. Upon completion of the DMSO quantitation process, samples were evaporated to dryness using the Virtis Genesis freeze-drier and shipped from the purification laboratory to the long-term storage laboratory. Individual purified solid samples were then redissolved in the same volume of DMSO to an average concentration of 3 mM/plate based on the mole amounts determined previously (see DMSO quantitation), in polypropylene 96-well deep plates (VWR International, Suwanee, GA), using a Biomek FX liquid handler (Beckman Coulter, Fullerton, CA) equipped with both a 96- and 384-pipet tip heads. Solutions were sonicated for 5 min (Aquasonic 550D sonicator, VWR), shaken for 5 min (Titer Plate Shaker, Lab-Line Instruments, Inc., Melrose Park, IL), and mixed using the Biomek FX. Every four 96-well plates was reformatted into several identical v-bottom polypropylene 384-well plates (Greiner Bio-One, Longwood, FL), named mother plates. These mother plates were sealed using a PlateLoc heat-sealer (Velocity 11, Palo Alto, CA) with pierceable heat seals (Velocity 11) and stored at −20 °C.

Periodically, one of the 384-well mother plates was thawed (~30 min, at room temperature), compounds were diluted with DMSO to a 1 mM average concentration per plate and mixed by repeat pipetting, and several daughter plates were made. Daughter plates were heat-sealed and stored at −20 °C. All daughter plates were in storage for 18 months or less for all the results presented in this study.

Microfluidics Platform Basics. *Primary Screening.* The HTS assays were developed on a microfluidic platform described elsewhere.[9] On the day of the biological screening, one daughter plate described above (1 mM average nominal concentration) was allowed to thaw to room temperature and then diluted 20-fold with pH 7.4 assay buffer to a nominal 50 μM with a final 5% DMSO concentration. This represents 5× the inhibitor concentration before addition of substrate and enzyme (using a Multidrop 384, Thermo Labsystems, Franklin, MA). Each compound was tested at 10 μM final nominal concentration in 1% DMSO in duplicate against any one enzyme. This approach yielded highly reproducible data. Any one primary screening campaign across the whole library of compounds was performed in less than 2 weeks. Because of limited capacity, aqueous quantitation could not be performed for the entire compound library every 2 weeks. This process was performed only once for each compound described herein (except for the compounds in the follow-up process, see below), ensuring that the particular daughter plate used in the quantitation process was in long-term storage for at least 6 months. Since these data were used only qualitatively, and were not used to adjust any biological results (as described below for IC_{50} experiments), this was considered acceptable. Periodically, a daughter plate was retrieved from the long-term DMSO storage system and allowed to thaw to room temperature, followed by a 20-fold dilution with pH 7.4 buffer to a 50 μM final nominal concentration (5% DMSO).

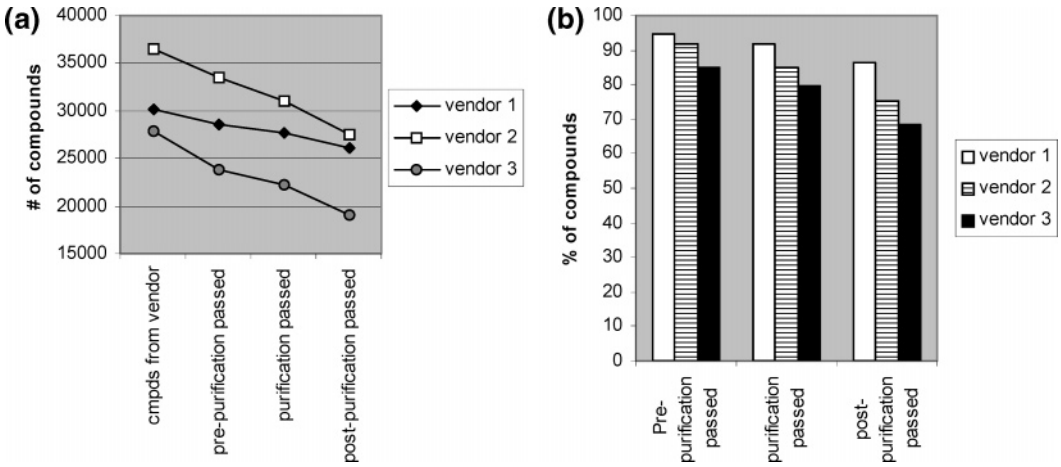

Figure 1. (a) Number of compounds passing each step of the purification process. (b) Percentage of compounds passing each step of purification process.

Aqueous quantitation values described herein were obtained from the 50 μM nominal daughter plate.

Enzymology/Follow-Up. Inhibitors identified in the primary screen were selected for further characterization. Compounds at a nominal 3 mM in 100% DMSO were cherry picked using a Biomek FX Span8 (Beckman Coulter) into two 384 v-bottom (Greiner) "intermediate" plates for each target in under 2 h. Both plates were then stored at −20 °C until ready for use, generally less than 1 week. The first intermediate plate, containing up to 40 compounds, was used for IC_{50} determinations while the second plate was used for quantitation. The IC_{50} intermediate plate was first thawed at room temperature for 30 min, and each compound was diluted 20-fold with assay buffer (containing no DMSO) from a nominal 3 mM in 100% DMSO to a nominal 150 μM in 5% DMSO. The plate was then incubated for 15 min at room temperature on a rotary shaker to allow for passive solubilization. Compounds were then serially diluted 2-fold in assay buffer containing 5% DMSO, generating eight inhibitor concentrations on the same plate. Immediately following dilution for all compounds, the IC_{50} intermediate plate was replicated into three identical assay plates (to allow K_i determinations). Following replication, enzyme and substrate were added, reducing the final compound concentration from 150 to 30 μM in 1% DMSO. The second intermediate plate was used for aqueous quantitiation and was typically processed within the same week as the enzymatic assay. This plate was thawed at room temperature for 30 min, followed by a 20-fold dilution with assay buffer and incubation for 15 min on a rotary shaker. This generated a 150 μM compound concentration in 5% DMSO that was used for analysis and was typically processed in less than 20 h. Aqueous quantitation values obtained from the 150 μM nominal intermediate plate were used to correct for the actual IC_{50}/K_i value for all nitrogen-containing compounds.

RESULTS AND DISCUSSION

A. Prepurification Analysis. Most high-throughput purification systems rely on purification without prior analytical evaluation of compounds or employ prepurification LC/MS analytical runs to sort out compounds that require purification.[17,18] Prepurification analysis has obvious advantages allowing removal of samples not worth purifying and identifying samples that require use of ELSD or UV response rather than MS to trigger fraction collection, thus facilitating downstream processing. However prepurification LC/MS analysis is expensive and reduces the overall throughput of the process, as well as requiring big initial investment into multichannel types of instrumentation.

Our approach was to employ a fast FIA/MS analysis prior to purifying any of the compounds purchased from vendors or synthesized in-house. Each compound was checked for identity and amount of targeted molecular ion. By performing this quick step, the more costly and lengthy purification analysis was avoided for the compounds that did not show the expected molecular ion. In addition, the amount of targeted molecular ion present in the chromatogram was utilized by Nanosyn's proprietary software resulting in assigning individual purification thresholds for each compound, thus reducing losses in the purification step. Results of the prepurification analysis of compounds acquired from different vendors are shown in Figure 1. We found that employing the FIA/MS analysis step is a cost-efficient solution, which greatly enhances the productivity of the whole process.

B. Purification. Although mass-directed HPLC purification is widely used in the industry, to our knowledge, very few laboratories employ this process to enhance the quality of large highly diverse compound collections (over 100 000) purchased from outside vendors. The main reason for this is the cost associated with purification and the need for ultrahigh throughput of purification (over 10 000 compounds per month) to keep up with HTS. We found that decreasing the scale of purification along with implementing a highly efficient, completely automated process allowed us to keep the cost low and achieve throughput of up to 15 000 compounds per month. Within a period of ~9 months 94 359 compounds from three vendors (~30 000 compounds each) were purified and characterized through our process.

Throughput of the purification process greatly depends on the quality of compounds in the library, and we asked whether there was a significant variability in libraries' qualities for different vendors. As shown in Figure 1a, vendor collections differ quite substantially. For example, although we acquired 6361 more compounds from vendor 2 than from vendor 1, only 1419 more

compounds from vendor 2 have been included in the screening library.

Overall between 68 (vendor 3) and 87% (vendor 1) of the initial number of compounds in a vendor collection passed through all steps of the purification process and have been included in the screening library (Figure 1b). We also found that libraries of compounds "historically" collected (vendor 3) have overall lower quality of initial compounds compared to more recently synthesized compound libraries (vendor 1) or a mixture of historical collection and recently synthesized compounds (vendor 2). In contrast, we found that collections containing "historically" collected compounds (vendors 2 and 3) are more appealing from the point of view of diversity, log *P* distribution, and other physical−chemical properties (data not shown).

C. Quantitation. Since it is virtually impossible to obtain and run calibration curves for every compound needing quantitation in an HTS setting, there is an obvious need to find detectors that have equimolar responses across a wide diversity of structures.[20] Two liquid chromatography HTS-amenable detectors are reported by their respective manufacturers to be linear across a large diversity of compounds: the ELSD and the CLND (Applicable to nitrogen-containing compounds only.). Several reports have directly compared the two detectors (most recently, ref 21) as well as comparing them with other analytical detectors[22] in terms of precision, linearity, sensitivity, and reliability. The CLND is more sensitive and more linear than the ELSD (∼4 times) but is less precise and less robust. Additionally, the CLND can only detect nitrogen-containing compounds, and important to note is the fact that when N−N bonds are present in the molecule, the absolute amount detected is somewhat underestimated because of the formation of N_2. This has clear implications in the accuracy of the measurement for these compounds. However, across a chemical series, where the N−N bonds are present in the same core, the amount of N_2 formed seems to be constant. This permits the relative quantitation within that series (unpublished results).

Both the CLND and ELSD detectors were evaluated in our laboratories. We found that the response linearity and to a lesser extent the lower LOD were critical factors in choosing a detector for this study. We present herein the results of quantitation of compounds in aqueous solutions using the CLND. Methods to increase the reliability and maintain a very low LOD are also described in the Experimental Section.

Since ∼97% of the compounds in our library contain nitrogen, the N-limitation of the CLND was acceptable. Of the N-containing compounds in our database, 21% have N−N bonds, and quantitation is underestimated to varying extents. Despite all this, we still found the CLND to be the best method for quantifying compounds, because of universality, sensitivity, and integration with LC separation techniques.

(a) DMSO Quantitation. Many pharmaceutical organizations store compounds as frozen DMSO (100%) solutions (4 or −20 °C) under controlled humidity or inert atmosphere. Several reports have recently been published on the stability and solubility of compound stored in DMSO solutions.[23−25] Initial DMSO solubility can be problematic, with an average of 10−15% and up to 20% of compounds from commercially available databases being poorly soluble.[24] More problems can arise from storage of compounds in DMSO. Repeated freeze−thaw cycles can decrease the linear percentage of compound left in solution,[23] similarly caused by water absorption during the time in storage.[25] Another issue is the impact that the concentration of compounds in DMSO can have on biological measurements.[25] A more complete study on this topic has been presented at a recent meeting,[26] where it has been shown that, for a diverse collection of ∼7000 compounds, the relative DMSO solubility was decreased for a 10 mM stock compared to a 2 or 3 mM stock. The conclusion was that concentrations between 2 and 5 mM are optimal concentrations to maximize DMSO solubility of a diverse HTS library.

In our laboratories, we chose a concentration of 3 mM in DMSO for the stock mother plates. These mother plates were further diluted to 1 mM with DMSO into daughter plates, which were used in the biological assays. The 3 and 1 mM concentrations, respectively, are average concentrations per each plate of compounds (see below).

Each sample purified in the system described above was automatically added to the DMSO quantitation queue. Since fractions collected during the purification process are in an acetonitrile/water mixture, and due to the variability of collected volumes and the uncontrolled evaporation occurring during the purification queue, accurate quantitation is not possible directly from the collection plates. Thus, upon completion of the purification sample queue, fractions were evaporated. Dried compounds were then redissolved in DMSO, shaken, and centrifuged. Following redissolution, compounds were characterized by FIA MS/ELSD/CLND for identity and quantity. Each of the compounds selected to become part of the compound library had a purity requirement of at least 90% (determined using the lowest number from two detectors, ELSD and UV 254) and quantity of at least 100 nmol.

Figure 2a shows the distribution of actual DMSO concentrations in daughter plates across a diverse set of ∼2000 compounds after purification. Compounds were brought up to a constant volume of DMSO, but the actual amount obtained through the purification process per compound well varied as shown. For each plate of compounds, the actual compound concentrations in their respective stock DMSO solutions varied between 0.25 (the minimum threshold for accepting a compound for testing) and 30 mM, but the average concentration in 100% DMSO across each daughter plate was 1 mM, prior to dilution in buffer. Since these compounds were quantified in DMSO, the variation is most probably due to the different amounts collected during the purification process, as opposed to differences in the dissolution process or compound precipitation from DMSO. Figure 2b shows a distribution of the ratios of the resulting DMSO concentrations of purified compounds to the calculated DMSO concentrations based on the vendor-provided weight information, assuming

(20) Yan, B. *Mod. Drug Discovery* **2004**, *(*Feb), 30−34.
(21) Allgeier, M. C.; Nussbaum, M. A.; Risley, D. S. *LC−GC North Am.* **2003**, *21* (4), 376−381.
(22) Petritis, K.; Elfakir, C.; Dreux, M. *J. Chromatogr., A* **2002**, *961* (1), 9−21.
(23) Kozikowski, B. A.; Burt, T. M.; Tirey, D. A.; Williams, L. E.; Kuzmak, B. R.; Stanton, D. T.; Morand, K. L.; Nelson, S. L. *J. Biomol. Screening* **2003**, *8* (2), 205−215.
(24) Balakin, K. V. *Curr. Drug Discovery* **2003**, *8*, 27−30.
(25) Cheng, X.; Hochlowski, J.; Tang, H.; Hepp, D.; Beckner, C.; Kantor, S.; Schmitt, R. *J. Biomol. Screening* **2003**, *8* (3), 292−304.
(26) Iyer, J. LRIG meeting, Jan 3003.

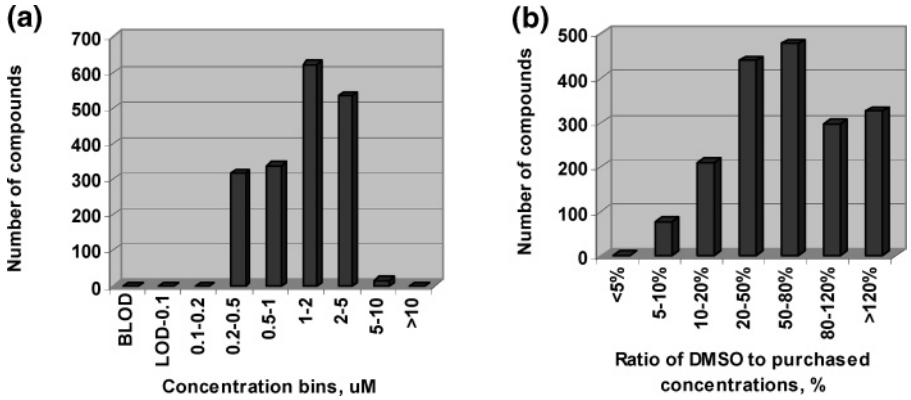

Figure 2. (a) Distribution of DMSO stock concentration values (mM) across ~2000 compounds. Average DMSO concentration across every plate of compounds is 1 mM. Binning of concentration values is done according to variation analysis performed (Table 1). (b) Ratio (%) of actual DMSO concentrations to concentrations calculated based on amount of compound purchased from each vendor (assuming compounds were 100% pure) for the same samples.

samples were 100% pure, for the same compounds in Figure 2a. The analysis performed has also shown DMSO concentration as high as 2.5-fold over those calculated based on vendor information, which can be explained by a low precision in weighing small (1 mg) samples. Concentrations lower than the expected 1 mM can be explained by both purity levels less than 100% and low weighing precision.

(b) Aqueous Quantitation. For perfectly water-soluble compounds, DMSO quantitation represents an upper limit on the concentration in buffer. If other biological assay components increase compound solubility, then the reported aqueous quantitation may reflect a lower concentration measurement. Since the CLND cannot be used to measure compounds in an assay mixture (due to contribution of other nitrogen-containing factors), our method of measuring after the initial solubilization step (with the addition of buffer to DMSO) represents the closest approximation of the actual assay concentration, as presented herein.

Variability of Sample Concentration Determination. A detailed 6-month study was performed on the various standards run to assess both the accuracy (how close the measurement is to the correct value) and precision (how exact and reproducible a measurement is) of the whole process. This study was performed with standards prepared from dry powders at various times to determine the variability of the weighing and pipetting processes.

In terms of instrument-to-instrument variation, we found the three CLND detectors we evaluated to be very different. The response factor (ratio of area under the curve to actual sample concentration) was 5 times higher on two of the detectors compared to the third, where the UV response was within a 5% window for the same samples. Only the two detectors that gave a low and similar response factor were accepted and used in this study. Since one of these detectors has been running in a high0throughput mode for only 2 months, all results presented in this study are from only the one detector, which has been running continuously for over 18 months; however, preliminary analysis shows very similar variations across the newer detector. A detailed analysis of the instrument-to-instrument variation is currently underway. Preliminary results show a very low contribution of this variation to the overall variation of the assay (data not shown).

Statistical analysis of the results showed that precision varied with the concentration of standards and rate of replication. The variability of CLND differed across the range of concentrations tested (Table 1). For a single-point estimate, the precision of measurement was ~25% at nitrogen concentrations of 50 μM and higher. Precision was 12−19% for two to three replicates. For concentrations lower than 50 μM, precision was worse. This lack of precision may be attributed to the automatic integration process employed. Upon manual integration of peaks, precision became better (data not shown). However, due to the large volume of analyses performed, manual integration of samples was not a practical solution.

Table 1. Precision (%) of the CLND Detector over Six Months of Continuous Operation (24 h/day, 7 days/week)[a]

	no. of replicate injections	nitrogen concentration, μM						
		10	20	50	100	200	500	1000
±PI width	1	191	86	27	23	24	21	20
±PI width	2	136	61	19	16	17	15	14
±PI width	3	111	50	16	13	14	12	12

[a] Precision results are presented as prediction interval (PI) widths.

In terms of nitrogen concentration, the LOD for our system was maintained at ~5 μM through the maintenance described. This gave a 10-fold window below the nominal concentration of the source plate for the primary screening (50 μM) and a 30-fold window below the nominal concentration for the follow-up screening (150 μM). All compounds below the LOD are shown in all graphs as values at that limit, which is clearly not accurate when trying to understand the extent of the solubility issue. This was done for the purpose of graphically representing all the data.

Precision results are shown below in Table 1 and are presented as prediction interval (PI) widths. The prediction interval is a range within which a future measurement is expected to fall. For example, from $n = 2$ measurements at 50 μM, the PI predicts with 95% confidence that the next measurement will be within 19% of the mean of the first two measurements.

The accuracy of the measurement was consistent using a weighed sample of caffeine at a concentration of 250 μM, quantitated using the propranolol standard curve with a variation of less than 15% (229−269 μM measured).

We assessed the variation of quantitation across a chemically diverse collection, which included several compounds containing

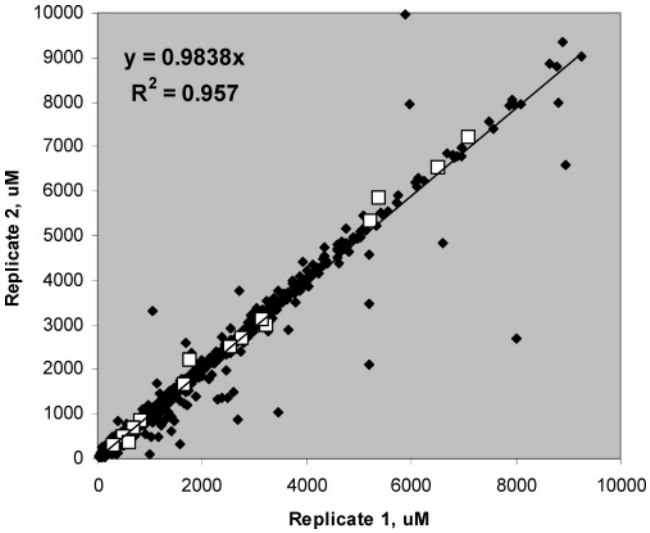

Figure 3. Within-plate replicate values across 900 different samples from 39 plates, run over a 6-month interval. White squares indicate the N−N-containing compounds.

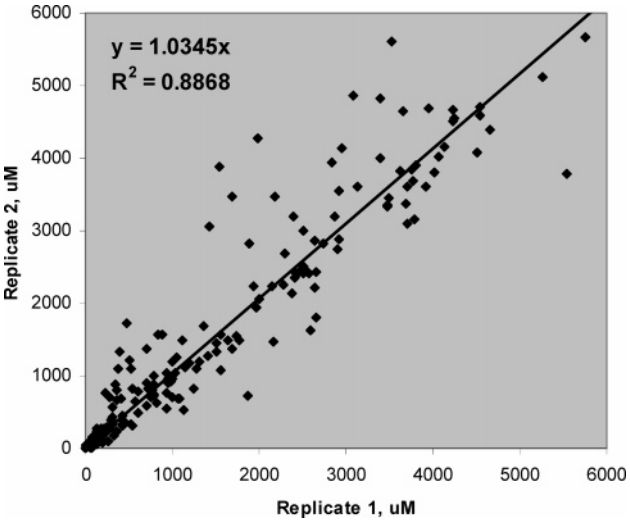

Figure 4. Between-plate replicate values across 400 different samples, from 5 plates, run over a 4-month interval, with independent controls.

N−N bonds. Nine-hundred random samples across 39 plates were sampled in duplicate: results are shown in Figure 3. Only 54 of the 900 samples were more than 30% different from each other, and each of the chromatograms for the 54 compounds in question had more than one peak. This result also supports the finding that, for compounds containing N−N bonds, replicate measurements fall within the same range of variability as non-N−N-containing ones.

We also assessed the variation of quantitation across a chemically diverse collection of 400 samples, taking into account the whole process variation from compound storage to liquid handling and detector variability. Each sample was analyzed as a single injection per run with the duplicate analysis being made from a different source plate at least 1 month after the first analysis. The results are shown in Figure 4. The increased variability found between-plate versus within-plate measurements is directly attributable to differences in compounds between batches.

Based on this analysis, all subsequent quantitation analyses for IC_{50} and K_i determinations were performed in triplicate and an average value was used. As part of the QC process employed, our automated data analysis template alerted the operator when there was a large difference among the replicate values. These samples were then manually integrated and analyzed.

Range of Concentrations. Figure 5a shows the distribution of buffer concentrations for a random selection of ~2000 compounds from Amphora's diversity library. For the purpose of representing both the DMSO and the buffer concentration distributions on the same graph, the DMSO concentrations (shown in Figure 2a) were extrapolated to the concentrations these compounds would have at the buffer level (using dilution factors) if they were 100% soluble in buffer. Figure 5b shows the distribution of solubility ratios for the same compounds in Figure 2. Both graphs illustrate the poorer solubility of compounds in buffer solutions (even those containing 5% DMSO) compared to 100% DMSO solutions.

This finding can have implications in the biological assessment involving any chemical, be it an HTS process or a single compound experiment. If the potency of any one compound is calculated using the DMSO stock solution concentration and extrapolated to the value at the screening level, there is a good chance that the final results are underestimated. Since compounds are present in their testing solutions at much lower levels than assumed (based on their stock solutions), this has an impact not only on

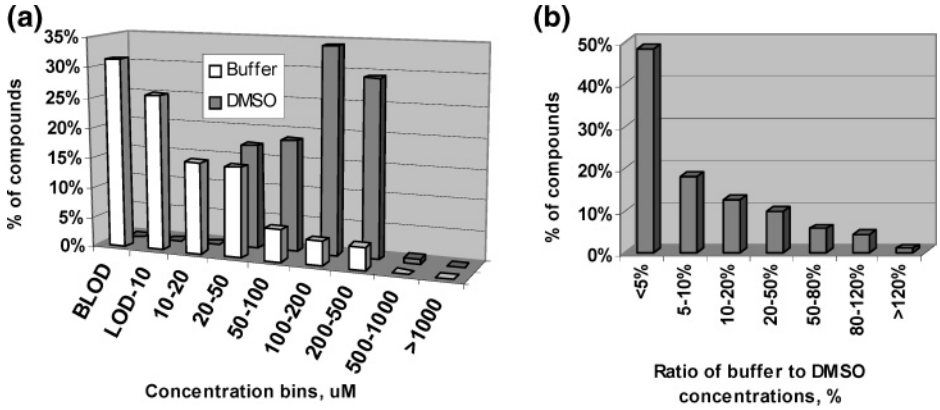

Figure 5. (a) Distribution of concentration values across ~2000 compounds. Values shown on the *x* axis are compound concentrations (μM). All the DMSO concentrations were converted to the buffer level (based on dilution factors). Binning of concentration values is done according to variation analysis performed (Table 1). (b) Ratio (%) of aqueous-to-DMSO concentrations for the same samples.

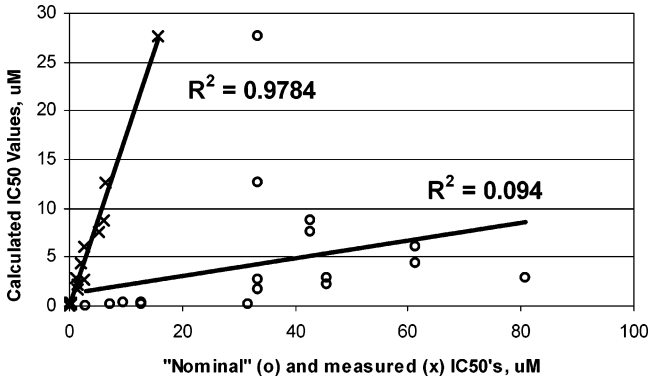

Figure 6. Nominal IC_{50} values (○) based on assumed compound concentration in buffer. Actual measured values (×) are based on quantitation in aqueous buffer on the *x* axis. The *y* axis represents the IC_{50} values calculated directly from the primary HTS inhibition (using a formula that includes the concentration measurements).

hit rates but also on the credibility of the negative data for any screen; varying solubilities of compounds in aqueous solutions can change the SAR conclusions within any chemical series. Some examples of the effect of the poor aqueous solubility of compounds on biological results are illustrated below.

D. HTS Significance. A convenient, nontargeted selection of 2797 samples was chosen from Amphora's diversity library for this analysis. These compounds were quantitated in buffer as described. Of the 2797 samples, 748 were below the limit of quantitation for the CLND. Out of the 748 samples that were below the LOD, only 30 were active against at least one of the 52 enzymes tested (with only 14 compounds active against more than one target), a "hit rate" of 4.0%. Out of the remaining 2049 compounds quantitated in the CLND, 654 were active against at least one enzyme, a "hit rate" of 31.9%.

The most plausible explanation for the difference in the two overall hit rates is the fact that the compounds below the LOD were present in too low an amount in the buffer solutions to have an effect in the enzymatic assay. This has implications in using the negative data for any HTS screen and having that information when performing analysis of an HTS screen offers a clear advantage.

Out of the above 2797 samples analyzed, 748 were below the LOD, which is 26.7% of the total number of samples, whereas only 155 chromatograms had more than one peak (5.5% of the total). Assuming that samples were below LOD because of the low solubility of the compounds in the aqueous solution, another conclusion drawn from this study was the fact that stability of compounds in buffer for short periods of time (~24 h, the average time the compound sits in the assay plates) seems to be less of a problem than solubility. This conclusion applies to a diversity library. We have seen examples of series of chemically similar compounds where stability was more of an issue (data not shown).

E. Structure−Activity Relationship implications. Rank ordering of IC_{50} and K_i values may change within a series as a result of aqueous quantitation. Typical screening procedures test the activities of compounds at one concentration to determine eligible candidates for more rigorous characterization. Follow-up assays for these activities include determining the IC_{50} value from dose−response curves at one substrate concentration ([S]) or by using multiple substrate concentrations, determining the mode of action and the inhibition constant (K_i) by nonlinear regression. Given that the mechanism, [S], and K_m values are known, the K_i can be calculated from the IC_{50} using the correct Cheng and Prusoff[27] relationship. This has been an accepted time-saving tool when evaluating potency for structurally related compounds and SAR evaluation. The benefit of utilizing quantitation in these studies is obvious as it corrects IC_{50} and K_i values to reflect true concentrations of potency and makes use of negative data if the compounds are generally insoluble.

Table 2. Rank Ordering Change of Potency Values within a Series of Chemically Similar Compounds When the Values Are Adjusted for the True Buffer Concentration, the DMSO Concentration, or the Assumed "Nominal" Concentration of Compounds in Solution

rank measd IC_{50}	rank nominal IC_{50}	rank DMSO IC_{50}	measd IC_{50} (μM)	nominal IC_{50} (μM)	DMSO IC_{50} (μM)
1	1	2	0.044	7.2	8.8
2	8	1	0.11	31.7	7.8
3	5	6	0.12	12.7	25.3
4	3	4	0.15	9.6	15.8
5	2	3	0.18	7.2	8.8
6	6	7	0.24	12.7	25.3
7	4	5	0.38	9.6	15.8
8	19	13	1.1	80.9	59.7
9	15	16	1.2	45.6	68.2
10	9	11	1.5	33.5	52.2
11	16	17	1.5	45.6	68.2
12	17	14	2.1	61.4	68.0
13	10	9	2.5	33.5	50.1
14	18	15	2.6	61.4	68.0
15	13	18	5.3	42.6	84.8
16	14	19	6.2	42.6	84.8
17	11	10	6.3	33.5	50.1
18	7	8	11.7	19.4	37.8
19	12	12	15.7	33.5	52.2

Table 2 illustrates the rank-ordering change within the same common pharmacophore series shown later in Figure 6, when the IC_{50} values were calculated using either the "nominal", DMSO, or buffer concentration. For this particular series, not only was the rank ordering changed when using the real concentrations versus extrapolated stock solution concentrations, but it allowed detection of relatively insoluble, but potent, active compounds. Without measuring the intrinsic potencies of these compounds, this would not have been a scaffold prioritized for chemistry follow-up work. After solving these solubility issues, this series is currently at the hit-to-lead stage in one of the drug discovery programs in our laboratory.

Calculated versus Measured IC_{50}'s. To increase data collection efficiency, compound quantitation and the Cheng and Prusoff[27] relationships were used to calculate IC_{50} values directly from primary inhibition values for a kinase assay. Compounds in a related series (Figure 6 and Table 2) had been previously found to be competitive with ATP, and the appropriate competitive equation was used. It was assumed that the reaction followed a random binding of substrate model. Because the ratio between the peptide concentration used in the assay and its K_m was 0.1, the peptide contribution to the IC_{50} was ignored. Experimentally determined IC_{50} values from dose−response curves were com-

(27) Cheng, Y.; Prusoff, W. H. *Biochem. Pharmacol.* **1973**, 22, 3099−3108.

pared to primary screen-calculated IC_{50} values adjusted for DMSO-soluble concentration (nominal) or adjusted with aqueous buffer quantitation (measured) (Figure 6). A comparison of the regression line fits (R^2) for these data shows the pronounced effect quantitation has on improving the fit ($R^2 = 0.98$). This experiment demonstrates that it is possible to utilize primary screening data to elucidate SAR trends when the true amount of compound is known and that low inhibitions due to low solubility do not negatively influence SAR interpretation.

CONCLUSIONS

Two major factors associated with compounds must be accounted for in high-throughput screening processes: sample quality and sample quantity. Though many laboratories are now addressing sample quality by employing MS-based purification methods, to our knowledge, few are looking at the quantity of compounds in solutions. More importantly, the concentration in aqueous solutions for biological assays is not being considered. Rather than ignore these issues, we have attempted to capture more analytical information regarding the solubility and stability of compounds used in the screening process.

We have described a streamlined system for purifying and quantitating compounds at various stages of the process, coupled with automated "traditional" storage methods (DMSO, −20 °C). All of our compounds are purified by HPLC using mass-directed fraction collection, then quality controlled and quantitated in their respective DMSO stock solutions by CLND/ELSD/MSD. We have also demonstrated a method for determining the concentration of small organic molecules in aqueous solutions. This method is routinely used in our laboratories for accurate IC_{50} and K_i determinations, as well as quantitation of primary screening compounds, which has the advantage of being simple, sensitive, and cost effective. The method works best if the starting stock solution was relatively pure (>90%) and above a certain concentration threshold.

Data presented herein exemplify only some of the processes potentially affected by the lack of available buffer quantitation data, such as HTS analysis and accurate potency determinations. However, these data can also be utilized in other ways, such as a training set for building more accurate aqueous solubility prediction models, determining SAR directly from a one-point activity determination, or helping in library selection and design.

ACKNOWLEDGMENT

The authors gratefully acknowledge the HTS Operations group at Amphora for support of this work, especially Steve Birkos for his commitment to running fast and precise enzymatic assays, Erica Schnars for processing the CLND data in record time, Charles Perez for his promptness in fixing any laboratory-related issues, John Dickson and Jose Mendoza for their skilled editing, and PAC (formerly Antek Instruments), in particular Mark E. Homan for all his help in turning the CLND into a 24/7 running instrument. We are grateful to the entire Analytical Chemistry group at Nanosyn for their hard work and efforts, which allowed achieving the record throughput of 10 000 purified compounds per month being shipped on a routine basis. We thank Robert Lagasca for keeping all the LC/MS instrumentation in good working condition and acknowledge the assistance provided by Helen Hua in running the instrumentation.

SUPPORTING INFORMATION AVAILABLE

Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review June 3, 2004. Accepted August 24, 2004.

AC0491859