Strategic Pooling of Compounds for High-Throughput Screening

Mike Hann,[†] Brian Hudson,[‡] Xiao Lewell,[†] Rob Lifely,*[,‡] Luke Miller,[‡] and Nigel Ramsden[§]

Department of Biomolecular Structure, Compound Diversity, and Exploratory Chemistry,
Glaxo Wellcome Research and Development, Medicines Research Centre, Gunnels Wood Road,
Stevenage, Hertfordshire SG1 2NY, U.K.

Bringing new medicines to the market depends on the rapid discovery of new and effective drugs, often initiated through the biological testing of many thousands of compounds in high-throughput screening (HTS). Mixing compounds together into pools for screening is one way to accelerate this process and reduce costs. This paper contains both theoretical and experimental data which suggest that careful selection of compounds to be pooled together is necessary in order to reduce the risk of reactivity between compounds within the pools.

INTRODUCTION

The pharmaceutical industry continues to have to invest huge resources into the discovery of a single new product. Current estimates of the costs of researching and developing a new medicine are of the order of £300 million. With increasing pressure on the industry to reduce this cost, the past 10 years has seen a move toward automation to increase the throughput in all stages of the research and development pipeline. Thus, high-throughput screening (HTS) of compounds from library synthesis and compound collections has become the norm in the industry.[1] However, like all new technologies, the introduction of these methods has brought a new set of problems. This paper reports our efforts to address one of these.

Estimates of the numbers of molecules that could theoretically be made and which are likely to have structures and a MW which are compatible with being suitable for consideration as drug molecules are very large (probably greater than 10^{18}). These are very much bigger numbers than those which are already known in the literature (the CAS database has just reached 1.8×10^{7}), and the number available to acquire from commercial sources or in-house collections is smaller again (typically of the order of 1×10^{6}). However, with the modern synthesis methods developing rapidly, it is likely that this number will grow. The cost of acquiring such large numbers of compounds is however only one part of the cost of running HTS campaigns. The cost of the biological assays can vary from a few cents to dollars per well so the cost of screening 1 million samples for just one project could be as much as $1 million. Thus, even if the logistics and cost of synthesis or acquisition could be achieved, it is unlikely that the industry will be able to afford to screen all the compounds available at one time. Aside from monetary concerns, it is highly inefficient to screen everything available if the same results can be achieved for less effort.

There are several ways to reduce the number of assays to be conducted. One method is prior selection or design of compounds based on preexisting knowledge of the molecular target for the required bioactivity [2−4] or of already known active molecules,[5] with the possible inclusion of more generic drug like properties.[6] While this current paper addresses some issues to do with the latter type of knowledge, the main aspect of the work described here is a complementary approach to reduce the number of assays that need to be carried out. This is to pool the samples prior to screening so that the number of assays required is reduced by a factor which is the number of samples combined into a pool. For example, if 10 compounds are pooled and assayed together, then the number of assays done is one-tenth of the number that would be done if the compounds were assayed individually.

While we had previously implemented a pooling strategy in which compounds were randomly assigned to these pools of 10 compounds, similar in concept to Devlin et al.,[7] we became concerned that intermolecule chemical reactivity might occur within these pools. This was because the pooled samples have a high organic concentration (10 mM) and, although stored long term at −20 °C in dimethyl sulfoxide (DMSO), they would also be held at +4 °C and finally at the assay temperature for significant periods of time. We were also concerned about the effects on the ionic status of the compounds when combining them with other compounds, which might alter dramatically the solubility of the compounds as originally prepared. Additionally, although not addressed in this paper, other possible undesirable effects of pooling such as competitive inhibition or allosteric effects should be considered.

With these concerns in mind, we developed an alternative pooling strategy in which potentially reactive molecules are kept apart. We have also tried to avoid changing the ionic character of compounds in the pools compared to their status as individual compounds. Additionally, we addressed the issue of the possible additive biological activities of compounds in pools by ensuring that very similar compounds were not mixed together, since when a pool is screened in a

---

[†] Department of Biomolecular Structure.
[‡] Compound Diversity.
[§] Exploratory Chemistry.

bioassay, it is the total bioactivity of the mixture which is observed in the experiment. If there are several weakly active compounds in a pool, then the sum of these activities is difficult to differentiate from the case where just one compound is responsible for all of the activity. Since *very* closely related compounds (as defined for instance by their 2D structure) are likely to have similar biological activity, we need to find a method for ensuring that such similar compounds are in different pools. In this way we should be able to reduce the number of false positives from screening the pools.

The most obvious and prevalent form of unwanted chemical reactivity was considered to be the reaction of electrophiles (E) and nucleophiles (N). Other possible broad classes of chemical reactions (e.g., rearrangement, oxidation/reduction, addition/elimination, radical) were felt to be less likely. This is because typical nucleophiles (e.g., amines) and electrophiles (e.g., esters) are quite prevalent in the types of molecules that we have in our collections. With this in mind we separated these classes of molecules so that when pooling was carried out, we did not mix them. In practice we created four chemical classes: electrophiles (E), nucleophiles (N), electrophile/nucleophiles (E/N), and inerts (I). The E/N classification was needed for molecules which contain both an electrophile and a nucleophile; such molecules contain the E or N in such a conformation, configuration, or state of reactivity that inter- or intramolecular reactivity is not possible. Because we were not able to easily predict which was the case, we decided that these molecules were best kept to a group on their own. The inert classification contained all molecules that were not in either E, N, or E/N. We used the Daylight SMARTS language[8] to define the E and N types based on the perception of chemists' views of functional group reactivity. The definitions that we used are in no way exhaustive of all known E's and N's, but we believe we have captured the most likely ones to be present in the molecules we typically screen.

Solubility was the physical property that we were most concerned would interfere with screening. In particular, we were concerned that compounds would be out of solution even before they reached the buffered biological assay medium. Unfortunately there are no good a priori predictors of molecular solubility, and, in particular, the effect of mixing compounds in pools has not, to our knowledge, been studied in this context. However, we reasoned that one of the most likely parameters to change the solubility of a compound in a pool was a dramatic change of the ionic state of the molecules. Thus, for instance, mixing a strong acid with a strong base would create a salt whose solubility properties might be very different to its constituent parts. We decided to classify all the compounds which were available for pooling as either acidic (A), basic (B), zwitterionic (A/B), or neutral (N). Even more so than the assignment of E and N by SMARTS, the definition of acidic and basic labels is very difficult in a generic sense because of the environmental effects of the rest of the molecule on the pK_a of a particular functional group. We also considered the fact that some acids and bases in the collection already exist as salts. We therefore subdivided A and B into an A_free, A_salt, B_free, and B_salt classification based on the recorded information in our compound registry file.

Having classified all molecules on both reactivity and ionic status, we could assign all molecules to their correct cells in a matrix (Figure 1), and by pooling only within an element of the matrix, reactivity and ionic interaction events should be minimized. If there were insufficient samples within a given matrix element to satisfy pooling requirements, then neutral molecules of the same chemical class were used to make up the numbers.

This paper describes the means by which compounds have been segregated according to their chemical class before pooling, taking into account likely solubility issues and the need to maximize structural diversity within a pool. We also provide experimental evidence to substantiate our ideas that intermolecule chemical reactivity can thereby be reduced.

MATERIALS AND METHODS

Compound Filters. Prior to application of the strategic pooling methodologies, inappropriate compounds were removed using a set of structural filters. These fall into three types.

(a) Basic Filters. A set of basic filters was used to remove the following, nondrug like features: inappropriate salts (e.g., cyanides, heavy metals); compounds outside the molecular weight range if 150−750; compounds containing any atom other than C, H, O, N, P, S, F, Cl, Br, and I; compounds with less than 10 atoms; compounds lacking at least one of the chemically tractable bonds C−N, C−O, and C−S.

(b) Hard Filters. A set of substructure search filters was used to remove compounds containing inappropriate functional groups. These comprise filters for reactive functional groups, unsuitable leads (i.e., compounds which would not be initially followed up), and unsuitable natural products (i.e., derivatives of natural product compounds known to interfere with common assay procedures).

The hard filters, and their SMARTS definitions, are defined in the Supporting Information.

(c) Soft Filters. In addition to the basic and hard filters the following two soft filters were applied:

Based in part on the work of Gillet et al.,[6] a genetic algorithm has been structured and trained to identify unsuitable compounds which the other filters would fail to find. This algorithm scores compounds for drug likeness, relative to a training set classified by medicinal chemists. It excludes compounds for which a rigid "hard" filter cannot be defined. Experience with the algorithm has identified a suitable value for this score, below which a compound can be considered unsuitable. These compounds are removed.

In addition, a complexity filter was applied, implemented as the percentage saturation of the standard daylight fingerprint. Compounds with low values have little functionality even though they might pass all the basic filters and have quite high molecular weights. At the other end of the scale, compounds with a high degree of functionality are also filtered. These are considered to be too complicated for consideration as lead molecules of biological interest. This provides a simple, but powerful, method for excluding the very simple and very complicated compounds without having to define what this means in any detail. Compounds with less than 10% saturated and greater than 60% fingerprint saturated were excluded.

STRATEGIC POOLING OF COMPOUNDS FOR HTS

J. Chem. Inf. Comput. Sci., Vol. 39, No. 5, 1999 **899**
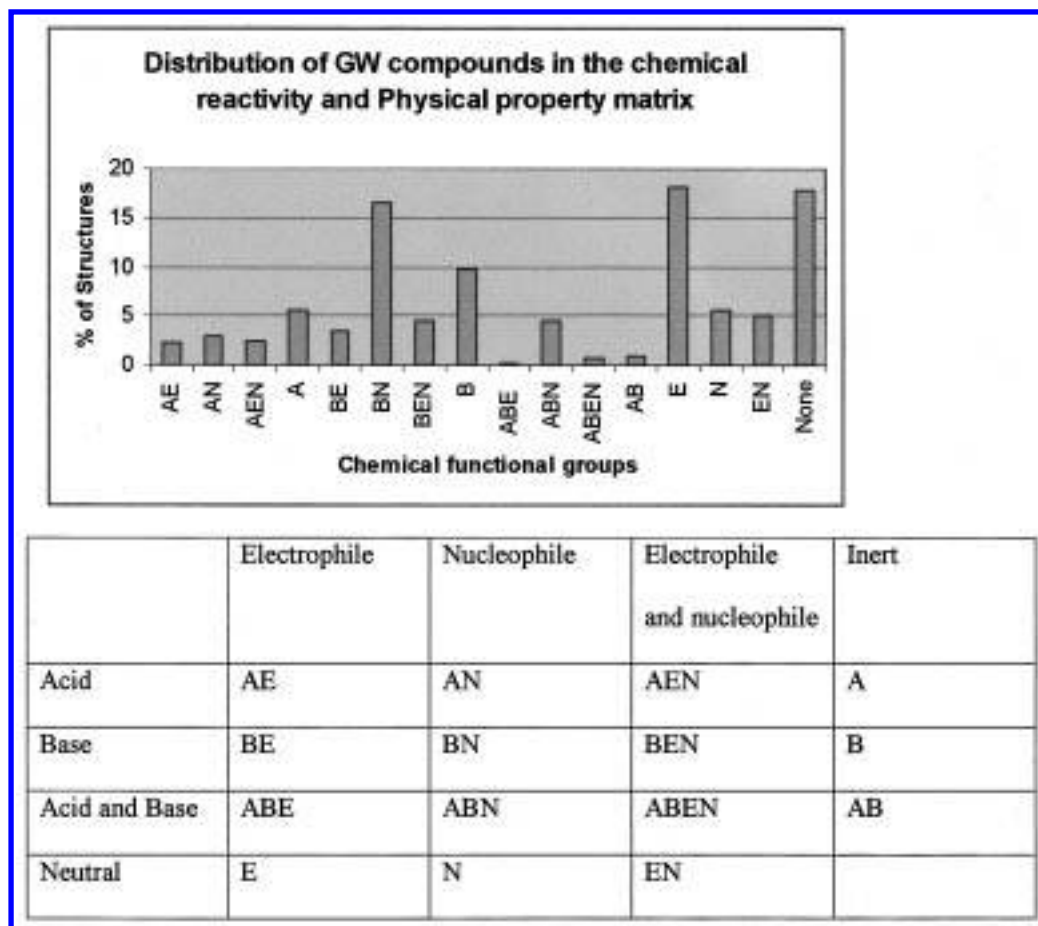
Figure 1. Separation of GW compounds into 16 chemically distinct categories according to whether a molecule contains any of the functional groups acids (A), bases (B), electrophiles (E), and nucleophiles (N).

Naturally, application of these filters is order dependent as, for example, the drug like compound filter also removes many with molecular weights outside the range specified in the basic filters. The filters were applied sequentially using the least CPU intensive filters prior to the more computationally demanding ones. The combined effect of all the filters was to remove about 25% of the potentially available compounds prior to the strategic pooling.

Separation of Compounds into Distinct Classes of Chemical Reactivity. Once compound filters had been applied to remove unsuitable samples, computational procedures were developed to identify reactive functional types within a molecule. Molecules were then separated into 16 different classes depending on whether a molecule contains one or several such functional types. Functional types (Figure 1) were defined as acids (A), bases (B), electrophiles (E), and nucleophiles (N). The definitions of reactive functional groups were written in SMARTS[8] and are given in the Supporting Information. The program for processing the reactive functionalities was written in C utilizing Daylight toolkit routines.

Burden Number To Order Compounds According to Similarity. The Burden number[9] for each structure was calculated using code developed by A. Rusinko, GW RTP.[10] This yielded a 1D descriptor for each compound which was used to order the compounds *within* each chemical class of the reactivity/ionization state matrix[11] (see Figure 1). We have shown by comparison with pairwise similarities of Daylight fingerprints that by ordering compounds in this

manner we ensure that very similar molecules are placed next to each other in sequence (data not shown). By the nature of the orthogonal pooling strategy (discussed below) this then minimizes the chance that these very similar compounds are present in the same wells.

Orthogonal Pooling. Once the compounds in each of the chemical classes (or matrix elements) shown in Figure 1 had been ordered by Burden number, orthogonal pooling was conducted by selecting a set of 8000 compounds from within a particular class. These were laid out onto master plates using 80 wells of 96 well plates (see Figure 2). The plate map describes the sequence of compounds in each of the 80 wells, and is derived from rows A to H and columns 1−10, the first position being A1 and the 80th position being H10. When 8000 samples are plated out in this fashion, a total of 100 plates are created, each well containing a unique discrete compound. These plates are nominally arranged into a 10 × 10 matrix (maintaining the plate order across rows to mirror the increasing Burden numbers of the samples on the plates). The first pooled plate containing 10 samples per well is created by mixing A1 from plate 1, A1 from plate 2, ..., A1 from Plate 10 to give A1 in the pooled plate. This continues until the 80th position in the pooled well is made from H10 from plate 1, H10 from plate 2, ..., H10 from plate 10. In this way, 20 pooled plates (each discrete compound is represented twice) are made from 100 discrete plates. Thus, the orthogonal pooling strategy that we use ensures that compounds with every 80th Burden number order are pooled in one orthogonal direction and every 800th in the other
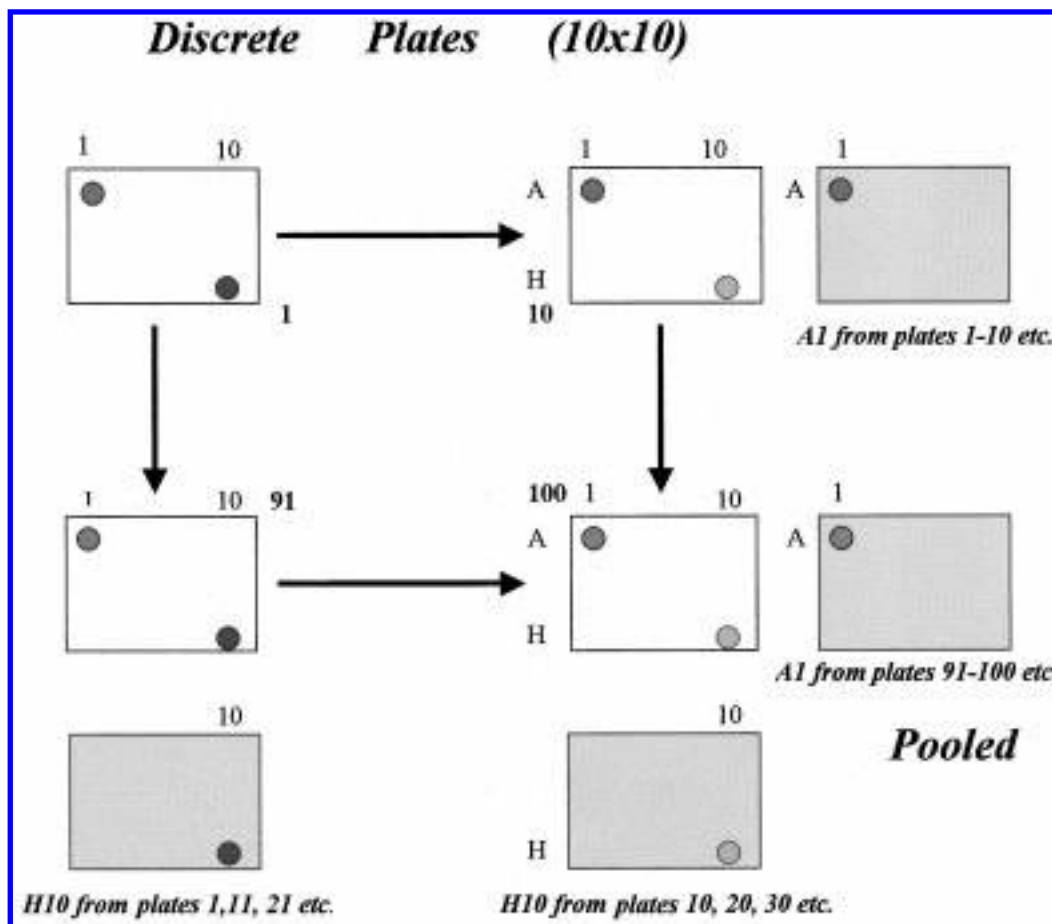
Figure 2. Schematic diagram of the method for orthogonal pooling.

direction. Provided that no very similar compounds occur within 80 Burden order numbers, this process ensures that very similar compounds are not put into the same well. In practice, as some chemical classes had less than 8000 compounds available for pooling, the number was made up to 8000 by using compounds classified as neutral. For example, compounds in the AE matrix element (acid/electrophiles) can be mixed with compounds in the E matrix element (neutral/electrophiles), etc. (Figure 1)

Liquid Chromatography−Mass Spectrometric Analysis (LC-MS). Liquid chromatography−mass spectrometric (LC-MS) analysis was performed using a Gilson 233XL autosampler (Anachem, Luton, Beds, U.K.), a Hewlett-Packard 1100 HPLC system (binary pump, vacuum degasser, and UV-diode array detector, Hewlett-Packard Ltd., Stockport, Cheshire, U.K.), and a Micromass Platform LC mass spectrometer fitted with an electrospray ionization probe (Micromass Ltd., Manchester, U.K.). The HPLC separation was carried out on a Supelcosil ABZ+ column (3.3 cm × 4.6 mm i.d., 3 μm particle size, Supelco U.K., Poole, Dorset, U.K.). Solvent flow through the HPLC system was set at 1.0 mL/min. This flow was split using an LC Packings Accurate postcolumn splitter (Presearch Ltd., Letchworth Garden City, Hertfordshire, U.K.) such that the flow into the mass spectrometer was at 20 μL/min. Two solvents were used for the HPLC separation. Solvent A consisted of ultrapure water containing 0.1% (v/v) formic acid (Fisher Scientific Ltd., Loughborough, Leicestershire, U.K.) and ammonium acetate (10 mM, Fisher Scientific). Solvent B consisted of acetonitrile/ultrapure water (90%/10% (v/v),

Rathburn Chemicals Ltd., Walkerburn, Scotland) containing 0.07% (v/v) formic acid and ammonium acetate (10 mM). The HPLC separation was carried out by gradient elution (column equilibrium: 0−1 min, 100% solvent A; 1 min−10 min, 100% solvent A − 100% solvent B; 10 min−14 min, 100% solvent B; 14−15 min: 100% solvent B − 100% solvent A) using diode array detection at 190−600 nm.

Mass spectrometric analysis was carried out using electrospray ionization (ESI) running in positive mode only. The mass spectrometer was set to scan from 100 to 1100 amu in 1 s, and the probe temperature was set at 80 °C.

Compound Pooling Validation Experiment. A set of 160 discrete compounds from our in-house collection was organized into 16 distinct chemical pool types (Figure 1). In cases where there were fewer than 10 compounds in a pool type, the set was made of up to 10 compounds with extra samples from pool 16 (neutral, inert samples). Each of the 10 compounds, dissolved in dimethyl sulfoxide (DMSO), was then mixed, either by making a pool of the 10 samples from within each pool type to give a final concentration for each compound of 0.5 mg/mL (strategic pooling) or by mixing one sample from each of the first 10 pool types to give six new pools, again with a final concentration for each sample of 0.5 mg/mL and thus increasing the chance that samples in each pool would react together (reactive pooling). Although not all possible combinations of functional classes in the reactive pools have been represented, the intention was to demonstrate the principle that mixing reactive compounds results in a significant increase in new compounds formed.

STRATEGIC POOLING OF COMPOUNDS FOR HTS

J. Chem. Inf. Comput. Sci., Vol. 39, No. 5, 1999 **901**

Table 1. Analytical Results for Strategic, Reactive, and Random Compound Pools

	time zero		month 1		month 3	
pool type	no. of degradation products	no. of new peaks	no. of degradation products	no. of new peaks	no. of degradation products	no. of new peaks
strategic[a]	19	2 (1%)	35	3 (2%)	47	14 (9%)
reactive[b]	10	6 (10%)	16	34 (57%)	22	47 (78%)
random[c] (solids)	1	1 (3%)	3	5 (17%)	4	6 (20%)
random[d] (liquids)	N/A	N/A	N/A	N/A	33	27 (27%)

[a] Sixteen pools with 10 compounds per pool. [b] Six pools with 10 compounds per pool. [c] Three pools with 10 compounds per pool. [d] Ten pools with 10 compounds per pool. The percentage of new peaks was calculated as % new peaks = total number of new peaks/total number of expected components × 100.
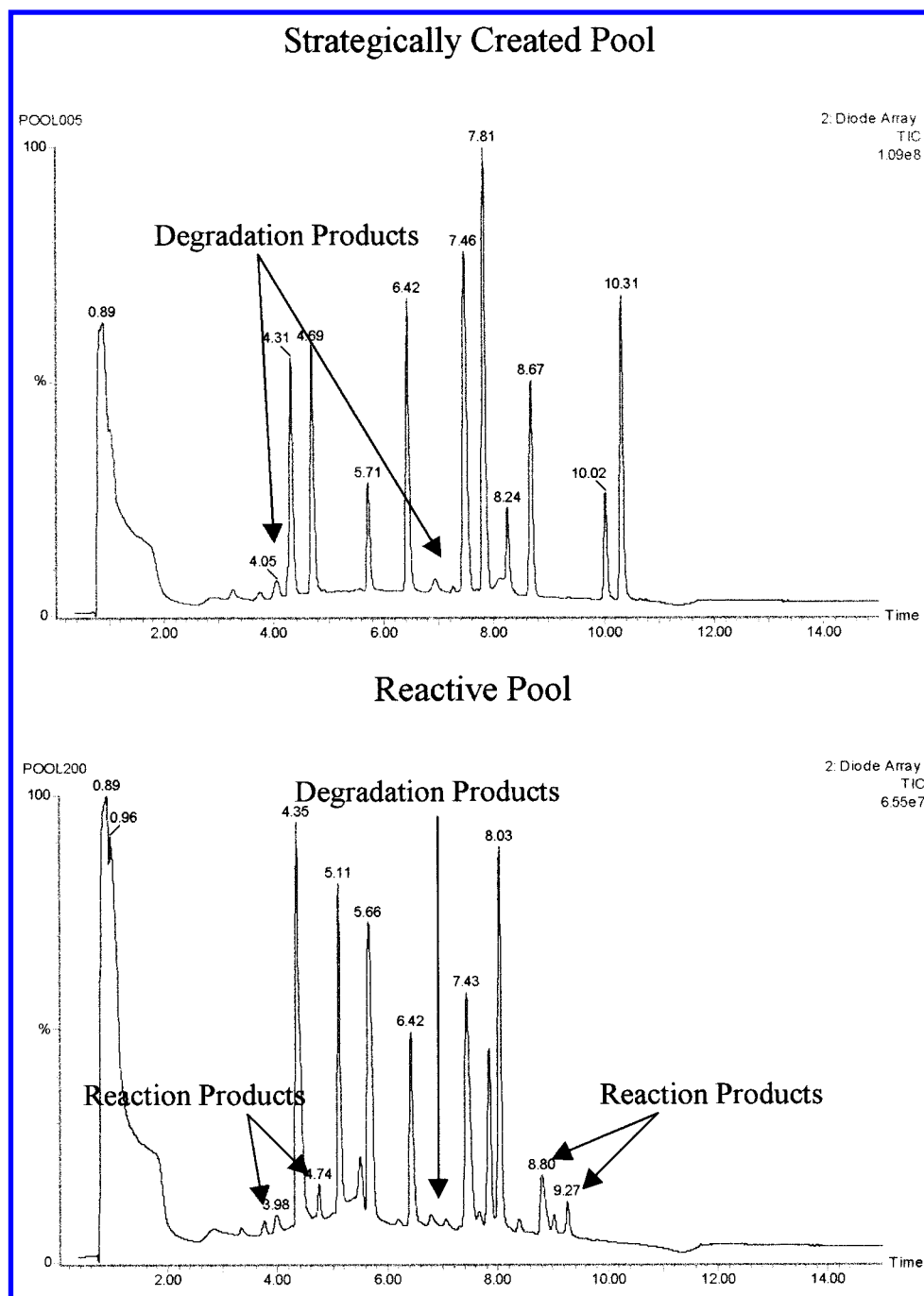
Figure 3. HPLC traces (UV detection) for a strategic pool and a reactive pool.

Ten random pools (10 compounds per pool) were also selected, together with the 100 discrete liquid stock solutions that were used to make these random pools. The random pools had been created 3 months prior to their analysis, and the discrete stock solutions used to create these pools had been stored at −20 °C for up to a year before the creation

of the pools. A time zero analytical result for these samples was therefore not available. However, the compounds that were present in three of these random pools were obtained as fresh solid material. These solids were dissolved in DMSO, and the solutions were then used to create three random pools (identical in origin to three of the random pools obtained from the liquid stores). All samples were stored in 96 well micronic blocks sealed with a foil lid at 4 °C and 17% relative humidity.

RESULTS AND DISCUSSION

Chemical Reactivity. Following filtering of the in-house compound collection to remove unsuitable molecules, the intermolecule chemical reactivity issue was addressed by separating the compounds into 16 chemical categories. Figure 1 shows the distribution of the compounds within these chemical categories.

Chemical Analysis of Experimentally Pooled Compounds. To test experimentally the idea of intermolecule reactivity, 160 compounds from the in-house collection were separated into the 16 chemical categories and pooled in various ways (strategic, reactive, and random pooling).

Table 1 shows the total number of degradation products and the number of unexplained extra peaks found at 0, 1, and 3 months, following LC-MS analysis of the 16 strategic pools, the 6 reactive pools, the 3 random pools created from freshly dissolved solid compounds, and the 10 random pools created from liquid stock solutions (3 month results only). Degradation products of each compound were identified as those extra peaks present in both the discrete and the pooled compound and accounted for the majority of peaks seen in the strategic and random pools. It is also apparent (3 month analytical data in Table 1) that there are many more unexplained extra peaks formed in the reactive pools (78%) than in the strategic (9%) or random pools (25%—composite results from solid and liquid random pools). These are presumably due to new product formation, as a result of reactions between the components in the pool.

In general, such reactions will give rise to compounds with molecular weights greater than 500. Although the reaction products observed are small in comparison to the peaks from the original components (typically the unexplained peaks were found to be 10% of the size of the original component peaks; see HPLC traces in Figure 3), their high molecular weight might result in unrepeatable biological activity when the samples are deconvoluted compared to the discrete samples that made up the pool.

Because of the large number of components in each pool, it was difficult to identify individual components that may have reacted together. It was clear, however, that some of these extra peaks had been formed from the reaction of amines with esters in the pool.

We have provided preliminary experimental evidence that our new pooling approach reduces the amount of contamination of samples that can occur through unwanted reactions within pools. The strategy to avoid changing the ionic status of compounds by only pooling similarly charged molecules should also aid in reducing solubility issues. Additionally, the method of pooling whereby similar structures are not mixed should help in the isolation of the active compound from a mixture. The combination of all these features has now been implemented in a new strategic pooling approach. The results of testing these pools in high throughput screening will be reported when available.

ACKNOWLEDGMENT

Supporting Information Available: Tables showing the hard filters and their SMARTS definitions. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Brown, D. Future pathways for combinatorial chemistry. *Mol. Diversity* **1997**, *2* (4), 217−222.
(2) Leach, A. R. Structure-based selection of building blocks for array synthesis via the World-Wide Web. *J. Mol. Graphics Modell.* **1997**, *15* (3), 158−160.
(3) Liu, D. X.; Jiang, H. L.; Chen, K. X.; Ji, R. Y. A New Approach To Design Virtual Combinatorial Library with Genetic Algorithm Based on 3D Grid Property. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 233−242.
(4) Kubinyi, H. Combinatorial and computational approachesin structure-based drug design. *Curr. Opin. Drug Discovery Dev.* **1998**, *1* (1), 16−27.
(5) Bures, M. G.; Martin, Y. C. Computational methods in molecular diversity and combinatorial chemistry. *Curr. Opin. Chem. Biol.* **1998**, *2* (3), 376−380.
(6) Gillet, V. J.; Willett, P.; Bradshaw, J. Identification of Biological Activity Profiles Using Substructural Analysis and Genetic Algorithms. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 165−179.
(7) Devlin, J. J.; Liang, A.; Trinh, L.; Polokoff, M. A.; Senator, D.; Zheng, W.; Kondracki, J.; Kretschmer, P. J.; Morser, J.; et al. High capacity screening of pooled compounds: Identification of the active compound without re-assay of pool members. *Drug Dev. Res.* **1996**, *37* (2), 80−85.
(8) Daylight Chemical Information Systems, Inc., Santa Fe, NM, at: http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html.
(9) Burden, F. R. Molecular identification number for substructure searches. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 225−227
(10) Rusinko, A. GlaxoWellcome, RTP, personal communication.
(11) We thank Stan Young, Glaxo Wellcome, RTP, for the initial suggestion of using this approach in a related pooling strategy.

CI990423O