Determining the optimal size of small molecule mixtures for high throughput NMR screening

Kelly A. Mercier[a] & Robert Powers[a],*

[a]*Department of Chemistry, University of Nebraska Lincoln, 722 Hamilton Hall, Lincoln, NE, 68522-0304, U.S.A*

Abstract

High-throughput screening (HTS) using NMR spectroscopy has become a common component of the drug discovery effort and is widely used throughout the pharmaceutical industry. NMR provides additional information about the nature of small molecule-protein interactions compared to traditional HTS methods. In order to achieve comparable efficiency, small molecules are often screened as mixtures in NMR-based assays. Nevertheless, an analysis of the efficiency of mixtures and a corresponding determination of the optimum mixture size (OMS) that minimizes the amount of material and instrumentation time required for an NMR screen has been lacking. A model for calculating OMS based on the application of the hypergeometric distribution function to determine the probability of a 'hit' for various mixture sizes and hit rates is presented. An alternative method for the deconvolution of large screening mixtures is also discussed. These methods have been applied in a high-throughput NMR screening assay using a small, directed library.

Introduction

Traditional methods of screening targets against chemical libraries include cell-based assays, surrogate systems, and systems to measure nucleic acid-protein interactions and receptor-ligand interactions (Fernandes, 1998). Libraries comprising hundreds of thousands of compounds can be screened in a short time period, where a particular robotic system has demonstrated a screening rate of 1,000 microtitre plates per 24 hours (Wallace, 1998). Although an efficient means to screen very large chemical libraries for activity against a specific protein target, these traditional techniques generally provide no indication of the mechanism of inhibition or verification that a binding interaction between the target and ligand have actually occurred.

*To whom correspondence should be addressed. E-mail: rpowers3@unl.edu

High-throughput screening (HTS) using nuclear magnetic resonance (NMR) spectroscopy has become a common component of the drug discovery effort and is widely used throughout the pharmaceutical industry because of the unique ability of NMR to provide direct evidence of a specific binding interaction between a potential chemical lead and the protein of interest (Klaus and Senn, 2003; Huth and Sun, 2002; Sem and Pellecchia, 2001; Roberts, 2000; Moore, 1999b; Moore, 1999a; Peng et al., 2001). Additionally, NMR may be used to evaluate the physical properties of a chemical lead, measure K_D's (Fielding, 2003) identify ligand binding sites (Roberts, 2000), and determine a co-structure (Clore and Gronenborn, 1994; Cooke, 1997; Kay, 1997; Roberts, 2000). A diverse number of NMR screening approaches have been developed, which include SAR by NMR (Shuker et al., 1996; Hajduk et al., 1997a; Hajduk et al., 1997c; Hajduk et al., 1999b; Johnson et al., 2003), SHAPES (Moore et al., 2004; Lepre et al., 2002;

Fejzo et al., 1999), and MS/NMR (Moy et al., 2001). NMR spectroscopy is a relatively insensitive technique requiring higher amounts of material and acquisition time compared to standard methods used in traditional HTS assays. Thus, a fundamental issue with NMR screens is a need to optimize the efficiency of sample throughput by achieving a balance between information content and resource utilization. As a result, NMR-based assays utilize chemical libraries that are significantly smaller in size compared to the hundreds of thousands to millions of compounds typically screened in an HTS assay.

An approach used to address the fundamentally lower throughput of NMR has been the development of small, directed compound libraries that are more amenable to NMR-based screens (Huth and Sun, 2002; Jacoby et al., 2003; Baurin et al., 2004; Villar et al., 2004). The SHAPES library is a typical example of the fragment based approach to NMR screening, where the library consists of a small, structurally diverse set of water soluble compounds that correspond to fragments or molecular frameworks of known drugs (Lin et al., 1997; Fejzo et al., 1999; Johnson et al., 2003; Lepre et al., 2002; Moore et al., 2004). A comparable approach to reduce the size of screening libraries is to use NMR-based assays as a secondary screen to validate hits from HTS assays. In this manner, the HTS assays reduce a large corporate library to a small, focused list composed of a few hundred to a few thousands compounds that is more applicable to an NMR screen (Hajduk and Burns, 2002; Jahnke and Widmer, 2004).

Since chemical libraries can still number in the thousands of compounds, even for directed libraries, mixtures of small molecules are often screened against a target to minimize resource utilization while increasing throughput (Jacoby et al., 2003; Chen and Shapiro, 1999; Lin et al., 1997; Meyer et al., 1997). Screening mixtures of 5 to 100 compounds have been described where it may be feasible for an NMR assay to screen upwards of hundreds of thousands of compounds (Devlin et al., 1996; Glick et al., 2003; Hann et al., 1999; Jacoby et al., 2003; Pratt Steven et al., 2004; Hajduk et al., 1999a). Thus, the use of mixtures makes an NMR assay readily amenable for screening smaller, directed libraries with the potential to screen larger, random libraries comparable to standard HTS assays.

There are some potential issues associated with the application of mixtures that impacts their wide-spread use in traditional HTS (Schriemer and Hindsgaul, 1998). A major concern is the observation that mixtures increases the inherent 'noise' of an HTS assay by either increasing the occurrence of false positives or false negatives (Glick et al., 2004; Pratt Steven et al., 2004). Another practical concern is the proper composition of the mixtures used for screening. Factors such as solubility, total organic concentration, structural diversity and compound reactivity may potentially limit the utility of mixtures in a screening endeavor (Brown et al., 2000; Brown and Martin, 1997; Schriemer and Hindsgaul, 1998; Hann et al., 1999). It is particularly challenging to design appropriate mixtures that adhere to these needs, especially for larger library and mixture sizes (Brown and Martin, 1997; Glick et al., 2003). Additionally, improvements in miniaturization and automation that continually improves the throughput of HTS assays diminishes any perceived advantage of mixtures (Dove and Marshall, 1999; Smith, 2002).

Despite these limitations, compound mixtures are routinely used in NMR assays because of the significant advantage that is achieved in increasing throughput. In general, mixture sizes in NMR screens appear to be chosen rather arbitrarily, where the simple application of mixtures achieves the main goal of improved throughput. No significant consideration has been given to determine an optimal mixture size (OMS) for an efficient implementation of an NMR-based assay to minimize resources. Deconvolutions of mixtures to identify the active compound results in an increase in the total number of NMR experiments required to screen the entire library. The impact of the deconvolution step may overwhelm any advantages in efficiency gained by screening a specific mixture size compared to alternatives. Thus, the size of the mixture chosen for a particular screen directly determines the total number of NMR experiments that are required and establishes the efficiency of the assay. An increase in mixture size results in a proportional decrease in the number of primary NMR experiments. But, basic probability indicates that the likelihood of finding a hit and the need to deconvolute a mixture will also scale with an increase in mixture size. Furthermore, the total number of deconvolution

experiments is a direct product of the size of the mixture. Thus, determining the OMS requires finding a balance between these two opposing trends: a decrease in primary experiments and an increase in deconvolution experiments as a function of mixture size. We describe a model based on the hypergeometric distribution function (Feller, 1968; Spiegel, 1992) that determines the optimum mixture size and corresponding efficiency of an NMR-based assay for a range of 'hit' rates.

Materials and methods

Statistical analysis

Because the hit rate is low within a random compound library, it is usually expected that only a small percentage of the screened mixtures will contain even a single hit. Random libraries usually have hit rates on the order of 0.1–0.5% (Dove and Marshall, 1999), but focused chemical libraries, such as SHAPES used for NMR assays, may exhibit significantly higher hit rates of 0.7–20% or higher (Lepre et al., 2002; Jahnke and Widmer, 2004). The hit rate, thus, plays a significant factor in determining the optimum size of a mixture.

An initial reasonable expectation would be that the number of NMR experiments collected for a given library size will continue to decrease, albeit slower, with increasing mixture size, as long as larger mixtures are practically attainable (Hann et al., 1999; Brown and Martin, 1997; Glick et al., 2003; Brown et al., 2000; Schriemer and Hindsgaul, 1998). Although it may seem that a direct relationship exists between mixture size and the number of experiments needed, we propose instead that identifying the OMS is best described through the use of a hypergeometric distribution function (Feller, 1968; Spiegel, 1992). The 'urn problem' is a classic illustration of the application of the hypergeometric distribution function. The urn problem involves sequentially selecting balls from an urn containing a fixed number of two different colored balls. The function describes the probability of pulling out one color verses another for a given sample size. This is perfectly analogous to the issue of defining the optimal mixture size where the probability of a mixture containing a hit is dependent on the distribution of inactive compounds (color one) and hits (color two) in a library. Again, this process is identical to creating a set of mixtures from a defined compound library (N) by randomly selecting n compounds from the library until N/n mixtures are generated and all the compounds in the library have been used. A hit is never guaranteed or limited in number in a mixture, and the hypergeometric distribution function accurately describes the distribution of hits as completely random in nature. The equation for the hypergeometric distribution function is defined as:

$$P(X = x) = h(x : n, M, N) = \frac{\binom{M}{x}\binom{N - M}{n - x}}{\binom{N}{n}}$$

(1)

where P is the probability of a mixture containing at least one hit, x is the number of hits present in a mixture, n is the number of compounds in the mixture, M is the number of hits present in the library, and N is the total population of the library.

The Winstats software program was used to determine the probability of a hit for a given sample (mixture) size and hit rate (12). A hypothetical chemical library of 200,000 compounds was used for all the simulations, where the observed trends scale proportionally with the size of the library. In the absence of deconvolution, the required number of experiments scales simply by N/n (Figure 1). When deconvolution is required, the probability of one or more hits being present in a mixture needs to be considered to determine the total number of experiments (T):

$$T = \left(\frac{N}{n}\right) + (N)(P)$$

(2)

The additional experiments attributed to deconvolution is simply determined by the number of mixtures containing one or more hits and the size of the mixture. This in turn is defined by the probability of a mixture having one or more active compounds times the total number of mixtures (N/n) and the total number of compounds per mixture (n). This simply reduces to (N)(P). After we completed our study, a similar analysis of preferred mixture size in the context of designing a combinatorial chemical library came to our attention (Teixido et al., 2000). Teixido et. al (2000) uses a simple and fundamentally flawed
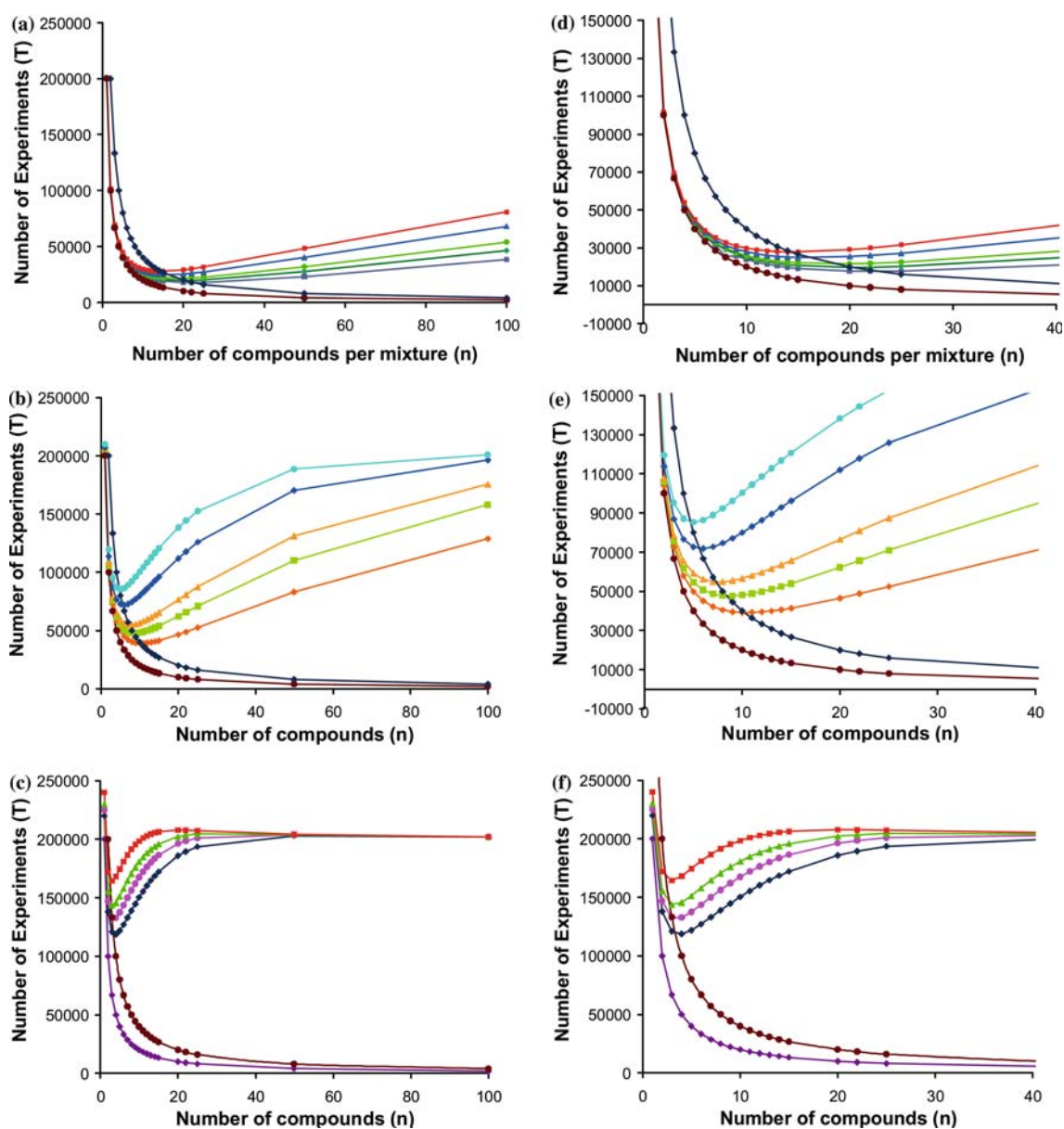
Figure 1. The OMS as a function of mixture size with no deconvolutions(◆—◆) and strategic pooling(●—●) are plotted in all graphs. The OMS as a function of mixture size requiring deconvolution are plotted at different hit rates: (a) 0.20%■—■, 0.25%◆—◆, 0.30%◆—◆,0.40%▲—▲, 0.50%■—■; (b) 1.0%◆—◆, 1.5%■—■, 2.0%■—■, 3.5%▲—▲, and 5.0%●—● and; (c) 10%◆—◆, 12.5%●—●, 15%▲—▲, 20%■—■. (d-f) Expanded view of graphs from a-c, respectully. The OMS for hit rates between 0.20–0.50% is 5–11. The OMS for hit rates between 1.0–5.0% is15–20. For hit rates >10%, the graphs (c, f) clearly shows that using mixture sizes for these larger hit rates can lead to much higher total number of experiments. For the larger mixture sizes proposed at these hit rates, using the singleton method is preferred as each mixture will have at least 1 hit with P = 1. For these cases, all of the mixtures would have to be deconvoluted leading to a great number of total experiments compared to screening without mixtures.

assumption especially in the context of NMR-based screens. Their analysis is based on the assumption that the active compounds are uniformly distributed throughout the mixtures where a mixture will contain either zero or one active compound. This is a reasonable assumption for very low hit rates and small mixture sizes where the hypergeometric distribution function indicates
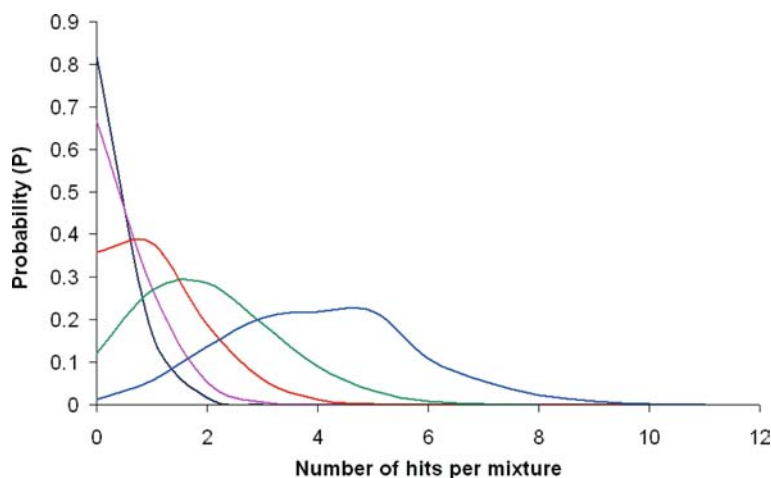
Figure 2. A probability distribution for different hit rates for mixtures of 20 compounds. The probability of having 0–10 hits per mixtures ultimately depends on the hit rate: 1% ——; 2% ——; 5% ——; 10% ——; 20% ——. For the larger hit rates, there is a higher probability of having more than 1 hit per mixture.

a vanishingly small probability of a mixture containing more than one hit, but this analysis provides erroneous results as both hit rate and mixture size rise. Figure 2 clearly illustrates the impact of hit rate on the distribution of hits per mixture. There is a point where the probability shifts from a majority of mixtures lacking an active compound to the situation where the most common event is for a mixture to contain two, three or more active compounds. The application of the hypergeometric distribution function clearly indicates that an increasing hit rate continually shifts the ratio of the number of hits per mixture to values greater than one.

For a given mixture size, it is *always* more efficient to screen a library by avoiding deconvolution, but it also suggests that there is a situation where it is more efficient to screen a larger mixture size requiring deconvolution compared to a smaller mixture size that avoids deconvolution. Thus, using an OMS (n_1) for a given hit rate is more efficient than using a smaller mixture size (n_2) that avoids deconvolution if the following condition holds:

$$\frac{1}{n_1} > \frac{1}{n_2} + P \qquad (3)$$

This condition holds for the typical 0.1–0.5% hit rates for random libraries but diminishes quickly with increasing hit rates associated with directed or focused libraries. Figure 3 shows how the OMS decreases with an increasing hit rate.

An alternative method to the single deconvolution step described above, is a bucket-sort approach that increases the efficiency of screening larger mixtures (>20) by diminishing the impact of deconvolution. The bucket-sort method subdivides large mixtures (n_1) with identified hits into a second set of smaller mixtures (n_2). As before, the hypergeometric distribution function determines the probability of identifying a hit in each of the mixtures sizes (P_1, P_2). Again, the total number of deconvolution experiments is based on the number of mixtures containing a hit and the size of each mixture. The total number of experiments using the bucket- sort method is:

$$T = \left(\frac{N}{n_1}\right) + \left(\frac{N}{n_2}\right)(P_1) + (N)(P_1)(P_2) \qquad (4)$$

The difference in the standard deconvolution method and the bucket-sort is clearly illustrated in Figure 4. For lower hit rates, the total number of experiments needed for the bucket sort method is significantly smaller. However, it is clear that as the hit rate increases, the bucket sort method begins to approach the same results as the standard method.

It is important to note that the hit rate (HR_2) for the second set of mixtures in the bucket sort approach is not equivalent to the hit rate (HR_1) for the library. The bucket-sort approach effectively
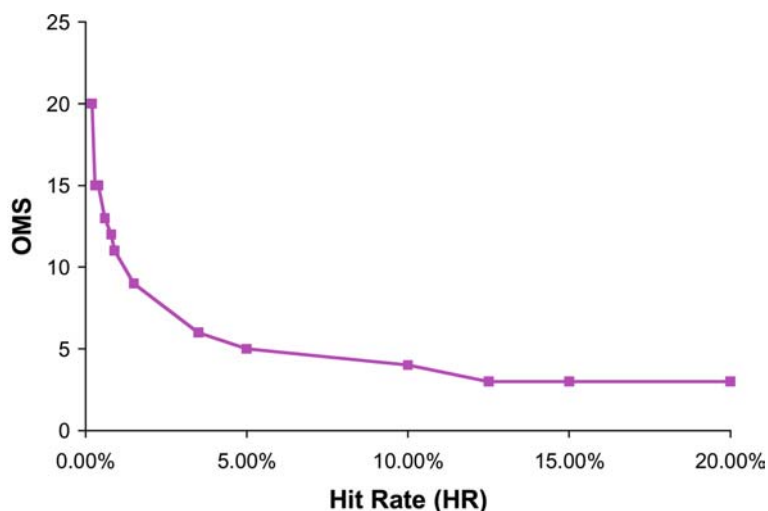
Figure 3. A plot of optimal mixture size (OMS) as a function of hit rate (HR).
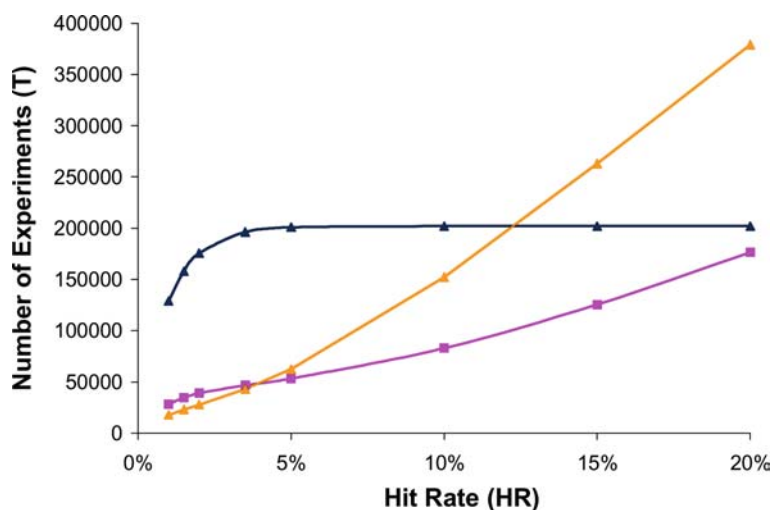

Figure 4. The difference between standard deconvolutions with mixtures of 100 (—▲—) and the bucket sort approach with mixtures of 100 and 10 (—▲—) and 100 and 5 (—■—). As the hit rate increases, the total number of experiments for the bucket sort approaches the results for the standard deconvolution method and may exceed the standard deconvolution method if mixture sizes are not optimized.

concentrates the number of hits in the second set of mixtures. This occurs because the total number of hits remains constant but the total number of compounds that is screened in the second set of mixtures is decreased since mixtures devoid of a hit are discarded. The two hit rates are related by the probability of identifying at least one hit in the initial large mixture size:

$$HR_2 = \frac{HR_1}{P_1} \qquad (5)$$

HR_2 can then be used to determine a probability for identifying a hit in the second smaller set of mixtures and a corresponding OMS for the second screen in the same way as before. Of course, there is a finite limitation to the utility of the bucket-sort approach. For a given library size, as hit rate and size of the large mixture increase, the probability of identifying at least one hit in every mixture rapidly approaches unity. Figure 5 illustrates the results of the OMS analysis as a function of mixture size for the second deconvolution step in the bucket-sort method.
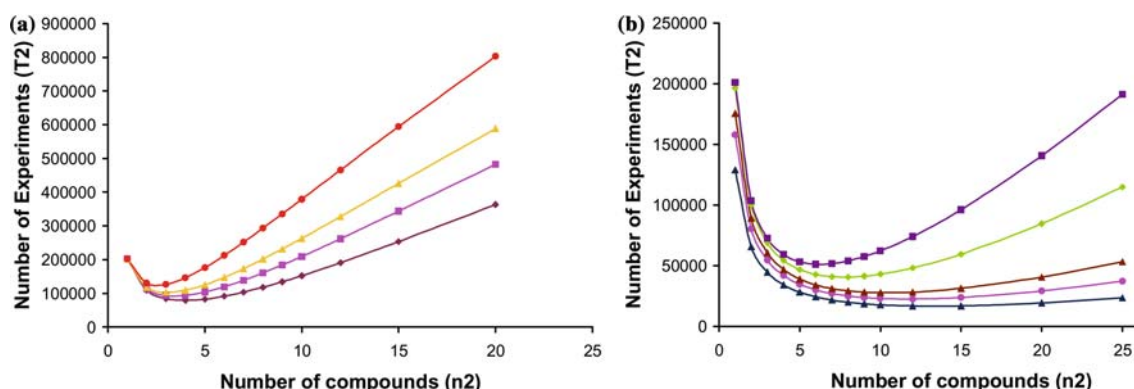
Figure 5. A plot of OMS as a function of mixture size for the second screening step using the bucket-sort method. (a) hit rates of: 10%—◆—; 12%—■—; 15%—▲—; and 20—●—. (b) hit rates (HR_2) of: 1%—▲—; 1.5%—●—; 2%—▲—; 3.5%—◆—; 5%—■—. OMS ranges from 6–12 for HR_2 between 1–5%. At higher hit rates, OMS is approximately 1–3 compounds. For consistency, each mixture that has a hit for the second screen was calculated assuming deconvolution, but this may not be necessary for smaller mixture sizes.

Compound library

A directed small compound library composed of approximately 300 compounds is used for the NMR screen. The compounds are structural diverse and soluble to 100 μM in an aqueous buffer. Based on the described OMS analysis and expected high hit rate (0.7–20%) for a directed chemical library, mixtures were designed to avoid a necessary deconvolution step. Avoiding deconvolution in an NMR screen requires identifying combinations of compounds that yield at least one unique NMR resonance per compound. A high resolution was desired so that a binding event could be absolutely assigned to a specific compound in the mixture by observing line-width changes in the ^{1}H NMR spectrum upon addition of the protein. Because the ^{1}H spectral window is fairly narrow, only mixture sizes of 3–4 compounds were readily obtained that fit this criteria. Screening this directed compound library by NMR has yielded a hit rate of 7.3%, which is consistent with the higher hit rates observed with other directed (0.7–20%) libraries. It is reasonable to expect that the absolute value for the hit rate will vary depending on the particular protein that is screened.

Special care was taken when designing the mixtures to minimize the possibility of chemical interactions between the compounds within the mixture. To ensure that no chemical reactions or interactions had taken place, reference NMR spectra were acquired for each individual mixture where chemical shifts and peak intensities were compared between the NMR spectra of the mixtures and the corresponding individual compounds. In all cases, the NMR spectrum for each compound in a mixture were essentially identical to its individual NMR spectrum. Additionally, each mixture is composed of structurally unique compounds to decrease the likelihood that multiple active compounds are present in the mixture.

NMR sample conditions

A 20 mM stock solution in D_2O or d-DMSO for each compound in the library has been generated and is stored at −80 °C. A reference ^{1}H NMR spectrum of each compound was collected to ensure reasonable solubility and stability. Additionally, reference spectra confirmed consistency of the NMR spectra for each compound in its corresponding mixture eliminating reactivity or interactions between the mixture compounds. The NMR samples contained 100 μM of compound, 5% DMSO and 20 mM of d-Bis Tris buffer at pH 7.0. The NMR spectra were collected on a Bruker 500 MHz Avance spectrometer equipped with a triple-resonance, Z-axis gradient cryoprobe and a BACS-120 sample changer. ^{1}H NMR spectra were collected with solvent presaturation and a total acquisition time of 12 minutes, where S/N ≥ 4 was required to keep the compound in the chemical library. Figure 6 is an example of the quality of the NMR data and the typical compounds comprising a mixture. Given these experimental conditions for the NMR screen, a
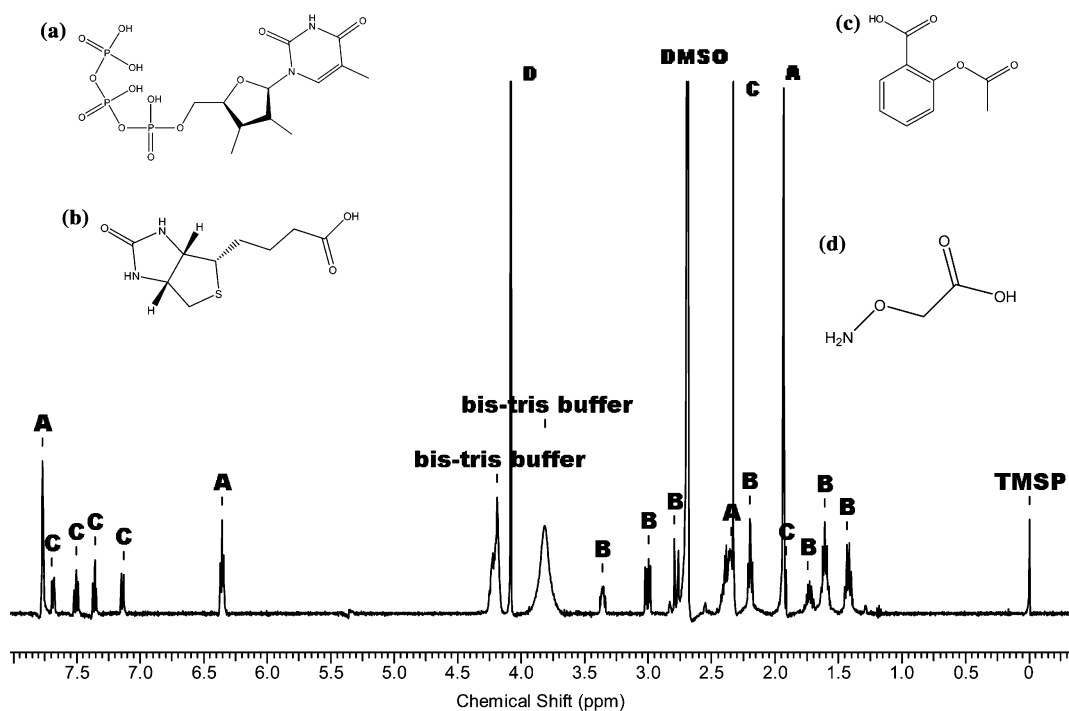
Figure 6. An example of a typical NMR spectrum for a mixture of 4 compounds from our screening library. Each NMR resonance has been assigned to one of the compounds in the mixture, d-DMSO (2.52 ppm) or TMSP (0.0 ppm). The compound names are as follows: a, Thymidine 5′-triphosphate sodium salt (TTP); b, Biotin (vitamin H); c, Acetylsalicylic acid; d, O-(Carboxymethyl)hydroxylamine.
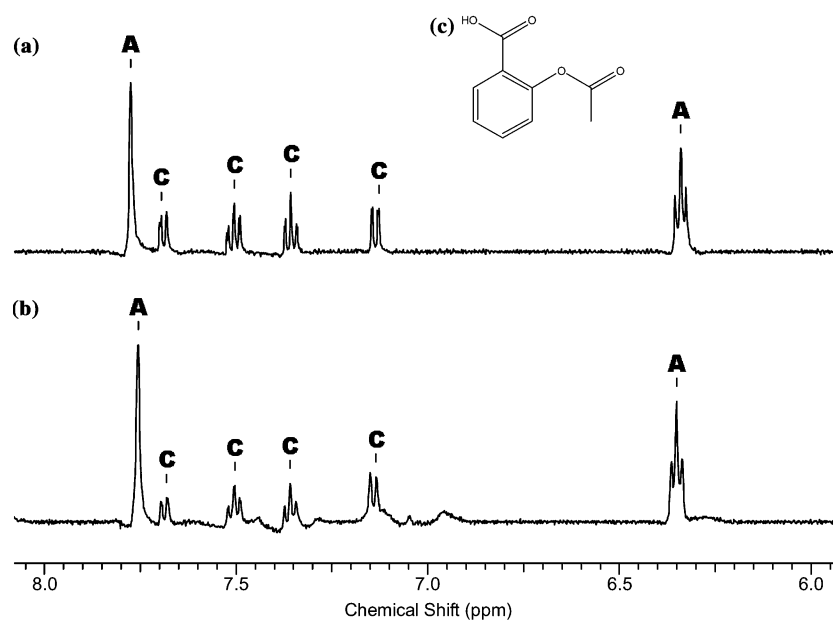

Figure 7. Example of a positive binding event from the 1D NMR line-broadening screen. The expanded aromatic region of Mixture 1004 (a) without and (b) with SAV1430 illustrating the induced change in linewidths. Only the NMR resonances assigned to acetylsalicylic acid broaden in the presence of SAV1430. The NMR resonances for the remaining compounds in the mixture are unchanged. The additional broad resonances in spectrum B are from the protein.

conservative estimate for the upper-limit for an observable K_D is ∼500 μM–1 mM. This is based on a conservative estimate that a 10–20% change in the line-width of NMR spectra is readily observable with a 15–25 Hz line-width for a modest size protein (15–25 kDa). The K_D upper-limit will increase for larger molecular-weight proteins since a larger line-width change will occur for a given fraction bound ligand.

After a reference NMR spectrum was collected for each mixture of compounds, 25 μM of SAV1430 from *Staphlococcus aureus* was added to each NMR tube and a second 1D ^{1}H spectrum of the mixtures was collected for comparison. A positive binding event is identified by an observed line-width change for an NMR resonance attributed to a specific compound in the mixture (Figure 7).

Discussion

Identifying the optimal mixture size (OMS)

Pharmaceutical companies have large libraries of compounds that range from the hundreds of thousands to millions of compounds, which have evolved from decades worth of synthetic efforts, from acquisition or more recently from combinatorial chemistry approaches (Armstrong, 1999; Gonzalez and Negulescu, 1998; Oldenburg, 1998; Fernandes, 1998; Kenny et al., 1998; Silverman et al., 1998). The inherently low-sensitivity and long acquisition times necessary to acquire even the simplest NMR experiment limits the feasibility of screening a corporate library using a one compound per sample approach. Thus, the advantage of mixtures is both apparent and paramount to the successful application of NMR in high-throughput assays that require screening thousands of compounds in a reasonable time-frame, where an order of magnitude improvement in throughput may be achieved compared to screening singletons (Meyer et al., 1997; Lin et al., 1997; Jacoby et al., 2003; Hajduk et al., 1999a; Dalvit et al., 2003; Chen and Shapiro, 1999). There are multiple acceptable paradigms that contribute to the proper design of individual mixtures in a library. The major concerns are: (i) minimizing reactivity and interactions of compounds, (ii) maximizing structural diversity, (iii) maximizing solubility and (iv) maintaining consistent physical properties (pH, ionic strength,

total organic concentration) (Schriemer and Hindsgaul, 1998; Hann et al., 1999; Jacoby et al., 2003; Brown and Martin, 1997; Gorse and Lahana, 2000; Brown et al., 2000). One desirable impact of the nature of corporate libraries and these design criteria is the result that mixtures can be considered generally random in composition. Random mixture composition implies that the likelihood of achieving a hit against a target in a particular mixture is also random and dependent on the total number of 'active' compounds in the entirety of the library.

The efficiency of screening does not simply scale with the size of mixtures but depends on the design of the NMR experiments and the nature of mixtures being used to screen the chemical library. The efficiency of an NMR screen can be measured by the total number of NMR experiments (T) that are required to screen the entire library and identify all the 'active' compounds. If deconvolution of the hits is unnecessary, than efficiency will simply scale as a function of the mixture size (N/n) (Figure 1). A more typical situation is for an NMR screen to require the deconvolution of a mixture to identify the hit(s). The impact of deconvolution on the total number of NMR experiments or the efficiency of the screen is simply determined from the number of mixtures with an identified hit times the size of the mixtures (Equation. 2).

Thus, the true test of the efficient use of mixtures is dependant on the probability of a hit occurring in any given mixture. As the probability of a hit being present in a mixture increases, the number of deconvolution experiments will also increase while decreasing the efficiency of the mixture size. The probability of a hit being present in a mixture will increase with both the size of the mixture (more chances to find a hit) and the hit rate (more number of hits) (Figure 2). This scenario is perfectly analogous to the classic 'urn problem' where the hypergeometric distribution function (Equation. 1) can be applied to determine the probability (P) that one or more 'hits' will be present in any given mixture based on the size of the mixture and the 'hit' rate. (Feller, 1968; Spiegel, 1992). The total number of NMR experiments required to screen a compound library can then be determined from the probability of finding a hit within a given mixture (Equation 2). As illustrated in Figure 1, plotting mixture sizes as a function of the total number of NMR experiments needed to

screen the entire library results in a minimum that identifies the optimal mixture size (OMS). The mixture size that corresponds to the minimum is directly dependent on the hit rate. OMS ranges between 5–20 compounds per mixture over a hit rate of 5% to 0.2%, where OMS decreases as the hit rate increases (Figure 3). The absolute hit rate is an unknown quantity prior to executing a screen, but an expected value can be predicted based on the composition of the screening library. For directed libraries that are routinely used in NMR higher hit rates are anticipated (0.7–20% or higher) (Lepre et al., 2002; Jahnke and Widmer, 2004). Our analysis of OMS would predict that a mixture size of 5 compounds would be a reasonable choice for the high hit rates predicted for NMR screens of directed compound libraries.

For smaller mixture sizes where the probability of a hit in a mixture is diminished the trend approximates N/n. Initially increasing mixture size improves the efficiency of the screen, but as the mixture size continues to increase deconvolution becomes a significant component of the total number of experiments. As a result, there is a point where a smaller mixture size is more efficient at minimizing the total number of experiments compared to larger mixtures because of the large negative impact of deconvolution. Consider a typical hit rate of 0.25% and a theoretical library comprising 200,000 compounds, for a mixture size of 5 compounds a total of 42,480 experiments are predicted. This total actually *increases* to 46,300 experiments for a mixture size of 100 because of the impact of deconvolution.

Increasing hit rate has a negative impact on the absolute efficiency of an NMR-based screen using mixtures. This arises because hit rate only contributes to the total number of necessary deconvolution steps resulting in a proportional increase in the total number of experiments required to screen the library regardless of mixture size. An increase in hit rate also results in a shift to a lower mixture size for the identified OMS. As hit rate rises, the probability of an active compound being present in a mixture also increases (Figure 2). This results in a corresponding increase in the number of deconvolution steps and a subsequent decrease in the efficiency of the mixture size. Conversely, a decrease in the mixture size diminishes the probability of identifying a hit since fewer chances are available for 'pulling' an active compound out of the library. Thus, a smaller OMS compensates for the higher hit probability to maintain a minimal number of NMR experiments.

While corporate libraries are screened as mixtures in traditional HTS assays (Schriemer and Hindsgaul, 1998; Appel et al., 1999; Pratt Steven et al., 2004), the use of mixtures has not achieved wide-spread acceptance because of a continuing improvement in throughput from miniaturization and automation which diminishes any perceived advantage of mixtures (Dove and Marshall, 1999; Smith, 2002). Concerns about increases in false positives and increases in total organic concentrations combined with practical challenges in designing chemical libraries and the need to deconvolute hits further limits the inherent value of mixtures in HTS (Glick et al., 2003; Schriemer and Hindsgaul, 1998; Pratt Steven et al., 2004; Glick et al., 2004; Brown et al., 2000; Brown and Martin, 1997). Nevertheless, the analysis of OMS in the context of NMR-based screens is also applicable to HTS assays. Our analysis of OMS clearly indicates that a significant improvement in efficiency on the order of 7–10 fold can be obtained compared to screening the library as singletons. While deconvolution limits the total gain that can be obtained from screening mixtures, a significant advantage is still achieved.

Conditions where deconvolution of mixtures should be avoided

Our analysis makes it clear that larger mixture sizes can be detrimental to the efficiency of the NMR screen due to the impact of the deconvolution step (Figure 1c, f). For higher hit rates (>5%), even small mixture sizes (<5) have a significant probability of containing a hit, and unlike low hit rates, the trend does not approximate the non-deconvolution (N/n) condition. As a result, a large improvement in efficiency is obtained by using smaller mixture sizes that avoids deconvolution except for the extreme case of a mixture size of one. Thus, while our analysis of OMS predicts a reasonable choice of a mixture size of 5 compounds for NMR screens with high hit rates, the analysis also indicates that the efficiency of the screen improves dramatically if deconvolution can be completely avoided even if a smaller mixture sizes (<5) are employed. Also, smaller mixture sizes tend to be more readily achievable when

factors such as solubility, reactivity, chemical diversity and spectral overlap are considered.

Additionally, as the size of the mixture increases the total number of NMR experiments required to screen the library asymptotically approaches $N + N/n$. This occurs at relatively small mixture sizes for the larger hit rates, and quickly exceeds the total number of NMR experiments (N) required if the compounds were simply screened as singletons. Obviously, these additional experiments would negatively impact the cost of the assay by a needless increase in instrument time, analysis time, and protein material and should be avoided. High hit rates result in the probability of a hit being present in any given mixture to rapidly approach unity, completely negating any value in using mixtures that requires deconvolution.

Hit rates that are observed in traditional HTS screens of large random compound libraries are typically $\leq 0.5\%$ (Dove and Marshall, 1999). This implies that screening HTS assays would be more efficient by using mixture sizes in the range of 15–20 compounds even when deconvolution is required. Conversely, chemical libraries screened by NMR have a high likelihood of having hit rates $> 5\%$. First, NMR is increasingly using small directed libraries with increased drug-like characteristics and a higher propensity to bind a protein because of the limited throughput of NMR screens compared to traditional HTS (Fejzo et al., 1999; Johnson et al., 2003; Lepre et al., 2002; Moore et al., 2004). Second, NMR screens tend to be more sensitive than traditional HTS where observing K_D's in the 100 μM to mM range are common (Hajduk and Burns, 2002). Finally, NMR-based assays are also increasingly being used as secondary assays to confirm hits identified from HTS (Hajduk and Burns, 2002; Jahnke and Widmer, 2004). In this context, the expected hit rate would be very high when using these enriched compound libraries where it is theoretically possible to approach 100%. In practice, the hit rates tend to be significantly below 100% due to a large number of undesirable mechanisms that led to an observed activity in a biological assay that does not involve a specific protein-ligand interaction (McGovern et al., 2002; McGovern et al., 2003; Seidler et al., 2003; Rishton, 1997). Nevertheless, the hit rate is still expected to be significantly larger than the random hit rate of 0.1–0.5%. Our analysis of OMS indicates that under conditions of

high hit rate ($>5\%$) the use of mixtures for NMR-based screens that require deconvolution is extremely limited. Nevertheless, as long as the hit rate is below 29%, the minimal mixture size of 2 compounds will still be more efficient than screening the library as singletons.

Based on this analysis we have designed a focused compound library that uses mixtures to maximize the efficiency of an NMR screen while avoiding deconvolution. One means to rapidly identify a protein-ligand interaction is by measuring a change in line-width (T_2) (Rossi et al., 1992; Hajduk et al., 1997b). In this case, deconvolution is unnecessary if at least one NMR resonance can be uniquely attributed to each compound in the mixture. This is routinely and readily achievable by using mixtures composed of 3–4 compounds (Figure 6). A binding interaction to the targeted protein can be unambiguously attributed to one compound in the mixture without the need for deconvolution (Figure 7). This approach has routinely been applied to a screening library composed of $\sim$300 compounds and 70 mixtures where a hit rate of 7.3% has been observed. Thus, using this library configuration only requires 70 NMR experiments to screen the entire library.

One potential problem with the application of mixtures is the possibility of false negatives. This issue may arise if two or more active compounds are present in the same mixture. As described previously, NMR screens of directed chemical libraries tend to yield high hit rates with a correspondingly higher probability of multiple hits being present in a single mixture (Figure 2). Thus, false negatives are a particular concern for NMR screens. The most likely source of a false negative would be the presence of both a tight and weak binder in the mixture, where the tight binder completely displaces any observable interactions of the weak binder with the protein. This is not an issue if multiple binders are present with comparable binding affinities. Tight binders with K_D's $\leq \sim$10 nM would efficiently displace weaker μM-mM binders given our NMR sample conditions and result in false negatives. While the situation is unavoidable, it is not detrimental to the value of the screen. In general, observing tight binding compounds with K_D's $\leq \sim$10 nM are relatively rare events in the early stages of drug discovery where NMR screens are commonly utilized.

Correspondingly, the occurrence of a false negative will have a low probability. Conversely, in optimized compound libraries where tight binding ligands may be more common, the lose of identifying a significantly weaker ligand would presumably be inconsequential to the success of the project since the focus would be on identifying the tight binding ligands. In fact, this is the general situation where observing a tight binding ligand at the expensive of a weaker ligand would not be perceived as detrimental since the tight binding ligand generally has inherently more value to the project.

The problems with mixtures and deconvolution at this high of a hit rate can be further illustrated by conceptionally screening our small directed library as mixtures of 10 compounds. The initial number of experiments would be decreased by 57% (30 experiments) but probability predicts that 16 of these experiments would exhibit a positive hit requiring 160 deconvolution experiments for a total of 190 experiments. The result is an additional 120 NMR experiments with a corresponding $\sim$2.7 decrease in the efficiency of the NMR-based assay by screening the library at a higher mixture size that requires deconvolution. Of course, this is still more beneficial than screening the library under the extremely inefficient condition of singletons, which would require $\sim$300 NMR experiments.

Situations where deconvolution improves efficiency

For a given mixture size, it is *always* more efficient to screen a chemical library by NMR that avoids the need for deconvolution. Nevertheless, as apparent by the plots in Figure 1, there are numerous situations where a larger mixture size (n_2) that requires deconvolution is more efficient than a smaller mixture size (n_1) that avoids deconvolution. An extreme example is a mixture size of one, where the vast majority of mixture sizes and hit rate combinations will always be more efficient. In effect, this scenario is the direct opposite of the situation observed for high hit rates, where it is more advantageous to avoid deconvolution. The major factor that contributes to this scenario is the number of required deconvolution steps which directly depends on the probability of identifying a hit in a mixture (Equation. 3). Since this probability increases

proportionally with both an increase in mixture size and an increase in hit rate, any advantage of screening with a larger mixture size is eventually lost for hit rates $> 5\%$. In general, hit rates for traditional HTS assays tend to be significantly less than 5% (Dove and Marshall, 1999) implying that larger mixture sizes that require deconvolution are typically more efficient for traditional HTS assays compared to smaller (≤ 5) mixture sizes that avoid deconvolution. Again, this is the opposite observation for larger hit rates where it will never be advantageous to screen a mixture that requires deconvolution compared to mixtures that avoid deconvolution. NMR-based assays that are screening directed SHAPE-like libraries or confirming HTS hits will maximize efficiency by using smaller mixture sizes that avoids deconvolution.

Strategic pooling of compounds is a common method that attempts to avoid the need to deconvolute mixtures (Devlin et al., 1996). Each compound in the library is screened twice where a compound is only identified as a hit if both mixtures containing the same compound exhibits activity. The unambiguous assignment of activity to a specific compound is achieved by designing sets of mixtures where only a single compound is present in two pairs of mixtures. The success of this approach is based on the assumption that the low hit rate in the screen effectively eliminates the occurrence of multiple active compounds being present in the same mixture. Also, the low hit rate results in a low number of active mixtures that minimizes serendipitous overlap between multiple mixtures. As an example, consider the two sets of mixtures *A1(a1,a2,a3,a4)*, *B1(b1,b2,b3,b4)*, *A2(a1,b1,c1,d1)* and *B2(a2,b2,c2,d2)* where each compound as required is grouped into two unique sets of compounds. If all four mixtures are identified as containing an active compound, it will not be possible to unambiguously assign the active compounds without deconvolution of at least one of the mixtures. This problem occurs because mixtures *A1* and *B1* have a compound present in both mixtures *A2* and *B2*, but only two of the four overlapping compounds (*a1,b1,a2,b2*) present in the two sets of mixtures are active.

Assuming the likelihood of serendipitous overlap and multiple hits per mixture is essentially zero for low hit rates, the effective efficiency of strategic pooling is $\sim(2N/n)$. Based on our analysis of OMS, strategic pooling only becomes more

efficient than standard deconvolution methods when:

$$\frac{1}{n} < P \qquad (6)$$

The probability (P) of observing a hit in mixture (n) will scale with increasing mixture size and hit rate, implying that the relative efficiency of strategic pooling improves with larger mixture sizes and larger hit rates. This is apparent in Figures 1a and 1f where the curve for strategic pooling is plotted together with the standard deconvolution curves. For a given mixture size at low hit rates ($\leq 0.5\%$), standard deconvolution is always more efficient than strategic pooling. This is also consistent with the previous observation that standard deconvolution closely follows the non-deconvolation situation (N/n) at low hit rates. Interestingly, at higher hit rates the efficiency of strategic pooling is effectively equivalent to the standard deconvolution approach. This is apparent in figures 1c-f where the strategic pooling curve nearly passes through the OMS points in the standard deconvolution curves. At all hit rates, strategic pooling only becomes more efficient than standard deconvolution when mixture sizes greater than the predicted OMS are used (Figure 1). But at higher hit rates and larger mixture sizes, the strategic pooling approach fails because the probability of finding multiple active compounds in a single mixture and the occurrence of pairs of mixtures with overlapping compounds is non-zero. This is clearly evident in Figure 2 where the probability of a mixture containing more than one active compound becomes the common event at higher hit rates. As a result, the strategic pooling approach would not be an advantage for NMR-based screens that are being used as a secondary assay to confirm active compounds from an HTS screen or utilizing a SHAPES-like library. Strategic pooling is also a limited utility in an HTS screen.

There are also some practical considerations in the routine application of a strategic pooling approach for larger mixture sizes (Glick et al., 2003). Inherent to the methodology is the requirement of designing ($2N/n$) mixtures from a library of N compounds where each mixture has a unique combination of compounds. The mixture design also needs to consider issues such as compound reactivity, solubility and diversity (Schriemer and Hindsgaul, 1998; Hann et al.,

1999; Brown and Martin, 1997; Brown et al., 2000). As mixture size and library size increase, finding two sets of mixtures that adheres to these requirements will increase dramatically in complexity where a practical solution may be intractable (Brown and Martin, 1997; Brown et al., 2000).

Bucket-sort approach to deconvolution

One mechanism to reduce the impact of the deconvolution step is to use a tiered approach to screening. A large mixture size (>20) is initially screened, where mixtures with an identified hit are sub-divided into a second set of smaller mixtures (≤ 10) for further screening. Compounds from the smaller mixtures with an identified hit are then screened as singletons. Effectively the bucket-sort method results in a two-step deconvolution process (Equation. 4). Figure 4 illustrates the dramatic decrease in the total number of experiments required for a mixture size of 100 compounds if the deconvolution step is first screened as mixtures of 5 compounds instead of singletons. The bucket-sort approach clearly results in a significant improvement in efficiency ($\sim$4–5) compared to standard deconvolutions and $\sim$2 fold improvement compared to the OMS.

Determining the optimal mixture size for the second set of mixtures is conceptionally similar to determining OMS for the simple deconvolution process (Figure 5). A major difference is the fact that the probability of identifying a hit in the mixture is not determined by the overall hit rate for the compound library (HR_1). The bucket-sort approach results in an increase in the effective hit rate (HR_2) for the second set of mixtures since the first screen simply removes a set of inactive compounds while moving all the active compounds forward in the assay. The impact is a relatively narrow OMS window of 6–12 compounds that is only meaningful for hit rates $\leq 5\%$. For higher hit rates, the OMS reduces to a range of 1–3 compounds per mixture. Also, the impact on the total efficiency of the NMR screen is relatively flat for low hit rates. For a hit rate of 1%, there is a minimal difference between mixture sizes 5 through 25. In essence, the impact of the bucket-sort approach is not strongly influenced by the OMS for the second set of mixtures at low hit rates. It appears the majority of the gain in

efficiency simply arises from applying two sets of mixtures to the deconvolution process. Figure 4 clearly illustrates this point, a mixture size of 5 and 10 compounds is used for the second mixture. At hit rates ≤5%, there are only minor differences in the efficiency of the bucket-sort approach using the two different mixtures sizes compared to the significant improvement over both standard deconvolution and OMS. But, as the hit rate rises, using a less than optimized second mixture size results in a dramatic decrease in the efficiency of the bucket-sort approach where the method quickly becomes less efficient than a standard deconvolution approach. Since the OMS reduces to 1–3 compounds per mixture at higher hit rates, the bucket sort approach effectively reduces to a standard deconvolution method.

The application of the bucket-sort approach to deconvolution suggests that it may be feasible for an NMR-based assay to screen a 200,000 compound library in ∼2–4 weeks assuming a random library with a low hit rate (≤0.5%) and typical NMR acquisition times. While the throughput still positions the efficiency of an NMR-based screen at the lower end of a traditional HTS assay, the added value of a confirmed protein-ligand binding interaction with a diminished number of false positives more than compensates for the lower throughput. In fact, current trends indicate that the preferred mode of executing an HTS assay is to screen smaller libraries that have been designed to contain compounds with improved drug-like characteristics and maximum structural diversity (Viswanadhan et al., 2002; Matter et al., 2001; Xu and Stevenson, 2000; Kubinyi, 2003; Jelic et al., 2003; Xue and Bajorath, 2000; Lewis et al., 2000; Willett, 2000; Spellmeyer and Grootenhuis, 1999; Gorse and Lahana, 2000). In principal, these smaller chemical libraries would be more amenable to an NMR-based assay using a bucket-sort approach.

While the bucket-sort approach represents a significant improvement in efficiency for NMR-based screens, it also adds layers of complexity requiring multiple interrelated compound libraries comprising different mixture sizes and compound combinations with corresponding compatibility issues. The bucket-sort method for large mixture sizes (≳50) yields the best efficiency results in regards to the total number of required experiments, and thus should be favored for screening large libraries, but may represent serious practical challenges in identifying appropriate mixtures. Clearly, identifying a few thousand mixtures composed of 50–100 compounds that are: (i) non-reactive, (ii) stable, (iii) don't aggregate or form micelle-like structures, (iv) maintain a reasonable solubility detectable by NMR, (v) are structurally diverse and (vi) don't negatively impact the stability of proteins is an extremely challenging endeavor. Most of these factors are not reliably or readily predictable from the simple knowledge of the compounds structure (Cheng and Merz, 2003; Chen et al., 2002; Taskinen and Yliruusi, 2003; Hann et al., 1999). Additionally, experimental data is very limited especially given the large number of novel compounds that comprise most corporate chemical libraries (Klan and Jindrich, 2000). In practice, confirming the compatibility of mixture combinations for small libraries is typically obtained by trial and error. The alternative is to assemble best estimates of appropriate compound mixtures and simply ignore or discard problematic or 'failed' mixtures. The expectation is that the number of failed or problematic mixtures would represent a small percentage of the entire library. This is a reasonable assumption when assembling small mixtures composed of 5–10 compounds, but as the mixture size is increased by an order of magnitude, problem mixtures may become more prevalent. Thus, the inherent difficulty in assembling an appropriate chemical library using large mixtures is the limiting factor in the application of a bucket-sort NMR screening approach.

Conclusion

A model based on the application of the hypergeometric distribution function for predicting the optimal mixture size for NMR-based screens was presented. The model indicates that OMS is directly dependent on the hit rate, where OMS decreases as the hit rate is increased. An OMS in the range of 15–20 compounds per mixture is predicted for typical random chemical libraries with hit rates of 0.35–0.2%. NMR-based assays are increasingly being used to screen small directed or focused compound libraries or as secondary assays to confirm HTS chemical leads that exhibit significantly higher hit-rates (>5%). For these

expected higher hit rates where the absolute hit rate is not known *a priori*, a mixture size of 5 compounds for an NMR screen is a reasonable target. But, our analysis also indicates that screening small directed libraries that require deconvolution of mixtures is extremely inefficient relative to screening smaller mixture sizes that avoid deconvolution. Thus, the efficiency of the NMR screen will be improved significantly by using mixture sizes composed of 3–4 compounds that do not require deconvolution compared to a mixture size of 5 compounds that requires deconvolution. Also, smaller mixture sizes tend to be more practically achievable when factors such as solubility, structural diversity, reactivity and spectral overlap are considered.

The opposite condition exists at low hit rates (≤0.5%), where there are numerous situations where an improvement in efficiency can be achieved by using larger mixtures sizes that require deconvolution. A bucket-sort approach that minimizes the impact of deconvolution using a tiered approach results in a dramatic improvement in efficiency relative to both OMS and single deconvolution step, but incurs the practical challenge of needing to assemble a large chemical library comprising mixtures of 50–100 compounds that are valuable to a screening effort. While this analysis of OMS was done in the context of NMR screening, the results presented herein are universally applicable to all types of screening methods including mass spectrometry and biological assays.

Acknowledgements

References

Appel, J.R. Johnson, J. Narayanan, V.L. and Houghten, R.A. (1999) *Mol. Divers.* **4**91–102.

Armstrong, J.W. (1999) *Am. Biotechnol. Lab.* **17**26–28.

Baurin, N. Aboul-Ela, F. Barril, X. Davis, B. Drysdale, M. Dymock, B. Finch, H. Fromont, C. Richardson, C. Simmonite, H. and Hubbard, R.E. (2004) *J. Chem. Inform. Comput. Sci.* **44**2157–2166.

Brown, R.D. Hassan, M. and Waldman, M. (2000) *J. Mol. Graph. Model.* **18**427–437.

Brown, R.D. and Martin, Y.C. (1997) *J. Med. Chem.* **40**2304–2313.

Chen, A. and Shapiro, M.J. (1999) *Anal. Chem.* **71**669A–675A.

Chen, X.-Q. Cho, S.J. Li, Y. and Venkatesh, S. (2002) *J. Pharmaceut. Sci.* **91**1838–1852.

Cheng, A. and Merz, K.M. Jr. (2003) *J. Med. Chem.* **46**3572–3580.

Clore, G.M. and Gronenborn, A.M. (1994) *Protein Sci.* **3**372–390.

Cooke, R.M. (1997) *Curr. Opin. Chem. Biol.* **1**359–364.

Dalvit, C. Hadden, D.T.A. Sarver, R.W. Ho, A.M. and Stockman, B.J. (2003) *Combin. Chem. High Throughput Screen.* **6**445–453.

Devlin, J.J. Liang, A. Trinh, L. Polokoff, M.A. Senator, D. Zheng, W. Kondracki, J. Kretschmer, P.J. and Morser, J. (1996) *Drug Develop. Res.* **37**, 80–5.

Dove, A. and Marshall, A. (1999) *Nat. Biotechnol.* **17**859–863.

Fejzo, J. Lepre, C.A. Peng, J.W. Bemis, G.W. Murcko, M.A. and Moore, J.M. (1999) *Chem. Biol.* **6**755–769.

Feller, W. (1968) In *An Introduction to Probability and its Applications*, Vol. 1 Wiley, New York, pp. 41–45.

Fernandes, P.B. (1998) *Curr. Opin. Chem. Biol.*, **2**, 597–603.

Fielding, L. (2003) *Curr. Top. Med. Chem. (HilversumNetherlands)* **3**, 39–53.

Glick, M. Klon, A.E. Acklin, P. and Davies, J.W. (2003) *Mol. Phys.* **101**1325–1328.

Glick, M. Klon Anthony, E. Acklin, P. and Davies John, W. (2004) *J. Biomol. Screen.: official journal of the Society for Biomolecular Screening* **9**32–6.

Gonzalez, J.E. and Negulescu, P.A. (1998) *Curr. Opin. Biotechnol.* **9**624–631.

Gorse, D. and Lahana, R. (2000) *Curr. Opin. Chem. Biol.* **4**287–294.

Hajduk, P.J. and Burns, D.J. (2002) *Combin. Chem. High Throughput Screen.* **5**613–621.

Hajduk, P.J. Dinges, J. Miknis, G.F. Merlock, M. Middleton, T. Kempf, D.J. Egan, D.A. Walter, K.A. Robins, T.S. Shuker, S.B. Holzman, T.F. and Fesik, S.W. (1997a) *J. Med. Chem.* **40**3144–3150.

Hajduk, P.J. Gerfin, T. Boehlen, J.-M. Haeberli, M. Marek, D. and Fesik, S.W. (1999a) *J. Med. Chem.* **42**2315–2317.

Hajduk, P.J. Gerfin, T. Boehlen, J.-M. Haeberli, M. Marek, D. and Fesik, S.W. (1999b) *J. Med. Chem.* **42**2315–2317.

Hajduk, P.J. Olejniczak, E.T. and Fesik, S.W. (1997b) *J. Am. Chem. Soc.* **119**12257–12261.

Hajduk, P.J. Sheppard, G. Nettesheim, D.G. Olejniczak, E.T. Shuker, S.B. Meadows, R.P. Steinman, D.H. Carrera, G.M. Jr. Marcotte, P.A. Severin, J. Walter, K. Smith, H. Gubbins, E. Simmer, R. Holzman, T.F. Morgan, D.W. Davidsen, S.K. Summers, J.B. and Fesik, S.W. (1997c) *J. Am. Chem. Soc.* **119**5818–5827.

Hann, M. Hudson, B. Lewell, X. Lifely, R. Miller, L. and Ramsden, N. (1999) *J. Chem. Inform. Comput. Sci.* **39**897–902.

Huth, J.R. and Sun, C. (2002) *Combin. Chem. High Throughput Screen.* **5**631–643.

Jacoby, E. Davies, J. and Blommers, M.J.J. (2003) *Curr. Top. Med. Chem. (HilversumNetherlands)* **3**, 11–23.

Jahnke, W. and Widmer, H. (2004) *Cell. Mol. Life Sci.* **61**580–599.

Jelic, D. Mesar, V. Basic, I. Nadramija, D. and Verbanac, D. (2003) *PharmaChem* **2**64–67.

Johnson, E.C. Feher, V.A. Peng, J.W. Moore, J.M. and Williamson, J.R. (2003) *J. Amer. Chem. Soc.* **125**15724–15725.

Kay, L.E. (1997) *Biochem. Cell Biol.* **75**1–15.

Kenny, B.A. Bushfield, M. Parry-Smith, D.J. Fogarty, S. and Treherne, J.M. (1998) *Prog. Drug Res* **51**245–269.

Klan, P. and Jindrich, J. (2000) *CHEMagazin* **10**21–22.

Klaus, W. and Senn, H. (2003) *Meth. Principles Med. Chem.* **16**417–437.

Kubinyi, H. (2003) *Nat. Rev. Drug Discov.* **2**665–668.

Lepre, C.A. Peng, J. Fejzo, J. Abdul-Manan, N. Pocas, J. Jacobs, M. Xie, X. and Moore, J.M. (2002) *Combin. Chem. High Throughput Screen.* **5**583–590.

Lewis, R.A. Pickett, S.D. and Clark, D.E. (2000) *Rev. Comput. Chem.* **16**1–51.

Lin, M. Shapiro, M.J. and Wareing, J.R. (1997) *J. Org. Chem.* **62**8930–8931.

Matter, H. Baringhaus, K.-H. Naumann, T. Klabunde, T. and Pirard, B. (2001) *Combin. Chem. High Throughput Screen.* **4**453–475.

McGovern, S.L. Caselli, E. Grigorieff, N. and Shoichet, B.K. (2002) *J. Med. Chem.* **45**1712–1722.

McGovern, S.L. Helfand, B.T. Feng, B. and Shoichet, B.K. (2003) *J. Med. Chem.* **46**4265–4272.

Meyer, B. Weimar, T. and Peters, T. (1997) *Euro. J. Biochem.* **246**705–709.

Moore, J. Abdul-Manan, N. Fejzo, J. Jacobs, M. Lepre, C. Peng, J. and Xie, X. (2004) *J. Synchrot. Radiat.* **11**97–100.

Moore, J.M. (1999a) *Curr. Opin. Biotechnol.* **10**54–58.

Moore, J.M. (1999b) *Biopolymers* **51**221–243.

Moy, F.J. Haraki, K. Mobilio, D. Walker, G. Powers, R. Tabei, K. Tong, H. and Siegel, M.M. (2001) *Anal. Chem.* **73**571–581.

Oldenburg, K.R. (1998) *Annu. Rep. Med. Chem.* **33**301–311.

Peng, J.W. Lepre, C.A. Fejzo, J. Abdul-Manan, N. and Moore, J.M. (2001) *Meth. Enzymol.* **338**202–230.

Pratt Steven, D. David Caroline, A. Black-Schaefer, C. Dandliker Peter, J. Xuei, X. Warrior, U. Burns David, J. Zhong, P. Cao, Z. Saiki Anne, Y.C. Lerner Claude, G. Chovan Linda, E. Soni Niru, B. Nilius Angela, M. Wagenaar Frank, L. Merta Philip, J. Traphagen Linda, M. and Beutel Bruce, A. (2004) *J. Biomol. Screen.: Official J. Soc. Biomol. Screen.* **9**3–11.

Rishton, G.M. (1997) *Drug Discov. Today* **2**382–384.

Roberts, G.C.K. (2000) *Drug Discov. Today* **5**230–240.

Rossi, C. Donati, A. and Sansoni, M.R. (1992) *Chem. Phys. Lett.* **189**278–80.

Schriemer, D.C. and Hindsgaul, O. (1998) *Combin. Chem. High Throughput Screen.* **1**155–70.

Seidler, J. McGovern, S.L. Doman, T.N. and Shoichet, B.K. (2003) *J. Med. Chem.* **46**4477–4486.

Sem, D.S. and Pellecchia, M. (2001) *Curr. Opin. Drug Discov Develop.* **4**479–492.

Shuker, S.B. Hajduk, P.J. Meadows, R.P. and Fesik, S.W. (1996) *Science* **274**1531–1534.

Silverman, L. Campbell, R. and Broach, J.R. (1998) *Curr. Opin. Chem. Biol.* **2**397–403.

Smith, A. (2002) Nature (London, United Kingdom), 418, 451, 453, 455, 457, 459, 461, 463..

Spellmeyer, D.C. and Grootenhuis, P.D.J. (1999) *Annu. Rep. Med. Chem* **34**287–296.

Spiegel, M.R. (1992) Theory and Problems of Probability and Statistics, McGraw-Hill, New York.

Taskinen, J. and Yliruusi, J. (2003) *Adv. Drug Deliver. Rev.* **55**1163–1183.

Teixido, J. Michelotti, E.L. and Tice, C.M. (2000) *J. Combin. Chem.* **2**658–674.

Villar, H.O. Yan, J. and Hansen, M.R. (2004) *Curr. Opin. Chem. Biol.* **8**387–391.

Viswanadhan, V.N. Balan, C. Hulme, C. Cheetham, J.C. and Sun, Y. (2002) *Curr. Opin. Drug Discov Develop.* **5**400–406.

Wallace, R.W. (1998) *Drug Discov. Today* **3**192 .

Willett, P. (2000) *Curr. Opin. Biotechnol.* **11**85–88.

Xu, J. and Stevenson, J. (2000) *J. Chem. Inf. Comput. Sci.* **40**1177–1187.

Xue, L. and Bajorath, J. (2000) *Comb. Chem. High Throughput Screen.* **3**363–372.