JOURNAL OF combinatorial CHEMISTRY

Subscriber access provided by American Chemical Society

Article

Ruminations Regarding the Design of Small Mixtures for Biological Testing

Jordi Teixido, Enrique L. Michelotti, and Colin M. Tice

J. Comb. Chem., **2000**, 2 (6), 658-674• DOI: 10.1021/cc0000390 • Publication Date (Web): 26 October 2000

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

$$\text{Screening Efficiency} = \frac{1}{\frac{1}{N} + \quad N * HR}$$

More About This Article

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

- Supporting Information
- Links to the 1 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

ACS Publications
High quality. High impact.

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

Ruminations Regarding the Design of Small Mixtures for Biological Testing

Jordi Teixido,*,† Enrique L. Michelotti,*,‡ and Colin M. Tice‡

Institut Quimic de Sarria, Ramon Llull University, Via Augusta 390, Barcelona 08017, Spain, and Rohm and Haas Company, 727 Norristown Road, Spring House, Pennsylvania 19477-0904

Received May 12, 2000

Synthesis and screening of compound mixtures offer avenues to increase throughput and reduce cycle time in the discovery of new drugs and agrochemicals. Equations are derived which show that the efficiency of synthesis and screening of mixtures is a function of the screening hit rate and the number of compounds in each mixture when simple one-step deconvolution by retesting the individual compounds in each active mixture is employed. Values of hit rate and number of compounds in each mixture which afford various levels of increased efficiency are delineated. Two-step deconvolution, in which the active mixtures from the first round of testing are subdivided into mixtures with fewer compounds for a second round of mixture screening prior to final testing of individual compounds, is shown to be more efficient than simple one-step deconvolution under most conditions. For optimum efficiency, the number of compounds in each mixture in the second round testing should be the square root of the number of compounds in each mixture in the first round. At high hit rates the efficiency of the double scan or indexed approach to deconvolution is shown to be higher than that of simple deconvolution. This discussion is oriented mainly toward mixtures of 4−20 compounds and screens which give hit rates in the 1−10% range. The equations describing efficiency are applied in the context of a 49-member amide library produced as mixtures of seven compounds. This library includes the commercial herbicide pronamide and was screened for herbicidal and insecticidal utility.

Introduction

It is generally accepted in the pharmaceutical and agrochemical arena that 20,000 or more compounds must be screened to find one commercial product. This has led to the ongoing development of techniques to rapidly synthesize and screen large numbers of compounds. In particular, in recent years advances in the design and synthesis of combinatorial libraries have led to the preparation and screening of large numbers of new compounds either as single entities or in mixtures.[1−12] Synthesis and/or screening of mixtures of compounds is one approach to increase throughput (number of compounds screened or synthesized per unit of time) and reduce the time to discover a new product.

Judicious design of mixtures of compounds can increase substantially the throughput with respect to the "one at a time" method.[13−16] When mixtures of compounds are synthesized and screened, a deconvolution step is necessary to determine which molecule(s) is(are) responsible for the biological activity.[17−22] Mixtures of compounds may be produced by combining a number of pure compounds prepared individually or synthesized as mixtures. Screening of mixtures has met with the greatest acceptance in new lead discovery where a low hit rate is anticipated, while screening single pure compounds remains the preferred mode for lead optimization.

When designing libraries consisting of mixtures of compounds, several issues must be confronted. First, the synthetic strategy should be robust enough to provide roughly equal molar concentration of each component in each mixture;[23−35] second, analytical methods should allow the identification and quantification of each component,[36−45] and finally, the biological test must be amenable to the evaluation of mixtures.[46−48] Also, consideration should be given to the possible biological interactions.[49]

In this paper we analyze in detail the pros and cons of synthesis and screening of mixtures and delineate conditions under which working with mixtures is likely to improve throughput using or slightly adapting the existing "one at a time" screening procedures. The first question to be addressed is: is it possible and advantageous to use standard, one compound at a time, screening procedures for testing mixtures? The advantage or maximum efficacy of this approach will arise from a compromise between the number of compounds present in the mixture and the expected hit rate (percentage of active compounds present in the library). We will analyze the screening process and its effectiveness in the testing of mixtures (section I). We will continue with considerations and benefits on the synthesis of discrete mixtures (sections II and III). Finally, we will also present the results obtained with a small library screened for agrochemical utility in the context of the theoretical analysis (section IV).

† Ramon Llull University.
‡ Rohm and Haas Company.

10.1021/cc0000390 CCC: $19.00 © 2000 American Chemical Society
Published on Web 10/26/2000

Discussion

I. Biological Screening. I.1. Biological Screening Assay: Quantity of Test Materials and Detection Limit. The sensitivity and detection limit threshold of a given biological assay will influence the nature of the mixture, in terms of both the amount of each compound in the mixture and the number of compounds per mixture, that can successfully and reliably be tested. The pharmaceutical industry has developed highly specific and very sensitive assays that can successfully detect/evaluate the activity of thousands or even millions of compounds simultaneously at nanomolar concentrations.[3] In this case the slow process is to identify the active compound(s) present in the mixture. In the agrochemical industry, this type of assay is less common and the majority of the assays utilized are in vivo whole organism biological tests designed originally for the evaluation of one compound at a time. Agrochemical assays also use larger quantities of material than can be produced on a single bead, and for this reason the split/mix with encoding technique is not applicable. In this discussion, we are especially interested in the latter, posing the question of whether these in vivo whole organism assays can be utilized for the biological characterization of discrete mixtures composed of 20 or fewer compounds. Special attention must be given to the total amount of test material applied to a given biological test, because the solubility of the mixture will be a function of the solubility of each individual compound and the number of components. Also, the specific limitations of each test will determine the upper limit of amount of test material that can be applied. The total number of compounds evaluated simultaneously in a mixture will also depend on the detection limit of the biological test in consideration. In many cases, compounds are evaluated at several different doses. Let's consider a generic biological test run at three or four different doses, e.g., 15, 60, 250, and 1000 ppm where 1000 ppm is the largest quantity of applied chemical tolerated by the test. If a four-compound mixture containing one compound active at 250 ppm is applied at 1000 ppm, the assay should give a response similar to that obtained by testing the active component alone at 250 ppm. Similarly, a 16-compound mixture containing one compound active at 60 ppm should give a response similar to that obtained by testing the active component alone at 60 ppm. If this is the case, biological evaluation of mixtures containing from 1 up to 16 compounds can be performed without modification of the existing test protocol, providing that missing a compound in a mixture that is only active at higher doses can be tolerated.

I.2. False Positives/False Negatives. The occurrence of false positives and false negatives is inherent to any biological test. When mixtures are being evaluated, the effects of false positives and false negatives are compounded. The biological activity of a single compound as a function of the concentration (C_m) can be written as a power series expansion:

$$\text{biological activity} = f(C_m) = k_o + k_m C_m + k_m^{(2)} C_m^{\;2} + k_m^{(3)} C_m^{\;3} + \ldots$$

For a mixture of n compounds, this expression can be written as a function of the concentration (C_1, C_2, C_3, ...C_n) of each compound:

$$\text{biological activity (mixture)} = \mathrm{f}(\mathrm{C}_1, \mathrm{C}_2, \mathrm{C}_3, \ldots \mathrm{C}_n) =$$
$$= \boldsymbol{k_o} + \boldsymbol{k_1 C_1} + \boldsymbol{k_2 C_2} + \boldsymbol{k_3 C_3} + \ldots \boldsymbol{k_m C_m} \ldots + k_{12} C_1 C_2 + k_{13} C_1 C_3 + k_{23} C_2 C_3 + \ldots k_{mn} C_m C_n \ldots + k_{123} C_1 C_2 C_3 + \ldots +$$
$$\boldsymbol{k_1^{(2)} C_1^{\;2}} + \boldsymbol{k_2^{(2)} C_2^{\;2}} + \boldsymbol{\ldots} \boldsymbol{k_m^{(2)} C_m^{\;2}} \boldsymbol{\ldots} + \boldsymbol{k_m^{(3)} C_m^{\;3}} \boldsymbol{\ldots} + \boldsymbol{k_m^{(p)} C_m^{\;p}} \boldsymbol{\ldots} + \ldots k_{pqr}^{(i+j+k)} C_p^{\;i} C_q^{\;j} C_r^{\;k} \ldots$$

The design of the mixture should consider any chemical, physicochemical, and particularly biological interactions between compounds. Synergism or additivity can lead to false positives and antagonism, depending on the assay, to false negatives. If the test procedures and the mixture design can minimize or eliminate the cross terms ($\boldsymbol{k_{mn\ldots}}^{(..)} \to 0$) in the above equation, then we can envision that the new factor introduced by the mixture approach is solely additive (**bold** terms). A false positive will lead to a waste of resources forcing a retest of the individual components of the mixture, which may require resynthesis and/or separation (deconvolution).

In this study we will consider that any chemical and physicochemical incompatibilities are minimized, or totally eliminated, by the design of the mixtures and by physical manipulation of the test sample. It is anticipated that chemically similar compounds will have also similar biological activities leading to additivity problems.[1]

Designing mixtures containing similar compounds will increase the global efficiency because active compounds are more likely to occur in a few mixtures and thus fewer mixtures will need to be deconvoluted. However, in small libraries it will also increase the number of false positives due to additivity of several weakly active compounds. Conversely, designing mixtures containing very dissimilar compounds will lead to active compounds occurring in a larger number of different mixtures. This will reduce the number of false positives but will also increase the number of testing operations needed to identify the active compound(s) present in the library, because more mixtures will need to be deconvoluted. Each case should be evaluated carefully to decide which is the most efficient approach.

In this analysis we will always assume the worst case for the number of testing operations needed to identify the active compound(s), that is: *Design the most dissimilar mixtures expecting similar active compounds to occur in different mixtures, leading to the largest number of mixtures to deconvolute*.

Our analysis will show that even in this case a considerable decrease in the number of testing operations needed to identify the active compound(s) is expected when the hit rate is below 10%.

I.3. Influence of the Number of Compounds in a Mixture (*N*) and the Hit Rate (HR) on the Screening Efficiency. Before discussing the general case, we will begin with a hypothetical example. Let's assume that we want to analyze a library of 1000 compounds in 250 four-compound mixtures ($M = 250$, $N = 4$). As stated above, the worst case for the number of operations needed to identify the active

Table 1. Screening Efficiency as a Function of the Hit Rate (HR)[a]

HR	no. of active compds[b]	no. of active mixtures[c]	deconvolution[d]	total no. of operations[e]	screening efficiency[f]
0.001	1	1	4	250 + 4 = 254	1000/254 = 3.94
0.01	10	10	40	290	3.45
0.025	25	25	100	350	2.86
0.05	50	50	200	450	2.22
0.075	75	75	300	550	1.82
0.10[g]	100	100	400	650	1.54
0.125[h]	125	125	500	750	1.33
0.15[h]	150	150	600	850	1.18
0.175[h]	175	175	700	950	1.05
0.020[i]	200	200	800	1050	0.95
0.25	250	250	1000	1250	0.80
>0.25	>250	250	1000	1250	0.80

[a] Hit rate (HR); defined as the fraction of active compounds in the library. [b] One thousand-compound library divided in 250 mixtures (*M*) of 4 compounds each (*N*). [c] Assuming 1 active compound per mixture. [d] Number of individual compounds to be retested (number of active mixtures multiplied by 4). [e] Number of operations to test the *M* initial mixtures plus the number of individual compounds to be retested in the deconvolution step. In this case 250 + (active mixtures) × 4. [f] Defined as in eq 1. [g] Hit rates lower than 0.1 show screening efficiencies > 1.5. [h] Borderline improvement in the screening efficiency. [i] Not an improvement in the screening efficiency.

compounds is when each active compound is present in a different mixture. To determine which are the biologically active compounds without considering replicates we need to run

$$1000/N = 1000/4 = 250 \text{ mixtures} \rightarrow 250 \text{ initial tests}$$

If the hit rate, HR, defined as the fraction of compounds in the library active at the concentration at which they are tested in the mixture, is 0.001, only 1 compound in the entire set in consideration is active and only one mixture will be active. We will then need to individually retest (deconvolute) the four components of the active mixture in order to identify the active compound. In this case, only 254 tests (250 initial tests + 4 retests) are necessary to find the active compound, instead of 1000 tests for the traditional "one compound at a time" approach. This corresponds to a 3.94× decrease in the number of tests that need to be run or, conversely, a concomitant 3.94× increase in the efficiency of biological screening.

Similarly, if HR = 0.01, there are 10 active compounds in the library. Thus, a maximum of 10 mixtures can show biological activity. We would then need to individually retest the 40 compounds present in the 10 biologically active mixtures. Therefore, a total of 290 tests will identify the 10 active compounds present in the library, which corresponds to a 3.45× improvement in the rate of biological evaluation, which is to say, a better use of screening resources. The analysis for different values of HR for this library is presented in Table 1. From Table 1, we can easily see that if HR ≥ 0.25, 25% of the compounds in the library are active, there are then more than 250 active compounds, and in the worst case, all 250 mixtures will have at least one active compound and more tests will be required to analyze the library than if the compounds were tested one at a time. Also, Table 1 shows that for HR below 0.1, 10% of the compounds in the library are active, and the increase in the screening efficiency is >1.5. When HR is between 0.175 and 0.20, the same number of tests are needed to evaluate the library using either the "one compound at a time" or the mixture approach and there is little increase or decrease in the testing throughput.

To analyze a general case, let us consider a library of *M* mixtures of *N* compounds each. The total number of compounds in the library is the product $M{*}N$. Obviously, the number of tests needed to screen the library in the "one compound at a time" approach is $M{*}N$. Using the mixture approach, initially we only need to run *M* tests and later retest individually (deconvolute) only the compounds present in the active mixtures.

If HR is the hit rate, previously defined as the fraction of active compounds in the library, then, $M{*}N{*}\mathrm{HR}$ is the number of active compounds present in the library. The worst case for the rate of biological evaluation, the screening throughput, is when each active compound is present in a different mixture and the product $M{*}N{*}\mathrm{HR}$ is also the maximum number of mixtures that show activity and must be deconvoluted. The most straightforward deconvolution method is to simply individually retest each compound in the mixture. Applying this protocol requires $(M{*}N{*}\mathrm{HR}){*}N = M{*}N^2{*}\mathrm{HR}$ additional tests since there are *N* compounds in each mixture. Then $M + (M{*}N^2{*}\mathrm{HR})$ is the total number of tests needed to screen a library of *M* mixtures of *N* compounds per mixture in the mixture approach.

In this analysis we define the screening efficiency as the ratio of the number of tests needed to analyze the library by the traditional "one compound at a time" approach, divided by the number of tests needed to screen the library by the *N*-compound mixture approach. Thus,

$$\text{screening efficiency} = \frac{M{*}N}{M + M{*}N^2{*}\mathrm{HR}}$$

Simplifying the formula and normalizing the library, we get the general expression

$$\text{screening efficiency} = \frac{M{*}N}{M + M{*}N^2{*}\mathrm{HR}} = \frac{M{*}N}{M{*}N\left[\frac{1}{N} + N{*}\mathrm{HR}\right]}$$

$$\text{screening efficiency} = \frac{1}{\frac{1}{N} + N{*}\mathrm{HR}} \qquad (1)$$

Equation 1 shows that the screening efficiency is only a function of *N* (the number of compounds in each mixture)

and HR (the hit rate), but it is independent of the number of mixtures M. We can analyze some special cases of HR:

(a) When HR $\geq 1/N$ then N*HR ≥ 1 and the denominator of eq 1 is always > 1 resulting in a screening efficiency that is always < 1. This corresponds to the situation in which all the mixtures have at least one active compound.

$$\text{screening efficiency} = \frac{1}{\frac{1}{N} + 1} = \frac{N}{N+1} < 1$$

(b) When HR $< 1/N$, some of the mixtures show no biological activity, then N*HR < 1 resulting in a screening efficiency that is greater than case (a) and might also be $>$ 1. This is illustrated in Table 1, in the case when $N = 4$ then $1/N = 0.25$. The actual break even point for the screening efficiency falls between 0.175 and 0.20 and not exactly at 0.25. When the screening efficiency is 1, the same number of biological tests is needed to evaluate the library by either the "one compound at a time" method or the mixture approach. Then from eq 1:

$$1 = \frac{1}{\frac{1}{N} + N^*\text{HR}}$$

which can be rearranged to

$$\text{For screening efficiency} = 1 \qquad \text{HR} = \frac{N-1}{N^2} \qquad (2)$$

For the hypothetical example of $N = 4$ we find that the turning point at which the screening efficiency = 1 occurs at HR = 0.1875, which is in agreement with the value between 0.175 and 0.20 indicated in Table 1. Similarly, eq 1 can be solved for any desired level of screening efficiency, in the most general case:

$$\text{HR} = \frac{N - (\text{screening efficiency})}{(\text{screening efficiency}) * N^2}$$

Resolving for specific cases, e.g.

$$\text{For screening efficiency} = 2 \qquad \text{HR} = \frac{N-1}{2N^2} \qquad (3)$$

$$\text{For screening efficiency} = 4 \qquad \text{HR} = \frac{N-4}{4N^2} \qquad (4)$$

The values of HR given by eqs 2, 3, and 4 for various values of N are tabulated in Table 2. Thus if a screening efficiency of 4 is required to justify use of mixtures, the maximum hit rate that can be tolerated is 0.0156 which occurs when $N = 8$. Furthermore, if any active compounds are expected, then N must be ≥ 5 if a screening efficiency of ≥ 4 is to be attained.

From eq 1 it is possible to study analytically (results shown in Table 3) and graphically (results presented in Figures 1 and 2) changes of the screening efficiency for different values of N and HR. It is important to note that when HR = 0.03−0.1, the maximum screening efficiency falls somewhere between $N = 4-6$, (see Figure 1). With increasing N, the curve sharpens and the turning point for the screening

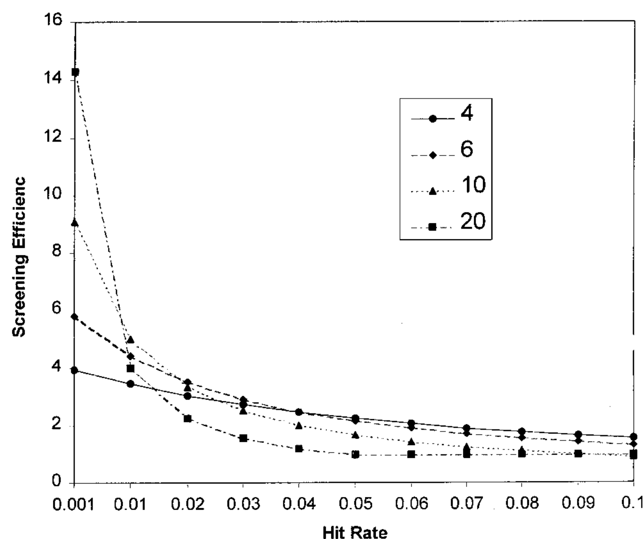


Figure 1. Screening efficiency as a function of the HR and N.

Table 2. Maximum Hit Rate (HR) as a Function of N To Achieve Various Levels of Screening Efficiency (Eqs 2, 3, and 4)

	allowable hit rates		
N	screening efficiency ≥ 1	screening efficiency ≥ 2	screening efficiency ≥ 4
2	≤0.2500	≤0.0000	
3	≤0.2222	≤0.0556	
4	≤0.1875	≤0.0625	≤0.0000
5	≤0.1600	≤0.0600	≤0.0100
6	≤0.1389	≤0.0556	≤0.0139
8	≤0.1094	≤0.0469	≤0.0156
10	≤0.0900	≤0.0400	≤0.0150

Table 3. Screening Efficiency as Function of N and HR (Eq 1)

	screening efficiency			
N	HR 0.001	HR 0.01	HR 0.05	HR 0.1
2	1.99	1.92	1.67	1.43
3	2.97	2.75	2.07	1.58
4	3.93	3.45	2.22	1.54
5	4.88	4.00	2.22	1.43
6	5.79	4.41	2.14	1.30
8	7.52	4.88	1.90	1.08
10	9.09	5.00	1.67	0.91
15	12.24	4.61	1.22	0.94
20	14.29	4.00	0.95	0.95

efficiency approaches the origin (Figures 1 and 2). From Figure 1 we can see that the maximum screening efficiency is equal to N, (HR = 0). Analyzing Figures 1 and 2 graphically for a given N and the maximum expected HR, the average screening efficiency is related to the area under the corresponding curve (application window). The greater the application window, the greater the screening efficiency. Accordingly, it is possible to increase the screening efficiency by using two curves at the same time, and this case will be treated in more detail later (see section I.4., eqs 7 and 8).

In eq 1 and Figures 1 and 2 we have considered a library with an infinite number of points although, in fact, we have discrete integers. Moreover, the HR of the library in a new assay is an unknown, as it is impossible to know the number of positive hits before screening. One solution is to test a

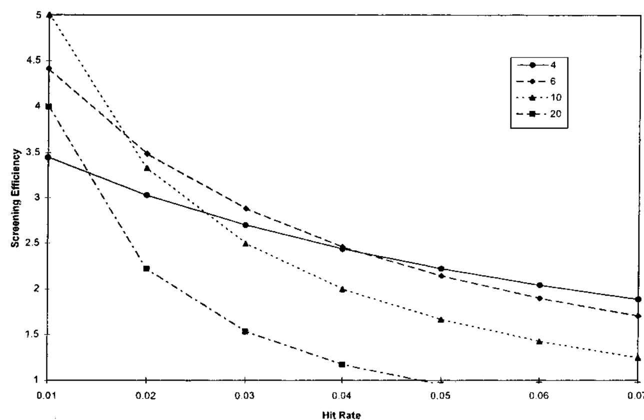


Figure 2. Screening efficiency as a function of the HR and N (expanded).

small set of mixtures similar to the proposed library to assess the HR of the new assay. However, this will decrease the overall efficiency of the process, and the resulting HR will depend heavily on how representative the selected mixtures are of the proposed library. For an existing test, it is possible to estimate an average value for the HR based on the test historical results, although we must keep in mind that this estimate depends heavily on the diversity and classes of compounds previously tested. If the compounds of the new libraries are drastically different from the compounds in historical testing, the library HR may vary greatly from the historical HR. In addition, certain types of assay targets are considered difficult and very low hit rates are obtained, e.g., protein−protein interactions while other targets give intermediate hit rates, e.g., enzyme inhibition tests. The first tier, whole organism tests commonly used in the agrochemical industry are typically run to yield hit rates in the range 0.01−0.10. This makes it worthwhile to reverse the argument and reanalyze eq 1 from the point of view of the number of mixtures that may show biological activity.

Let's again consider a library of M mixtures of N compounds, as before the total number of compounds is $M*N$. *Multiplying* both sides of eq 2, which relates HR to N when the screening efficiency is 1, by $M*N$, constant for a given library, we get

$$M*N*\mathrm{HR} = \left[\frac{N-1}{N^2}\right]*M*N$$

$$\text{number of active compounds} = M - \frac{M}{N}$$

or

$$\text{maximum number of active mixtures} = M - \frac{M}{N} \qquad (5)$$

As before $M*N*$HR is the number of active compounds present in the library and, in our analysis, the maximum number of mixtures showing biological activity. Taking HR from eq 2, now eq 5 becomes the maximum number of mixtures that can show biological activity and still will be faster to analyze the library by the mixture approach than by the traditional "one at a time" method (screening efficiency > 1). Let us again consider some special cases:

(a) First consider the case in which there is a greater number of mixtures than there are compounds in each mixture, i.e., $M > N$ or alternatively $M/N > 1$. In practice, the expression $M - M/N$ must be an integer. For example, in a library of 60 compounds divided into 15 mixtures ($M = 15$) of four compounds each ($N = 4$), resolving yields:

$$M - \frac{M}{N} = 15 - 3.75 = 11.25 \rightarrow 11$$

Then 11 is the maximum number of mixtures that can show biological activity and still the mixture approach will provide faster analysis of the library than by the traditional "one at a time" method. Eleven active compounds in a library of 60 compounds corresponds to 18% of the compounds in the library being active (HR = 11/60 = 0.18). Now consider the same 60-compound library formatted in 10 mixtures ($M = 10$) of six compounds each ($N = 6$). Then $M - M/N = 10 - 10/6 = 8.33$, and the maximum number of active compounds that may be present in the library and still allow a gain in the screening efficiency is 8 (HR = 0.13).

(b) Consider now $M \leq N$. In this case we have an equal number or fewer mixtures than compounds present in each of the mixtures. In many published studies this arrangement has been implemented: small numbers of mixtures each containing thousands or millions of compounds have been used in the first round of screening.[1,34] Obviously, $M/N \leq 1$ and from eq 5, $M - M/N$ will always lie between M and $M - 1$. Consider again the same 60-compound library but now let us divide the library in six mixtures of 10 compounds each. Resolving $M - M/N = 6 - 6/10 = 5.4$. In this case only five compounds (HR = 0.08) may show biological activity and still permit a gain in the screening efficiency. Obviously, when there is no active compound present in the library, no deconvolution is necessary and the $M < N$ arrangement will result in a substantially better screening efficiency. When only one mixture shows biological activity, the total number of tests necessary to identify the active compound is always $M + N$ whether $M > N$ or $M \leq N$. However, when more than one mixture shows biological activity, the $M > N$ arrangement always leads to faster library characterization. Therefore, as a general rule when deconvolution consists of screening all compounds in the mixture, we can state that, except when there is no active compound in the library, to gain in screening efficiency libraries should be designed such that

$$M \geq N \qquad (6)$$

i.e., to gain in screening efficiency the number of mixtures must be equal to or greater than the number of compounds in each mixture, except when there is no active compound present in the library.

Following the same procedure we can calculate screening efficiencies as a function of various values of active mixtures and N. Table 4 shows the calculated screening efficiencies for different combinations of M, the number of active mixtures, and N for a hypothetical 60-compound library. As we can see, when some activity is expected, the $M = 10$, $N = 6$ combination is equal to (only one active compound present in the library), or better than (more than one

Table 4. Screening Efficiency[a] as a Function of the Number of Active Mixtures and the Number of Compounds in Each Mixture (N)

		screening efficiency[e] (%)[f]								
		number of active mixtures[d]								
M[b]	N[c]	1	2	3	4	5	6	7	8	9
20	3	2.60(5.0)	2.3(10)	2.1(15)	1.88(20)	1.71(25)	1.58(30)	1.46(35)	1.36(40)	1.78(45)
15	4	3.20(6.7)	2.6(13)	2.2(20)	1.94(27)	1.71(33)	1.54(40)	1.40(47)	1.28(53)	1.18(60)
12	5	3.50(8.3)	2.7(17)	2.2(25)	1.88(33)	1.62(42)	1.43(50)	1.28(58)	1.15(67)	1.05(75)
10	6	3.80(10)	2.7(20)	2.1(30)	1.76(40)	1.50(50)	1.30(60)	1.15(70)	1.03(80)	0.94(90)
6	10	3.80(17)	2.3(33)	1.7(50)	1.3(67)	1.07(83)	0.91(100)			

[a] In a 60-compound library. [b] Number of mixtures. [c] Number of compounds in each mixture. [d] Number of active mixtures. [e] Number of operations in the "one at a time" approach (60) divided by the number of operations in the mixture approach. [f] Percentage of active mixtures.

Table 5. Maximum Number of Active Mixtures Tolerated by the Mixture Approach as a Function of the Number of Mixtures (M) and the Number of Compounds in Each Mixture (N)[a]

M[b]	N[c]	active mixtures[d]	%[e]	M[b]	N[c]	active mixtures[d]	%[e]
60	2	30	50	10	12	9	90
40	3	26	66	8	15	7	87
30	4	22	75	6	20	5	83
24	5	19	80	5	24	4	80
20	6	16	83	4	30	3	75
15	8	13	87	3	40	2	66
12	10	10	90	2	60	1	50

[a] Calculations made in a 120-compound library. [b] Number of mixtures. [c] Number of compounds in each mixture. [d] Calculated from eq 3 (maximum number of active mixtures = $M - {}^{M}/_{N}$) and rounded to the lower integer. [e] Percent of active mixtures with which the mixture approach still shows a gain in the screening efficiency.

compound in the library shows biological activity), the $M = 6$, $N = 10$ combination in terms of screening efficiency, i.e., $M \geq N$ is the best choice. As the number of active mixtures increases, the screening efficiency decreases more rapidly with N. For all values of N, the screening efficiency falls under 1.5 with more than six active compounds in the library (HR $\geq$ 0.1), falls under 2 with more than three active compounds in the library (HR $\geq$ 0.05), and never exceeds 4.

Similarly we can expand the analysis to a 120-compound library. Results are shown in Table 5. In this case as N decreases we can have a higher number of active mixtures, which is to say more active compounds present in the library, and still have an acceptable screening efficiency. Table 5 shows very clearly that $M \geq N$ leads to improved screening efficiency. The above rules and formulas are directly applicable for large numbers of compounds formulated as small mixtures where deconvolution involves only the synthesis of all compounds in the mixture where discrete numbers can be treated as continuous. Table 6 shows the number of operations needed to analyze small libraries, <49 compounds, as a function of the number of mixtures that show activity. An increase of at least 50% in the screening rate (screening efficiency = 1.5) is highlighted. Inspection of Table 6 makes it clear that the starting point for the mixture approach to afford a gain in screening efficiency is $M \geq 3$ and $N \geq 3$.

Identical considerations of HR, M, and N apply to the number of reactions required to produce a library for screening (vide infra II.4).

Table 6. Number of Operations Needed To Analyze the Library as a Function of the Number of Mixtures That Show Activity

M[a]	N[b]	M*N[c]	Number of Active Mixtures										
			0	1	2	3	4	5	6	7	8	9	10
			Number of Operations Needed to Evaluate the Library										
2	2	4	2	4	6								
3	2	6	3	5	7	9							
3	3	9	3	6	9	12							
4	3	12	4	7	10	13	16						
4	4	16	4	8	12	16	20						
5	4	20	5	9	13	17	21	25					
6	4	24	6	10	14	18	22	26	30				
7	4	28	7	11	15	19	23	27	31	35			
8	4	32	8	12	16	20	24	28	32	36	40		
9	4	36	9	13	17	21	25	29	33	37	41	45	
10	4	40	10	14	18	22	26	30	34	38	42	46	50
11	4	44	11	15	19	23	27	31	35	39	43	47	51
12	4	48	12	16	20	24	28	32	36	40	44	48	52
5	5	25	5	10	15	20	25	30					
6	5	30	6	11	16	21	26	31	36				
7	5	35	7	12	17	22	27	32	37	42			
8	5	40	8	13	18	23	28	33	38	43	48		
9	5	45	9	14	19	24	29	34	39	44	49	54	
10	5	50	10	15	20	25	30	35	40	45	50	55	60
6	6	36	6	12	18	24	30	36	42				
7	6	42	7	13	19	25	31	37	43	49			
8	6	48	8	14	20	26	32	38	44	50	58		
7	7	49	7	14	21	28	35	42	49	56			

[a] Number of mixtures. [b] Number of compounds per mixture. [c] Total number of compounds in the library.

We can also analyze special cases that might occur during the deconvolution process. Let us assume that the deconvolution of the most potent mixture yields a novel highly active compound. One might then choose not to deconvolute the remaining active mixtures. How much would the screening efficiency increase if only that most active mixture were deconvoluted? To analyze this possibility we can again consider the hypothetical example of Table 1, a library of 1000 compounds divided into 250 four-compound mixtures. If the first round of testing yielded 10 active mixtures and only one active mixture is deconvoluted. the screening efficiency will increase from 3.45 to 3.94 (Table 1). It can easily be seen in Table 1 that the larger the number of active mixtures, the larger the gain in screening efficiency realized by deconvoluting only one mixture. We can also analyze a general case. Consider the same library of M mixtures of N compounds each. The total number of tests needed to analyze the library in the "one compound at a time" approach is $M*N$; however, if only one mixture will be deconvoluted the total number of tests necessary to identify the active compound is $M + N$. We have previously defined the screening efficiency as the ratio of the number of tests needed to analyze the library by the traditional "one compound at a

time” approach divided by the number of tests needed to screen the library by the N-compound mixture approach; similarly we can define the screening efficiency with deconvolution of only one mixture as the ratio of the number of tests needed to analyze the library by the traditional “one compound at a time” approach ($M*N$), divided by the number of tests needed to screen the library by the N-compound mixture approach deconvoluting only the most active mixture ($M + N$). Thus in this case

$$\text{screening efficiency deconvoluting one mixture} = \frac{M*N}{M+N}$$

To estimate the increase in screening efficiency realized by deconvoluting only the most active mixture instead of all the active mixtures present in the library, we can utilize in any given case the above equation and eq 1.

I.4. Two-Step Deconvolution. On the basis of the previous analysis we can envisage increasing the overall throughput by using two curves at a time (see Figures 1 and 2). This will enlarge the application window of the method. We will first divide the library into M mixtures of N compounds per mixture as before but with a large value of N. After biological evaluation the compounds present in the active mixtures are now retested in mixtures containing n components in each mixture where $n < N$. That is to say, perform a first set of tests using the larger “N” available, and then retest the compounds in the active mixtures formatted as mixtures with a smaller “n”. Within this sublibrary the hit rate will be substantially higher than for the entire library. We designate this as HR2. Finally, the individual compounds in the mixtures active in the second biological evaluation are retested one at a time. This type of protocol, generally with more than two steps, has found considerable application with very large libraries produced by split and mix solid phase synthesis and is usually termed iterative deconvolution.[3] Recursive deconvolution is a particular implementation of iterative deconvolution that reduces the amount of resynthesis needed to complete the deconvolution procedure.[50] The method of synthesis defines the composition of the mixtures used in the various steps of deconvolution. Following the same process used above, let us first analyze a 1000-compound library formatted as 100 mixtures ($M = 100$) of 10 compounds in each mixture ($N = 10$). If 10% of the mixtures are active (HR = 0.01) and still assuming one active compound in each mixture, then there will be up to 10 active mixtures of 10 compounds each. If in the second step we use mixtures of five compounds each ($n = 5$) then we divide the 100 compounds from the 10 active mixtures into 20 mixtures of five compounds each, of which only 10 mixtures will show biological activity (we still assume the worst case, i.e., each active compound occurs in a different mixture), then the hit rate of the second round of testing HR2 will be $10/20 = 5/10 = 0.5$. If in the second step we instead divide the 100 compounds from the 10 active mixtures into 50 mixtures of two compounds each ($n = 2$), then the number of active mixtures in the second round of testing will again be 10. If we again assume one active compound per mixture, the second hit rate HR2 is brought to $10/50 = 2/10 = 0.2$. In general we can say that the maximum value of the second hit rate HR2 is n/N, which occurs when there is one active compound per mixture. Let us again consider a generic library of, as before, $M*N$ compounds. The number of operations to biologically evaluate the library by the “one compound at a time” method is $M*N$. If the library is tested as M mixtures of N compounds each, then, as before, $M*N*\text{HR}$ is the number of active mixtures. Dividing the compounds in the active mixtures into new smaller mixtures of n components leads to $(M*N^2*\text{HR})/n$, the total number of mixtures to be tested in the second round of testing, and

$$\frac{M*N^2\text{HR}}{n} * \frac{n}{N}$$

is the total number of active mixtures in the second round of testing. Thus finally

$$\frac{M*N^2\text{HR}}{n} * \frac{n}{\text{N}} * n$$

is the total number of compounds to test individually in a third phase of deconvolution. Then similarly to eq 1 we can express the screening efficiency in the double step testing method as

$$\text{screening efficiency} = \frac{M*N}{M + \frac{M*N^2\text{HR}}{n} + \frac{M*N^2\text{HR}}{N} * \frac{n}{N} * n}$$

resolving,

$$\text{screening efficiency} = \frac{1}{\frac{1}{N} + \frac{N}{n}*\text{HR} + n*\text{HR}} \quad (7)$$

As in eq 1, we can consider special cases for n and N. As in eq 1, for the screening efficiency to be > 1, HR must be $< 1/N$. When $n = 1$, step 2 is the final deconvolution step and eq 1, not eq 7, is valid. If $\text{HR} \geq 1/N$, all mixtures from step 1 are active but step 2 can still identify the active compounds if $\text{HR} < 1/n$. In other words if $1/n > \text{HR} \geq 1/N$ then

$$\text{screening efficiency} = \frac{1}{\frac{1}{N} + \frac{1}{n} + n*\text{HR}} \quad \text{for} \quad 1/N \leq \text{HR} \leq 1/n \quad (8)$$

If HR is also $\geq 1/n$, then all the mixtures in step 2 are also active, and the complete library must be deconvoluted. Thus when $\text{HR} \leq 1/n$, from eq 8

$$\text{screening efficiency} = \frac{1}{\frac{1}{N} + \frac{1}{n} + 1} < 1$$

In fact, considerations similar to those with the simple, one-step deconvolution still apply. It is interesting to note that to take full advantage of the technique for all compounds, N must be a multiple of n (vide infra). Experimentally, perhaps the best strategy will be to prepare first the n-compound mixtures and combine these to produce the N-compound mixtures for the first round of testing.

Utilizing eq 7 we can further study how a given value of N can determine the different possible values of n and hence

Table 7. Number of Operations in the Two-Step Deconvolution Approach

m^b	n^c	number of operations
2	50	52
4	25	29
5	20	25
10	10	20
20	5	25
25	4	29
50	2	52

[a] Calculated starting with mixtures of 100 compounds each ($N = 100$). [b] Number of mixtures in step 2, double scan method. [c] Number of compounds in each mixture in step 2, double scan method.

m, the number of second round mixtures. As a simple example, if $N = 16$ there are only three possible pairs of values for n and m: 8,2; 4,4; or 2,8. In the general case n,m it is clear that $N = n*m$ (in this example 2*8, 4*4, and 8*2). Replacing N in eq 7 by $n*m$

$$\text{screening efficiency} = \frac{1}{\frac{1}{N} + \frac{n*m}{n}*\text{HR} + n*\text{HR}} = \frac{1}{\frac{1}{N} + m*\text{HR} + n*\text{HR}} = \frac{1}{\frac{1}{N} + (n + m)*\text{HR}} \quad (9)$$

The difference between eq 9 and eq 1, the simple one-step approach, is that HR is multiplied by $n + m$ instead of N. In every case $n + m < N$ because $N = n*m$, which requires that the two-step approach is always better than the one-step approach except in the case of $N = 4$ when there is only one value for $n = m = 2$ and $n + m = n*m$. In this case the screening efficiency is the same in the one-step and two-step approaches. The numbers of tests required to analyze the library are exactly the same, but now testing more mixtures instead of single compounds in the deconvolution step.

Since $n + m = m + n$ in the two-step approach, it is the same to test n mixtures of m-compounds and then deconvolute or to test m mixtures of n-compound and then deconvolute. For instance, in the case of $N = 16$, the number of operations needed to test eight mixtures of two compounds each and then deconvolute is the same as to test two mixtures of eight compounds each and then deconvolute. In both cases the number of operations is 8 + 2 (or 2 + 8). The maximum screening efficiency occurs when $n + m$ is minimum, and recalling basic mathematics,[51] this happens when $n = m = \sqrt{N}$. In the previous example, 4 + 4 = 8 is less than 2 + 8 = 10. We can further exemplify this with the following case: for any given library of compounds let us assume $N = 100$, then the possible values for n and m are shown in Table 7. From Table 7 we can easily see that the maximum screening efficiency occurs when $n = m = 10 = \sqrt{100}$.

In general we can state that *in the range of validity of eq 7, if* $n*m = N$, *it is equivalent to use the pairings (N,n) or (N,m) to run the tests. The maximum screening efficiency is achieved when* $N = n^2$ *(i.e.,* $n = m$*), as in the pairings N,m = 100,10; 25,5; 16,4; or 9,3.*

Table 8. Possible Reagent Combinations for the Preparation of Three-, Four-, Five-, and Six-Compound Mixtures

p^a	q^b	N^c	p^a	q^b	N^c
3	1	3	1	5	5
1	3	3	6	1	6
4	1	4	1	6	6
1	4	4	2	3	6
2	2	4	3	2	6
5	1	5			

[a] Number of reagents A. [b] Number of reagents B. [c] Number of compounds per mixture.

II. Synthesis. II.1. General Considerations. We will focus our analysis on the design of synthetically prepared mixtures, i.e., reacting reagent A_1 with different reagents B_q to afford products A_1B_1, A_1B_2,...A_1B_q in the same reaction flask. Ideally the mixture should be prepared by a one-pot reaction where a set of reactants A_p and B_q would reach completion to give a "pure" mixture of the expected products A_pB_q without any byproducts This objective can frequently be realized in solid phase synthesis where reactions are driven to completion by use of excess reagents.[15] In our analysis, we assume that chemistry development has optimized the reaction conditions, so as to give a "pure" mixture. Also we will only consider reactions in which multiple reagents yield a single product:

$$A + B + ... \rightarrow C \quad (10)$$

In the case in which multiple reagents lead to multiple products,

$$A + B + C + ... \rightarrow D + E + ... \quad (11)$$

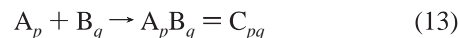
we assume that chemistry development has included implementation of a purification system that delivers only the desired products, for instance D, to be isolated and sent to biological screening in which the analysis of the latter case will be identical to the case in eq 10.

For a two-component reaction yielding one product such as

$$A + B \rightarrow C \quad (12)$$

the commercial availability of either starting materials will affect the strategy selected. In general the more readily available and hopefully the cheaper compound should be used in excess (vide infra, section II.2).

II.2. Mixture Design. Let us consider a $p*q$ combinatorial library arising from combining "p" reagents A and "q" reagents B.



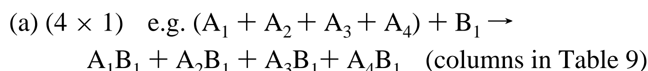
$$A_p + B_q \rightarrow A_pB_q = C_{pq} \quad (13)$$

If we want to make mixtures of $N = 3-6$ compounds, solving the equation $p*q = N$ for integer values of p, q, and N values, we find the possible combinations of reagents: For a mixture of three components, there are only two possibilities: $q = 1$, $p = 3$, and $q = 3$, $p = 1$. In the case of four-way mixtures, there are three possible combinations: $q = 4$, $p = 1$; $q = 1$, $p = 4$; and $q = 2$, $p = 2$. Similarly we can analyze four-, five-, and six-compound mixtures. These combinations are summarized in Table 8.

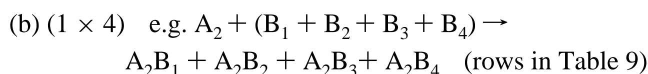
Table 9. Sixteen-Compound Combinatorial Library

	B_1	B_2	B_3	B_4
A_1	C11	C12	C13	C14
A_2	C21	C22	C23	C24
A_3	C31	C32	C33	C34
A_4	C41	C42	C43	C44

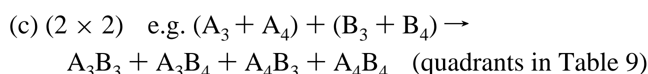
As a working example we will consider a $(A_{1-4})*(B_{1-4})$ square 16-compound combinatorial library (Table 9) that can be produced as four four-compound mixtures ($N = 4$) in which reagents B are commercially available. There are three possibilities for synthesizing the 16-compound library $(A_{1-4}) \times (B_{1-4})$ as four-compound mixtures:



(a) (4×1) e.g. $(A_1 + A_2 + A_3 + A_4) + B_1 \rightarrow A_1B_1 + A_2B_1 + A_3B_1 + A_4B_1$ (columns in Table 9)

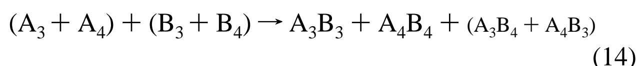


(b) (1×4) e.g. $A_2 + (B_1 + B_2 + B_3 + B_4) \rightarrow A_2B_1 + A_2B_2 + A_2B_3 + A_2B_4$ (rows in Table 9)



(c) (2×2) e.g. $(A_3 + A_4) + (B_3 + B_4) \rightarrow A_3B_3 + A_3B_4 + A_4B_3 + A_4B_4$ (quadrants in Table 9)

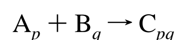
Assuming that B is commercially available but A is not, case (a) (4×1) is more favorable than case (b) (1×4) because the cheapest reagent can be used in excess if needed to drive the reaction to completion. Case (c) (2×2) on first examination seems more favorable than either (a) or (b) because it needs only four operations such as weighing, transferring, etc., instead of five as in cases (a) and (b) to prepare the final mixture (vide infra, section II.3). However, kinetic factors will generally influence case (c); for instance, if A_3 is more reactive than A_4 and B_3 is more reactive than B_4, products A_3B_3 and A_4B_4 will be likely produced in greater amounts than A_3B_4 and A_4B_3.



$$(A_3 + A_4) + (B_3 + B_4) \rightarrow A_3B_3 + A_4B_4 + (A_3B_4 + A_4B_3) \quad (14)$$

In the special case where the chemistry involved proceeds such that all the expected compounds are produced in similar equimolar amounts, case (c) can be a useful approach. In the well-studied case of peptide bond formation, attempts have been made to compensate for the lower reactivity of certain amino acids by increasing their concentration.[52-54] In general, however, it will be synthetically simpler to optimize cases (a) and (b) than (c); therefore, the safest mixture designs are the $(N \times 1)$ or $(1 \times N)$ arrays (columns or rows of Table 9, respectively).

II.3. Number of Weighings. As shown in the prior section, the synthesis of mixtures may also be advantageous from the point of view of the number of times reagents need to be weighed to produce the library. Consider a reaction in which reagents A and B are combined to produce a product C:



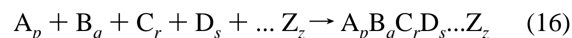
$$A_p + B_q \rightarrow C_{pq}$$

When preparing a single compound C_{pq}, we need to weigh out reagents A_p and B_q. To produce a library of N final products C one at a time, we need to run N reactions and weigh out A and B once for each reaction. Thus, the total number of weighings for all reagents is $2N$. When preparing a mixture of N compounds in a single reaction from p reagents A and q reagents B, where $N = p*q$, the total number of weighings for the two starting reagents is $p + q$. The ratio between the number of weighings needed in the single compound per reaction ("one at a time" approach) and the N compounds per reaction (mixture approach) will measure the reaction setup efficiency. Thus,

$$\text{reaction setup efficiency} = \frac{\text{number of weighings for "one at a time" approach}}{\text{number of weighings for mixture approach}}$$

$$\text{reaction setup efficiency} = \frac{2N}{p+q} = \frac{2*p*q}{p+q} \quad (15)$$

Equation 15 shows the reaction setup efficiency in terms of the number of weighings required for a two-reagent reaction. More generally we can consider a multicomponent reaction such as



$$A_p + B_q + C_r + D_s + \ldots Z_z \rightarrow A_pB_qC_rD_s\ldots Z_z \quad (16)$$

As for the specific case of a two-reagent reaction, when preparing a single compound per reaction we need to weigh each reagent A_p, B_q, C_r, D_s, ... Z_z once. N single reactions lead to N final products, and the number of weighings for all reagents is just $R*N$, where R is the number of reagents used. When preparing a mixture of N compounds per vessel from p reagents A, q reagents B, r reagents C, s reagents D, ... z reagents Z then $N = p*q*r*s*...*z$, the total number of weighings for the R starting materials is just $p + q + r + s + ...$ The ratio between the number of weighings needed in the single compound per vessel ("one at a time" approach) and N compounds per vessel (mixture approach) will measure the reaction setup efficiency. Thus,

$$\text{reaction setup efficiency} = \frac{R*N}{p+q+r+s+\ldots} = \frac{R*(p*q*r*s\ldots)}{(p+q+r+s+\ldots)} \quad (17)$$

Multicomponent reactions are very appealing for library production, because several points of diversity can be introduced in a single reaction. For instance, the three-component Passerini reaction[26,55] and the four-component Ugi reaction[56,57] have been widely used to generate large libraries of compounds. We can derive reaction setup efficiency equations for these three- and four-component reactions analogous to eq 15 for two-component reactions. For instance, the reaction setup efficiency for the Passerini reaction is

$$\text{reaction setup efficiency} = \frac{3*(p*q*r)}{(p+q+r)} \quad (18)$$

and for the Ugi reaction is

$$\text{reaction setup efficiency} = \frac{4*(p*q*r*s)}{(p+q+r+s)} \quad (19)$$

Table 10 shows the reaction setup efficiency calculated as a function of the number of reagents in each vessel for a two-, three-, and four-reagent reactions.

Table 10. Reaction Setup Efficiency as a Function of the Number of Reagents

N	*p*	*q*	*r*	*s*	reaction setup efficiency	*N*	*p*	*q*	*r*	*s*	reaction setup efficiency
2	2	1			1.33	12	4	3	1		4.50
3	3	1			1.50	12	3	2	2		5.14
4	4	1			1.60	16	16	1	1		2.67
4	2	2			2.00	16	8	2	1		4.36
6	6	1			1.71	16	4	4	1		5.33
6	3	2			2.40	16	4	2	2		6.00
8	8	1			1.78	2	2	1	1	1	1.60
8	4	2			2.67	3	3	1	1	1	2.00
12	12	1			1.85	4	4	1	1	1	2.29
12	6	2			3.00	4	2	2	1	1	2.67
12	4	3			3.43	6	6	1	1	1	2.67
16	16	1			1.88	6	3	2	1	1	3.43
16	8	2			3.20	8	8	1	1	1	2.90
16	4	4			4.00	8	4	2	1	1	4.00
2	2	1	1		1.50	8	2	2	2	1	4.57
3	3	1	1		1.80	12	12	1	1	1	3.20
4	4	1	1		2.00	12	6	2	1	1	4.80
4	2	2	1		2.40	12	4	3	1	1	5.33
6	6	1	1		2.25	12	3	2	2	1	6.00
6	3	2	1		3.00	16	16	1	1	1	3.37
8	8	1	1		2.40	16	8	2	1	1	5.33
8	4	2	1		3.43	16	4	1	1	1	6.40
8	2	2	2		4.00	16	4	2	1	1	7.11
12	12	1	1		2.57	16	2	2	2	2	8.00
12	6	2	1		4.00						

We can evaluate certain special cases: (a) when $p = N$ and $q = r = s = ...z = 1$ then the reaction setup efficiency is

$$\text{reaction setup efficiency} = \frac{R^*N}{p + 1 + 1 + ...1} = \frac{R^*N}{N + (R - 1)} \quad (20)$$

$$\text{when } N \text{ is large: reaction setup efficiency} \rightarrow R$$

(b) when $p = q = r = s = ... = z$ then

$$\text{reaction setup efficiency} = \frac{R^*p^R}{R^*p} = p^{R-1} \quad (21)$$

and when $p, q, r, s, ...z$ have been designed to have similar values then

$$\text{reaction setup efficiency} \rightarrow p^{R-1} \quad (22)$$

As we can see, the more similar the values of p, q, r, s, ..., the larger the reaction setup efficiency. For instance from Table 10, when $N = 16$, $p = 16$, and $q = r = s = 1$, the reaction setup efficiency = 3.37; however, when $N = 16$ and $p = q = r = s = 2$, then the reaction setup efficiency = 8. It should be reemphasized here that the effect of kinetic factors on the product distribution discussed in section II.2 may detract from the better reaction setup efficiency found when $p = q = r = s ... = z$.

II.4. Library Production Efficiency. Analysis of the number of reactions required to complete library synthesis and deconvolution in the mixture approach as opposed to synthesizing one compound at a time exactly parallels the argument developed in section I.3 for the number of tests required to screen this library. Thus eq 1 can be recast in terms of library production efficiency:

$$\text{library production efficiency} = \frac{1}{\frac{1}{N} + N^*\text{HR}} \quad (23)$$

It should be noted that these equations refer to synthesis and screening of the library in a single biological assay. If the library is to be screened in multiple assays, requiring different deconvolution pathways, both the screening and library production efficiencies will be lower (vide infra section IV.3).

Table 11. p^*q-Compound Combinatorial Library with One Active Member

	B_1	B_2	…..	B_{q-1}	B_q
A_1	C11	C12		C1q-1	C1q
A_2	**C21**	C22		C2q-1	C2q
…..					
A_p	Cp1	Cp2		Cpq-1	Cpq

Table 12. p^*q-Compound Combinatorial Library with Two Active Members

	B_1	B_2	…..	B_{q-1}	B_q
A_1	C11	C12		C1q-1	C1q
A_2	**C21**	C22		C2q-1	***C2q***
…..					
A_p	***Cp1***	Cp2		Cpq-1	**Cpq**

III. Double Scan Method. Considering a p^*q combinatorial library arising from combining "p" reagents A and "q" reagents B, as a grid of p^*q cells (Tables 11 and 12) each row can be synthesized as a q-compound mixture reacting a single reagent A with all reagents B_{1-q}. Similarly, each column can be synthesized as a p-compound mixture by reaction of a single reagent B with all A_{1-p}. In the double scan method,[13,46,58-61] also referred to as the indexed method, both the row and column mixtures of a library are synthesized, combining simultaneously the (p*1) technique with the (1 * q). Hence, if a single compound in the library is active (say **C21**, Table 11) then two mixtures, row 2 and column 1, should be active and the compound at the intersection is uniquely identified as the most active compound in the library and should be synthesized as a single pure compound for testing. However, if two compounds (say **C21** and **Cpq**, Table 12) are active, four mixtures are expected to be active (rows 2 and p and columns 1 and q). Therefore, four compounds, not only **C21** and **Cpq** but also ***Cp1*** and ***C2q***, are identified as potentially active and all four compounds must be synthesized and tested individually. The double scan method is a special case of positional scanning.[1,4]

The efficiency of the double scan method can be compared to the simple deconvolution approach that synthesizes only the q-($N \times 1$)-columns (or only the p-($1 \times N$)-rows) and resynthesizes all the individual compounds in the active mixtures. For the 25-member library depicted in Table 13, the traditional (5×1)-mixture approach would require initial synthesis of only the five column mixtures, while the double scan method requires synthesis of 10 mixtures. If only a single active compound exists in the library, then the simple approach would require synthesis of the five individual components of the single active column-mixture for a total of 10 reactions and 10 tests. The double scan method would

Table 13. Possible Dispositions of 0, 1, 2, and 3 Active Compounds in a 25-Compound Library[a]

Case I - no active compounds.

	B_1	B_2	B_3	B_4	B_5
A_1					
A_2					
A_3					
A_4					
A_5					

Case II - one active compound (1 row, 1 column)

	B_1	B_2	B_3	B_4	B_5
A_1					
A_2					
A_3					
A_4		X			
A_5					

Case III - two active compounds (1 row, 2 columns)

	B_1	B_2	B_3	B_4	B_5
A_1					
A_2					
A_3					
A_4		X		X	
A_5					

Case IV - two active compounds (2 rows, 1 column)

	B_1	B_2	B_3	B_4	B_5
A_1		X			
A_2					
A_3					
A_4		X			
A_5					

Case V - two active compounds (2 rows, 2 columns)

	B_1	B_2	B_3	B_4	B_5
A_1					
A_2				X	
A_3					
A_4		X			
A_5					

Case VI - three active compounds (1 row, 3 columns)

	B_1	B_2	B_3	B_4	B_5
A_1					
A_2		X	X		X
A_3					
A_4					
A_5					

Case VII - three active compounds (3 rows, 1 column)

	B_1	B_2	B_3	B_4	B_5
A_1					
A_2		X			
A_3		X			
A_4					
A_5		X			

Case VIII - three active compounds (2 rows, 2 columr

	B_1	B_2	B_3	B_4	B_5
A_1		X		X	
A_2					
A_3					
A_4		X			
A_5					

Case IX - three active compounds (2 rows, 3 columns)

	B_1	B_2	B_3	B_4	B_5
A_1					
A_2		X	X		
A_3					
A_4				X	
A_5					

Case X - three active compounds (3 rows, 2 columns)

	B_1	B_2	B_3	B_4	B_5
A_1		X			
A_2				X	
A_3					
A_4		X			
A_5					

Case XI - three active compounds (3 rows, 3 columns)

	B_1	B_2	B_3	B_4	B_5
A_1					
A_2		X			
A_3			X		
A_4				X	
A_5					

[a] The active mixtures are light shadowed. The intersections of the active column (5*1) and row (1*5) mixtures are dark shadowed. A possible arrangement of the active compounds is highlighted with "X".

require synthesis of a single compound at the intersection of the active column and the active row for a total of 11 reactions, making the simple deconvolution slightly more efficient. When multiple active compounds exist in the library, the efficiency of the double scan method may be greater than that of simple deconvolution depending on the unknown arrangement (same or different lines) of the active compounds. A number of these cases are depicted in Table 13, and the results are summarized in Table 14. Note that three active compounds out of 25 corresponds to a hit rate of 0.12 which is close to the limit of usefulness of the regular technique with $N = 5$, HR < 0.16 (section I.3., eq 5). When three active compounds are present, the library production efficiency of simple column mixture deconvolution varies from 1.25 to 2.50, while with the double scan method it varies across a narrower range from 1.32 to 1.92. If special precautions are used to design dissimilar mixtures, this will tend to favor cases I, II, V, and XI.

Table 14. Summary of Possible Outcomes

		no. of reactions run		library productn efficiency	
case	no. of active compds	column mixture deconvoltn	double scan	column mixture deconvoltn	double scan
I	0	5	10	5	2.5
II	1	10	11	2.5	2.27
III	2	15	12	1.67	2.08
IV	2	10	12	2.5	2.08
V	2	15	14	1.67	1.79
VI	3	20	13	1.25	1.92
VII	3	10	13	2.50	1.92
VIII	3	15	14	1.67	1.79
IX	3	20	16	1.25	1.56
X	3	15	16	1.67	1.56
XI	3	20	19	1.25	1.32

If at least one active compound is expected, the chemistry is compatible, and the 10 reactions can be run simultaneously, the double scan approach can be considered. Nevertheless, it can be shown that the same results can be obtained with savings in reagents, although not in reactions or screens run, to first synthesize, for example, the N-compound column mixtures and, after determining which mixtures are active, to synthesize the n-compound row mixtures using *only* those n B reagents that led to the activity. This strategy can be termed the partial ($N \times 1, 1 \times n$)-double scan.

Recalling eqs 1 and 2, it can be shown that increasing N will result in a better performance of the double scan-indexed strategy, if the test is compatible with N-compound mixtures. Since the arrangement of the active compounds is unpredictable, a single efficiency formula for the indexed strategy cannot be derived. However, squared arrangements of $p = q = N$ reagents to give a library of N^2 compounds can be easily discussed here. In the case of $N - 1$ active compounds lying in different rows and columns, $(N - 1)^2$ must be resynthesized and tested individually to complete deconvolution. Thus, $2N + (N - 1)^2 = N^2 + 1$ tests are needed to complete the evaluation of the whole N^2-library. Any other disposition of these $N - 1$ (or fewer) active compounds will result in greater efficiency. Nevertheless, it is possible to find the upper and lower limits for the efficiency that correspond to the active compounds lying in the same line (row or column) or in completely different row and column positions. The number of reactions to produce the library or test the library one compound at a time is N^2. The number of reactions to initially produce or test the library as N compound mixtures is $2N$. If all the active compounds occur in the same row or column (Table 13, cases I, II, III, IV, VI, VII), then an additional HR*N^2 reactions or tests of individual compounds are needed to complete deconvolution. Thus the upper limit for the efficiency of library production or screening is given by eq 24:

$$\text{upper efficiency limit} = N^2/(2N + \mathrm{HR}*N^2) = 1/((2/N) + \mathrm{HR}) \quad (24)$$

However, when the active compounds all occur on different rows and columns (Table 13, cases V, XI) an additional $(\mathrm{HR} * N^2)^2$ reactions or tests of individual compounds are required to complete deconvolution. Thus the lower limit for the efficiency of library production or screening is given by eq 25:

$$\text{lower efficiency limit} = N^2/(2N + (\mathrm{HR}*N^2)^2) = 1/((2/N) + (\mathrm{HR}*N)^2) \quad (25)$$

The actual efficiency for other arrangements (Table 13, cases VIII, IX, X) will be intermediate between the upper and lower limits of efficiency.

Comparing eq 1 with eqs 24 and 25, we can conclude that for zero or one active compound the simple ($N \times 1$) deconvolution technique always performs better, but for more than one active compound the double scan method can sometimes result in a better efficiency. In addition, the double scan method strategy can be useful at slighter higher rates than the simple deconvolution technique. Consequently, the ($N \times 1$, $1 \times N$) double scan technique or, if possible, the partial ($N \times 1$, $1 \times n$)-double scan method may be a useful choice.[62]

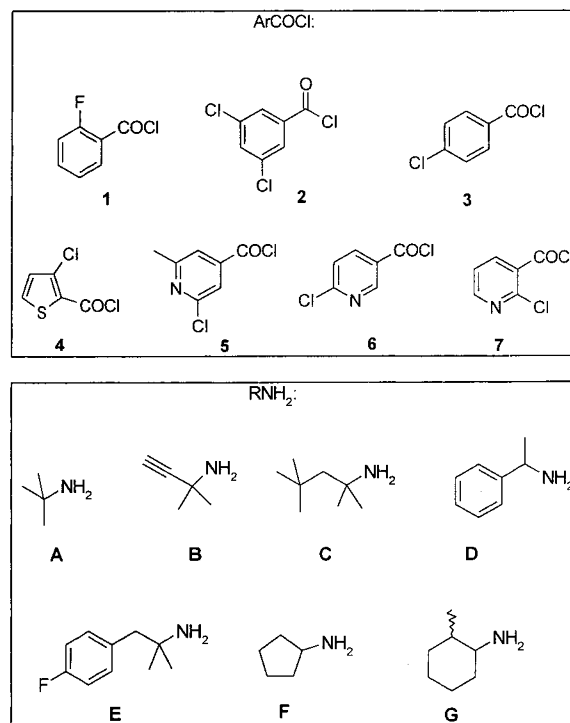
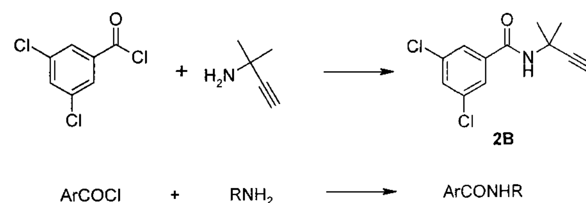


Figure 3. Reagents used for library synthesis.

Scheme 1. Synthesis of Pronamide and Analogues



IV. Experimental Example. IV.1. Library Synthesis. A practical example will serve to further illustrate the application of the eqs derived above in the agrochemical context. The preemergence mitotic inhibitor herbicide pronamide **2B**[63] can be prepared from 3,5-dichlorobenzoyl chloride and 1,1-dimethylpropargylamine in dichloromethane or THF with an organic base such as pyridine using well-established chemistry (Scheme 1). Furthermore, the chemistry is sufficiently general that related amides can be prepared efficiently under similar conditions, making this an ideal chemistry for synthesis of mixtures. Indeed amide formation has been a popular reaction in combinatorial chemistry.

A library of 49 amides including pronamide was prepared as a series of mixtures using the acid chlorides and amines shown in Figure 3. The acid chlorides selected are all aromatic, and the amines are all moderately hindered primary amines. Thus this library simulates to some degree a biased library that might have been designed for follow up of a weak screening hit. Fourteen reactions were run: first, 7 mmol of each acid chloride ArCOCl was individually reacted with a mixture containing 1 mmol each of the seven amines **A**−**G** to afford seven products (vide infra, the rows in Tables 14 and 15) designated **1(A−G)**, **2(A−G)**, ... **7(A−G)**.[46,58] Second, 7 mmol of each amine was individually reacted with

Table 15. Herbicide Screening Results

Mixture		(1-7)A	(1-7)B	(1-7)C	(1-7)D	(1-7)E	(1-7)F	(1-7)G
	Activity[a]	5	78	0	25	0	22	47
1(A-G)	0							
2(A-G)	51		2B					
3(A-G)	0							
4(A-G)	23							
5(A-G)	0							
6(A-G)	0							
7(A-G)	0							

[a] average % control of *Echinochloa crus-galli, Digitaria sanguinalis, Setaria viridis* and *Lolium multiflorum* preemergence at 4800 g ha^{-1}.

Key to Table 15

(dark shading)	Active Single Compound
(light shading)	Most Active Mixture

Table 16. Insecticide Screening against *Epilancha varivestis* at 300 ppm

Mixture		(1-7)A	(1-7)B	(1-7)C	(1-7)D	(1-7)E	(1-7)F	(1-7)G
	Activity[a]	-	-	+	+	+	+	-
1(A-G)	-							
2(A-G)	-							
3(A-G)	-							
4(A-G)	+			+	?	-	-	
5(A-G)	-							
6(A-G)	-							
7(A-G)	-							

[a]. + = active, - = inactive, ? = not tested.

Key to Table 16

(dark shading)	Predicted Active Single Compound
(light shading)	Active Mixture

a mixture containing 1 mmol each of the seven acid chlorides **1−7** to afford seven products (vide infra, the columns in Tables 15 and 16) designated **(1−7)A**, **(1−7)B**, ... **(1−7)G**. Each compound is produced twice, e.g. **2B** occurs in the mixtures **2(A−G)** and **(1−7)B**. GC and GC-MS were used to confirm the presence of the expected products in each mixture of amides, but no attempt was made to quantify each component. Analysis was rendered more difficult by the inclusion of acid chlorides **6** and **7** which have the same molecular weight, and by the inclusion of amine **G** which is a mixture of stereoisomers.

IV.2. Screening Results. The 14 mixtures were screened in first tier whole organism herbicide and insecticide screens normally used for single pure compounds. The amount of each mixture weighed for the assays was the same as the weight normally used for a single compound. Thus, on average, each compound will constitute about one-seventh of the mixture sample, and its effective application rate in the assay will be one-seventh of the normal application rate. In the herbicide screen, the results were scored as percent control while in the insecticide assay they were scored in a binary fashion as active or inactive.

When pronamide is tested as a pure single compound in the greenhouse, it requires ≥ 600 g ha^{-1} preemergence to show good activity against grasses (Table 17). A novel compound with this level of activity in conjunction with other desirable properties might be considered a viable lead for further synthesis. In mixtures **2(A−G)** and **(1−7)B** the effective rate of pronamide ≅ 685 g/ha when mixtures are tested at 4800 g ha^{-1} (Table 15). The percent control scoring used in this assay allows the single most active row and column to be identified, and the expected result was obtained—mixtures **2(A−G)** and **(1−7)B** were the most active (Table 15). This result indicates that this whole organism assay can successfully be used to test mixtures of at least seven compounds, provided that missing compounds only active at >600 g ha^{-1} is acceptable. It should be noted that when the same mixtures were retested at 1200 g ha^{-1}, corresponding to an effective rate of 170 g ha^{-1} for each compound in the mixture, no measurable weed control was recorded. Furthermore, if this screen had been run in a binary fashion where the cutoff between active and inactive corresponded to 40% control or 20% control, the results would not have unequivocally indicated pronamide as the most active compound.

Table 17. Preemergence Herbicidal Activity of Pronamide **2B** on Grass Weeds[a]

dose[b]	AM[c]	BYG[d]	CRB[e]	FOX[f]	RYE[g]
600	73	100	100	90	0
300	15	10	50	0	0

[a] Data are expressed as % control. [b] Expressed in g ha^{-1}. [c] Average % control of BYG, CRB, FOX, and RYE. [d] *Echinochloa crus-galli.* [e] *Digitaria sanguinalis.* [f] *Setaria viridis.* [g] *Lolium multiflorum.*

In an insecticide screen against *Epilancha varivestis* (Mexican bean beetle) which was scored in a binary fashion as active/inactive, one row and four columns were active (Table 16). Three of the four compounds at the intersections were synthesized and tested at 300 ppm on *Epilancha varivestis*: **4C** was active while **4E** and **4F** were inactive. Thus the activity of columns **(1−7)E** and **(1−7)F** may represent false positives. Unfortunately, the active compound **4C** was inactive on other insect species tested at lower rates in a follow up screen. Reports in the literature have indicated that, as observed here in this insecticide screen, hits identified by the indexed strategy have also been of lower potency than expected, presumably because of the additive effects of several weak compounds.[58] Therefore, although the whole organism assay can be used to test mixtures of at least seven compounds, it could be more efficient to reduce the number of compound in the mixture ($N = 4$ or 5) to reduce potential additive effects.

IV.3. Pronamide Library Results and Discussion. The compound mixtures synthesized in this example library are by necessity composed of groups of similar compounds since one reagent—the acid chloride or the amine—is constant in any given mixture. As discussed in section I.2, this should reduce the number of mixtures requiring deconvolution but increase the likelihood that several weakly active compounds, rather than one potent compound, are responsible for the activity of the mixture.

It is instructive to compare the efficiency of both the simple mixture deconvolution approach and the double scan method to the 49-compound amide library in terms of library production and screening separately and as an overall process. If we consider that the library contains a single herbicidally active compound, pronamide **2B**, then the hit rate is 1/49 = 0.0204. Applying eq 23 gives the library production efficiency of simple mixture deconvolution as measured by the number of reactions required:

$$\text{library production efficiency} = \frac{1}{\frac{1}{N} + (N^*\text{HR})} = \frac{1}{\frac{1}{7} + (7^*0.0204)} = 3.5$$

Similarly eq 1 gives the screening efficiency of simple mixture deconvolution as measured by the number of biological tests required:

$$\text{screening efficiency} = \frac{1}{\frac{1}{N} + (N^*\text{HR})} = \frac{1}{\frac{1}{7} + (7^*0.0204)} = 3.5$$

Applying eq 24 for the upper limit of library production efficiency of the double scan technique:

$$\text{library production efficiency} = \frac{1}{\frac{2}{N} + \text{HR}} = \frac{1}{\frac{2}{7} + 0.0204} = 3.27$$

In this case with a single active compound, eq 25 for the lower limit library production efficiency reduces to the same value. The value for screening efficiency in this example is also 3.27. As stated before (section II.5) when only one active compound is present in the library, simple mixture deconvolution outperforms the double scan technique (Table 18).

Table 18. Library Production Efficiency Comparison for Herbicide Discovery

	one at a time synthesis	simple mixture deconvolution		double scan
		row mixtures	column mixtures	
reactions required	49	14	14	15
efficiency[a]	1	3.5	3.5	3.27

[a] Efficiency = 49/reactions run to complete mixture strategy.

Table 19. Library Production Efficiency Comparison for Insecticide Discovery

	one at a time synthesis	simple mixture deconvolution		double scan
		row mixtures	column mixtures	
reactions required	49	14	35	18
efficiency[a]	1	3.5	1.4	2.72

[a] Efficiency = 49/reactions run to complete mixture strategy.

Examining the results obtained in screening for insecticidal activity gives a very different picture (Table 19). If only the row mixtures were synthesized and deconvoluted in the simple way, then simple mixture deconvolution again gives both a library production efficiency and a screening efficiency of 3.5. However, if only the column mixtures were synthesized, then four mixtures must be deconvoluted and the resulting library production and screening efficiencies drop drastically to 1.4, albeit still remaining above 1. The double scan strategy results in an intermediate efficiency of 2.72. The efficiency of the mixture approach using simple deconvolution or the double scan method is compromised by the apparent false positive results obtained with **(1−7)E** and **(1−7)F**. For the purposes of applying eqs 1, 15, and 24, the hit rate can be considered to be 4/49 = 0.0816 which gives the results already shown in Table 19. Note that eq 25 is not applicable since the four apparent hits are confined to one row mixture in the library. Assigning the true hit rate is complicated by the lack of data on compound **4D** (Table 16), making it unclear whether there was one hit **4C** (HR = 0.0204) or two hits **4C** and **4D** (HR = 0.0408) in the library. Examination of historical data for this assay gives a HR ≈ 0.05 for single random compounds purchased for lead discovery and screened at the full rate. Undoubtedly the hit rate for random compounds screened at one-seventh of this rate would be substantially lower. In any event, the actual maximum efficiency experienced can be significantly compromised by the presence of apparent false positives (columns E and F). One strategy to reduce the impact of false positives would be to retest all active mixtures prior to deconvolution either in the same screen or in a second tier screen. Permitting mixtures to progress into expensive higher tier screens may be considered undesirable.

The overall library production efficiency of the library for both herbicide and insecticide discovery together are considered in Table 20. From this perspective, the double scan technique outperforms the simple mixture deconvolution

Table 20. Overall Library Production Efficiency for Both Herbicide and Insecticide Discovery

		simple mixture deconvolution[a]		
	one at a time	row mixtures	column mixtures	double scan
reactions required	49	21	42	19
efficiency[b]	1	2.33	1.17	2.58

[a] Reactions required to complete strategy for herbicide and bean beetle insecticide screens. [b] Efficiency = 49/reactions run to complete mixture strategy.

Table 21. Overall Screening Efficiency for Both Herbicide and Insecticide Discovery

		simple mixture deconvolution		
	one at a time	row mixtures	column mixtures	double scan
screens required	98	28	49	33
efficiency[b]	1	3.5	2.0	2.97

[a] Screens required to complete strategy for herbicide and bean beetle insecticide screens. [b] Efficiency = 98/screens run to complete mixture strategy.

strategy even if the row mixtures were fortuitously selected for synthesis. This is because there are now two rows that must be deconvoluted **2(A−G)** and **4(A−G)**. The overall screening efficiency for both assays is presented in Table 21. It remains high if simple deconvolution is combined with fortuitous choice of the row mixtures but is very poor if the column mixtures are selected. The double scan technique gives an intermediate result and is the conservative choice. Generally libraries of compounds are prepared for use in multiple screens, whether run in parallel as is often the case for traditional agrochemical screening or serially as is more typical in pharmaceutical HTS screening, and thus any discussion of the viability of using mixtures must recognize the impact of performing multiple rounds of deconvolution for each assay.

Application of eq 1 to this library shows that even with five active compounds in the library, corresponding to HR = 0.10, the synthetic efficiency is 1.17 > 1. However, the practical utility of efficiencies slightly above 1 is minimal because of the additional tracking burden associated with deconvoluting mixtures.

Conclusions

We have analyzed the factors that affect the increase in throughput (screening efficiency and library production efficiency) experienced when libraries of compounds are screened or synthesized as mixtures and the mixtures are deconvoluted to single compounds in one or two steps. In addition we have compared this to the double scan or indexed library approach. The screening efficiency and the library production efficiency attained by assaying a library of compounds as mixtures and deconvoluting is a function of only two factors, the hit rate of the compounds in the library (HR) and the number of compounds in each mixture (N). In order for assaying of mixtures to increase throughput (screening efficiency > 1), HR must be less than $1/N$. In practice, a screening efficiency of at least 2 would be desirable to justify the use of mixtures which requires that each mixture contain at least 3 compounds ($N \leq 3$) and that no more than about five percent of the compounds in the library be active (HR $\leq$ 0.05). The maximum screening efficiency is obtained when there are no active compounds in the library (HR = 0) and is equal to the number of compounds in each mixture (N). With very low hit rates (HR < 0.01) the screening efficiency $\rightarrow N$, and the best results are obtained with the largest value of N that the assay can tolerate and still allow application of the desired quantity of each compound in the mixture. When formatting a library into mixtures for screening when the hit rate is expected to be in the 0.01 to 0.15 range, a greater tolerance for high values of HR while still maintaining improved screening efficiency is realized when the number of mixtures (M) is greater than the number of compounds in each mixture (i.e., $M \geq N$).

Application of a two-step deconvolution procedure is always more efficient than simple one-step deconvolution, except when $N = 4$. In two-step deconvolution the active N-compound mixtures are divided into n-compound mixtures ($n < N$) and reassayed. Finally the individual compounds from the active n-compound mixtures are assayed one at a time. The maximum screening efficiency is achieved when n is the square root of N.

When mixtures of compounds are synthesized in a single reaction, fewer weighings are required than to synthesize the same number of compounds individually. The reduction in the number of weighings and the resulting gain in reaction setup efficiency is most evident with multicomponent reactions and cases where equal numbers of each reagent are used.

The double scan approach, also known as the indexed approach, is less efficient than simple, one-step deconvolution when the hit rate is low. At higher hit rates the double scan approach can be more efficient than simple deconvolution. Its actual efficiency varies depending on the arrangement of active compounds in the library.

This analysis has focused on the synthesis of a library for a single screen. As demonstrated in the pronamide-based example library, when a library is made for assaying in multiple screens the resulting overall efficiency will be lower than that for a single assay because of the need for separate rounds of deconvolution for each assay.

Acknowledgment. The authors are indebted to Drs. Ted T. Fujimoto and Charles H. Reynolds for helpful discussions and suggestions.

References and Notes

(1) Houghten, R. A.; Pinilla, C.; Appel, J. R.; Blondelle, S. E.; Dooley, C. T.; Eichler, J.; Neftzi, A.; Ostresh, J. M. Mixture-based Synthetic Combinatorial Libraries. *J. Med. Chem.* **1999**, *42*, 3743−3778.

(2) Van Drie, J. H.; Lajiness, M. S. Approaches to Virtual Library Design. *Drug Discovery Today* **1998**, *3*, 274−283.

(3) Venton, D. L.; Woodbury, C. P. Screening Combinatorial Libraries. *Chemometrics and Intelligent Laboratory Systems* **1999**, *48*, 131−150. See also Houghten, R. A.; Pinilla, C.; Appel, J. R.; Dooley, C. T.; Cuervo, J. H. Generation and Use of Synthetic Peptide Combinatorial Libraries for Basic Research and Drug Discovery. *Nature* **1991**, *354*, 84−86.
(4) Snider, M. Screening of Compound Libraries ... Consomme or gumbo? *J. Biomolecular Screening* **1998**, *3*, 169−170.
(5) Maehr, H. Combinatorial Chemistry in Drug Research from a New Vantage Point. *Bioorg. Med. Chem.* **1997**, *5*, 473−491.
(6) Lam, K. S.; Lebl M.; Krchnak, V. The "one-bead-one-compound" Combinatorial Library Method. *Chem. Rev.* **1997**, *97*, 411−448.
(7) Choong, I. C.; Ellman; J. A. Solid-Phase Synthesis Applications to Combinatorial Libraries. *Annu. Rep. Med. Chem.* **1996**, *31*, 309−318.
(8) Balkenhohl, F.; Vondembusschehunnefeld, C.; Lansky, A.; Zechel, C. Combinatorial Synthesis of Small Organic-Molecules. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 2289−2337.
(9) Patel, D. V.; Gordon, E. M. Applications of Small-Molecule Combinatorial Chemistry to Drug Discovery. *Drug Discovery Today* **1996**, *1*, 134−144.
(10) Thompson, L. A.; Ellman, J. A. Synthesis and Applications of Small-Molecule Libraries. *Chem. Rev.* **1996**, *96*, 555−600.
(11) Beutel, B. A. Discovery and Identification of Lead Compounds from Combinatorial Mixtures. In *Topics in Drug Design and Discovery;* Trainor, G. L., Ed.; Academic Press: New York, 1997.
(12) Tan, D. S.; Foley, M. A.; Stockwell, B. R.; Shair, M. D.; Schreiber, S. L. Synthesis and Preliminary Evaluation of a Library of Polycyclic Small Molecules for Use in Chemical Genetic Assays. *J. Am. Chem. Soc*. **1999**, *121*, 9073−9087.
(13) Berk, S. C.; Chapman, K. T. Spatially Arrayed Mixture (spam) Technology: Synthesis of Two-Dimensionally Indexed Orthogonal Combinatorial Libraries. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 837−842.
(14) Fauchere, J. L.; Boutin, J. A.; Henlin, J. M.; Kucharczyk, N.; Ortuno, J. C. Combinatorial Chemistry for the Generation of Molecular Diversity and the Discovery of Bioactive Leads *Chemom. Intell. Lab. Syst.* **1998**, *43*, 43−68.
(15) Terrett, N. K.; Gardner, M.; Gordon, D. M.; Kobylecki, R. J.; Steele, J. Combinatorial Synthesis: the Design of Compound Libraries and Their Application to Drug Discovery. *Tetrahedron* **1995**, *51*, 8135−8173.
(16) Pirrung, M. C. Spatially Addressable Combinatorial Libraries. *Chem. Rev.* **1997**, *97*, 473−488.
(17) Schriemer, D. C., Hindsgaul, O. Deconvolution Approaches in Screening Compound Mixtures. *Comb. Chem. High Throughput Screening* **1998**, 155−170.
(18) Konings, D.; Wyatt, J. R.; Ecker, D. J.; Freier, S. M. Deconvolution of Combinatorial Libraries for Drug Discovery-Theoretical Comparison of Pooling Strategies. *J. Med. Chem.* **1996**, *39*, 2710−2719.
(19) Konings, D.; Wyatt, J. R.; Ecker, D. J.; Freier, S. M. Strategies for Rapid Deconvolution of Combinatorial Libraries: Comparative Evaluation Using a Model System. *J. Med. Chem*. **1997**, *40*, 4386-4395.
(20) Pirrung, M. C.; Chau, J. H. L.; Chen, J. L. Discovery of a Novel Tetrahydroacridine Acetylcholinesterase Inhibitor Through an Indexed Combinatorial Library. C*hem. Biol.* **1995**, *2*, 621−626.
(21) Boger, D. L.; Chai, W. Y.; Jin, Q. Multistep Convergent Solution-Phase Combinatorial Synthesis and Deletion Synthesis Deconvolution. *J. Am. Chem. Soc*. **1998**, *120*, 7220−7225.
(22) Haoyun, A. N.; Haly, B. D.; Cook, P. D. Discovery of Novel Pyridinopolyamines with Potent Antimicrobial Activity: Deconvolution of Mixtures Synthesized by Solution-Phase Combinatorial Chemistry. *J. Med. Chem*. **1998**, *41*, 706−716.
(23) Krchnak, V.; Lebl, M. Synthetic Library Techniques Subjective Biased and Generic Thoughts and Views. *Mol. Diversity* **1996**, *1*, 193−216.
(24) Ostergaard, S.; Holm, A. Synthesis and Screening of an Indexed Motif-Library Containing non-Proteinogenic Amino Acids. *J. Pept. Sci.* **1997**, *3*, 123−132.
(25) Panunzio, M.; Villa, M.; Missio, A.; Rossi, T.; Seneci, P. Solution Phase Library of Perhydrooxazin-4-ones. *Tetrahedron Lett.* **1998**, *39*, 6585−6588.
(26) Kim, S. W.; Bauer, S. M.; Armstrong, R. W. Multicomponent Solution Phase Synthesis of Dehydroamino Acid Derivatives Based on the Passerini Reaction. *Tetrahedron Lett.* **1998**, *39*, 7031−7034.
(27) Andrus, M. B.; Turner, T. M.; Asgari, D.; Li, W. The Synthesis and Evaluation of a Solution-Phase Indexed Combinatorial Library of non-Natural Polyenes for Multidrug Resistance Reversal. *J. Org. Chem*. **1999**, *64*, 2978−2979.
(28) Boger, D. L.; Goldberg, J.; Jiang, W.; Chai, W.; Ducray, P.; Lee, J. K.; Ozer, R. S. Andersson, C. Higher Order Iminodiacetic Acid Libraries for Probing Protein−Protein Interactions. *Bioorg. Med. Chem.* **1998**, *6*, 1347−1378.
(29) Merritt, A. T. Solution Phase Combinatorial Chemistry. *Comb. Chem. High Throughput Screening* **1998**, *1*, 57−72.
(30) Loughlin, W. A. Combinatorial Synthesis: a Heterocyclic Chemist's Perspective. *Aust. J. Chem.* **1998**, *51*, 875−893.
(31) An, H.; Haly, B. D.; Cook, P. D. New Piperazinyl Polyazacyclophane Scaffolds, Libraries and Biological Activities. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2345−2350.
(32) Wang, T. M.; An, H.; Vickers, T. A.; Bharadwaj, R.; Cook, P. D. Synthesis of Novel Polyazadipyridinocyclophane Scaffolds and Their Application for the Generation of Libraries. *Tetrahedron* **1998**, *54*, 7955−7976.
(33) An, H.; Wang, T. M.; Mohan, V.; Griffey, G. H.; Cook, P. D. Solution Phase Combinatorial Chemistry. Discovery of 13- and 15-membered Polyazapyridinocyclophane Libraries with Antibacterial Activity. *Tetrahedron* **1998**, *54*, 3999−4012.
(34) Carell, T.; Wintner, E. A.; Sutherland, A. J.; Rebek, J., Jr.; Dunayevskiy, Y. M.; Vouros, P. New Promise in Combinatorial Chemistry: Synthesis, Characterization, and Screening of Small-Molecule Libraries in Solution. *Chem. Biol*. **1995**, *2*, 171−83.
(35) Martin, E. J.; Blaney, J. M.; Siani, M. A.; Spellmeyer, D. C.; Wong, A. K.; Moos, W. H. Measuring Diversity: Experimental Design of Combinatorial Libraries for Drug Discovery. *J. Med. Chem*. **1995**, *38*, 1431−1436.
(36) Bleicher, K.; Lin, M. F.; Shapiro, M. J. Wareing, J. R. Diffusion Edited NMR: Screening Compound Mixtures by Affinity NMR to Detect Binding Ligands to Vancomycin. *J. Org. Chem*. **1998**, *63*, 8486−8490.
(37) Bonetto, V., Eriste, E.; Sillard, R. Combination of Fast Chromatographic Separations, Biosensor-Based Functional Assays, and Mass Spectrometry for Characterization of Complex Peptide Mixtures. *J. Protein Chem.* **1998**, *17*, 555−556.
(38) Terrett, N. Combinatorial Chemistry Screening Mixtures Using Affinity NMR. *Drug Discovery Today* **1998**, *3*, 242−242.
(39) Lin, M. F.; Shapiro, M. J.; Wareing, J. R. Screening Mixtures by Affinity NMR. *J. Org. Chem*. **1997**, *62*, 8930−8931.
(40) Meyer, B.; Weimar, T.; Peters, T. Screening Mixtures for Biological Activity by NMR. *Eur. J. Biochem.* **1997**, *246*, 705−709.

(41) Olah, T. V., McLoughlin, D. A.; Gilbert, J. D. The Simultaneous Determination of Mixtures of Drug Candidates by Liquid Chromatography Atmospheric Pressure Chemical Ionization Mass Spectrometry as an in vivo Drug Screening Procedure. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 17−23.
(42) Meyer, B.; Weimar, T.; Peters, T. Screening Mixtures for Biological Activity by NMR. *Eur. J. Biochem.* **1997**, *246*, 705−709.
(43) Blom, K. F.; Combs, A. P.; Rockwell, A. L.; Oldenburg, K. R.; Zhang, J.-H.; Chen, T. Direct Mass Spectrometric Determination of Bead Bound Compounds in a Combinatorial Lead Discovery Application. *Rapid Commun. Mass Spectrom*. **1998**, *12*, 1192−1198.
(44) Daley, D. J.; Scammell, R. D.; James, D.; Monks, I.; Raso, R.; Ashcroft, A. E.; Hudson, A. J. High-Throughput LC-MS Verification of Drug Candidates from 96-well Plates. A*m. Biotechnol. Lab*. **1997**, *15*, 24, 26, 28.
(45) Van Breemen, R. B.; Huang, C.-R.; Nikolic, D.; Woodbury, C. P.; Zhao, Y.-Z.; Venton, D. L. Pulsed Ultrafiltration Mass Spectrometry: a New Method for Screening Combinatorial Libraries. *Anal. Chem*. **1997**, *69*, 2159−2164.
(46) Pirrung, M. C.; Chen J. Preparation and Screening Against Acetylcholinesterase of a non-Peptide Indexed Combinatorial Library. *J. Am. Chem. Soc*. **1995**, *117*, 1240−1245.
(47) Neustadt, B. R.; Smith, E. M.; Lindo, N.; Nechuta, T.; Bronnenkant; A.; Wu, A.; Armstrong, L.; Kumar, C. Construction of a Family of Biphenyl Combinatorial Libraries: Structure−Activity Studies Utilizing Libraries of Mixtures. *Bioorg. Med. Chem. Lett*. **1998**, *8*, 2395−2398.
(48) Eliseev, A. V.; Curr. O. Emerging Approaches to Target-Assisted Screening of Combinatorial Mixtures. *Drug Discovery Dev*. **1998**, *1*, 106−115.
(49) Groten, J. P.; Tajima, O.; Feron, V. J.; Schoen, E. D. Statistically Designed Experiments to Screen Chemical Mixtures for Possible Interactions. *Environ. Health Perspect.* **1998**, *106*, 1361−1365.
(50) Erb, E.; Janda, K. D.; Brenner, S. Recursive Deconvolution of Combinatorial Chemical Libraries. *Proc. Natl. Acad. Sci. U.S.A*. **1994**, *91*, 11422−11426.
(51) In rectangles of the same area the minimum semiperimeter occurs in the square, i.e., $n = m$.
(52) Furka, A. Chemical Synthesis of Peptide Libraries, in *Combinatorial Peptide and non-Peptide Libraries*; Jung, G., Ed.; VCH: New York, 1996; Chapter 4, pp 111−137.
(53) Ostresh, J. M.; Winkle, J. H.; Hamashin, V. T.; Houghten, R. A. Peptide Libraries: Determination of the relative reaction rates of protected amino acids in competitive couplings. *Biopolymers* **1994**, *34*, 1681−1689.
(54) Eichler, J.; Houghten, R. A. Identification of substrate-analogue trypsin inhibitors through the screening of synthetic peptide combinatorial libraries. *Biochemistry* **1993**, *32*, 11035−11041.
(55) Armstrong, R., W.; Combs, A. T.; Tempest, P.A.; Brown, S. D.; Keating, T. A. Multiple-component Condensation Strategies for Combinatorial Library Synthesis. *Acc. Chem. Res*. **1996**, *29*, 123−131.
(56) Strocker, A. M.; Keating, T. A.; Tempest, P. A.; Armstrong, R. W. Use of a convertible isocyanide for generation of Ugi reaction derivatives on solid support: synthesis of α-acylaminoesters and pyrroles. *Tetrahedron Lett*. **1996**, *37*, 1149−1152.
(57) Keating, T. A.; Armstrong, R. W. Molecular diversity via a convertible isocyanide in the Ugi four-component condensation. *J. Am. Chem. Soc*. **1995**, *117*, 7842−7843.
(58) Smith, P. W.; Lai, J. J. Y. Q.; Whittington, A. R.; Cox, B.; Houston, J. G.; Stylli, C. H.; Banks, M. N.; Tiller, P. R. Synthesis and Biological Evaluation of a Library Containing Potentially 1600 Amides and Esters − A Strategy for Rapid Compound Generation and Screening. *Bioorg. Med. Chem. Lett.* **1994**, *4*, 2821−2824.
(59) Deprez, B.; Williard, X.; Bourel, L.; Coste, H.; Hyafil, F.; Tartar, A. Orthogonal Combinatorial Chemical Libraries. *J. Am. Chem. Soc*. **1995**, *117*, 5405−5406.
(60) Lyga, J. W.; Ali, S. F.; Webster, C. A.; Simmons, D. A.; Plummer, J.; Galietti, M. E. Synthesis and evaluation of an *indexed* combinatorial library for lead optimization. http://argon.ch.ic.ac.uk/ectoc/echet96/papers/004/index.htm.
(61) Tice, C. M.; Teixido, J. Screening indexed combinatorial libraries for agrochemical discovery. *Book of Abstracts,* 217th National Meeting of the American Chemical Society, Anaheim, CA, March 21−25, 1999; American Chemical Society: Washington, DC, 1999; AGRO-058.
(62) Houghten, R. A.; Appel, J. R.; Blondelle, S. E.; Cuervo, J. H.; Dooley, C. T.; Pinilla, C. The Use of Synthetic Peptide Combinatorial Libraries for the Identification of Bioactive Peptides (reprinted from *Biotechniques* **1992**, *13*, 412−421). *Pept. Res.* **1992**, *5*, 351−358.
(63) An, M. A.; Vaughn, K. C. Pronamide disrupts mitosis in a unique manner. *Pestic. Biochem. Physiol*. **1987**, *28*, 182−93.

CC0000390