

# How Large Does a Compound Screening Collection Need To Be?

Michael J. Lipkin<sup>\*1</sup>, Adrian P. Stevens<sup>1,4</sup>, David J. Livingstone<sup>2,3</sup> and C. John Harris<sup>1</sup>

<sup>1</sup>*BioFocus DPI, Chesterford Research Park, Saffron Walden, Essex, UK*

<sup>2</sup>*ChemQuest, Sandown, UK*

<sup>3</sup>*Centre for Molecular Design, University of Portsmouth, UK*

<sup>4</sup>*Current address: Accelrys Ltd., 334 Cambridge Science Park, Cambridge, UK*

**Abstract:** Increasingly, chemical libraries are being produced which are focused on a biological target or group of related targets, rather than simply being constructed in a combinatorial fashion. A screening collection compiled from such libraries will contain multiple analogues of a number of discrete series of compounds. The question arises as to how many analogues are necessary to represent each series in order to ensure that an active series will be identified. Based on a simple probabilistic argument and supported by in-house screening data, guidelines are given for the number of compounds necessary to achieve a “hit”, or series of hits, at various levels of certainty. Obtaining more than one hit from the same series is useful since this gives early acquisition of SAR (structure-activity relationship) and confirms a hit is not a singleton. We show that screening collections composed of only small numbers of analogues of each series are sub-optimal for SAR acquisition. Based on these studies, we recommend a minimum series size of about 200 compounds. This gives a high probability of confirmatory SAR (i.e. at least two hits from the same series). More substantial early SAR (at least 5 hits from the same series) can be gained by using series of about 650 compounds each. With this level of information being generated, more accurate assessment of the likely success of the series in hit-to-lead and later stage development becomes possible.

**Keywords:** Hit rate, diversity, library size, chemotype, series, parallel array, combinatorial chemistry, structure activity relationship.

## INTRODUCTION

When combinatorial chemistry first appeared on the pharmaceutical scene the emphasis was on the production of chemical arrays (or libraries) containing large numbers of compounds, either as singletons or in mixtures [1-3]. Its development, in conjunction with a number of key automation improvements such as high throughput screening (HTS), meant that hundreds of thousands of compounds could be routinely tested for biological activity [4, 5]. Diversity was considered to be a key design element in these early libraries. However, although the numbers of compounds involved in these early libraries appeared large on the human scale it was soon recognized that they represented only a very tiny fraction of the possible ‘drug-like’ molecules which could exist [6, 7]. Furthermore, screening performance often proved to be disappointing in comparison with that of ‘historical’ screening collections which had been built up by organizations over time and which often contained many distinct sets of compounds assembled without directed synthesis, such as “cherry picking” from other libraries, by purchase from commercial suppliers and/or from pre-existing in-house collections. Finally, hits identified from these early libraries often had poor properties that made them unsuitable as lead compounds (e.g., unsuitable functional groups, poor solubility, etc). Appreciation of this fact and the limited success of early screening campaigns soon led to the application of experimental design strategies [8, 9] to combinatorial chemistry libraries.

## EXPERIMENTAL DESIGN STRATEGIES

Two key concepts have emerged in library design: ‘drug-likeness’ and ‘focus’. The ‘drug-likeness’ concept refers to the realization that compounds within combinatorial libraries should have properties that are approximately equivalent to (or ideally mirror) those properties in drug molecules that confer their activity and discriminate them from common organic compounds. This is exemplified well by the original work of Lipinski and co-workers [10, 11], and the concept has been developed further by other groups (for examples, see references [12-15]). The second concept of ‘focus’ refers to the incorporation of as much information about the therapeutic target as possible into the design prior to synthesis. In effect, this latter concept is a direct attempt to ‘steer’ the design into a region of chemistry space that is likely to be associated with a desired therapeutic activity. In doing this, the intrinsic diversity of the resultant design is reduced. Indeed, it has been noted that the degree of diversity in a library is inversely related to the amount of information known about the target [16, 17].

Focused libraries can be sub-divided into two types, ‘target-focused’ and ‘ligand-focused’ [18]. A target-focused library makes use of structural information about a particular protein target or family of targets. It may be known, for example, that certain chemical groups make favorable interactions with a binding pocket or, conversely, that a particular molecular feature prevents binding. A ligand-focused library may be based on an active hit which has been identified in a primary screen or may be built on the basis of a pharmacophore which has been constructed from the analysis of screening results. In either case, a library will usually contain

\*Address correspondence to this author at the BioFocus DPI, Chesterford Research Park, Saffron Walden, Essex, CB10 1XL, UK;  
E-mail: Mike.Lipkin@glpg.com

at least one or more series of analogues that share a common chemotype, or molecular skeleton. In the extreme a diverse library will contain only one example of each chemotype whereas a focused library will consist of a series of analogues of only one chemotype.

### FACTORS AFFECTING SERIES SIZE

A collection of compounds generated using such design strategies will consist of a number of analogue series. How many representatives of each series are required? The answer, of course, depends largely on the context in which the question was asked. For example, is the critical matter one of cost or the likelihood of finding an active compound in a given series? As any series increases in size then the cost of synthesis and acquisition will rise as will the cost of testing it. The probability of finding an active or selective compound is also likely to rise as the size of the series increases although this is crucially dependent on the design of the series. The ideal situation is that the series should be sufficiently small to minimize the cost of synthesis, acquisition and testing, but large enough so as to sample a sufficient region of the chemically tractable space to be sure of identifying the required active compounds.

### FACTORS AFFECTING SERIES SIZE: SYNTHETIC EFFICIENCY

A further factor in series design is that chemical diversity appears to increase in a non-linear fashion as the series size increases. In an analysis of 12 series synthesized at the Novartis Institutes for Biomedical Research (NIBR), Jacoby and co-workers [19] showed that diversity, as measured by the variance of the information content of Daylight fingerprints, increased most rapidly up to a size of 1,000 compounds (Fig. 1). A “pragmatic conclusion for both the diversity and cost optimum” was that a series size of 2,000 is

preferable although for overall cost and other “practical reasons” it was generally better to double the size of the existing series rather than start a new project.

The NIBR combinatorial chemistry groups have, in recent years, synthesised and purified some 80,000 drug-like pure compounds, subdivided into around 40 individual combinatorial libraries and designed to explore new chemical space. The individual combinatorial library sizes as reported by Jacoby *et al.* vary between 500 and 5000 with most containing around 3,000 compounds. Interestingly, they contrasted these with lead optimisation series which, in a 2002 survey [20] of published series, were shown to contain fewer than 100 compounds in three quarters of the cases and were rarely more than 1,000 compounds.

### FACTORS AFFECTING SERIES SIZE: STRUCTURE-ACTIVITY RELATIONSHIP (SAR) EFFICIENCY

So, in terms of chemical diversity at least, it seems that there is some guidance in the literature with regards to size with respect to the efficiency of sampling of chemistry space *vs* enumerated size. However, the question of the likelihood of finding a hit has not been explicitly addressed. In this study we have set out to answer this question so that this information can be used as a guide to series size.

One can view an HTS screen as a purchase of knowledge; this knowledge is primarily contained in the hits obtained as a result of the screen. Clearly, one wants to obtain the maximum amount of knowledge for the minimum cost but both these quantities can be difficult to pin down. The screening side of the equation would seem to be most amenable to quantification but even here there can be complexities. For example, it may be more cost effective to screen more compounds if screening fewer would involve a cherry-picking operation from a screening collection. This would

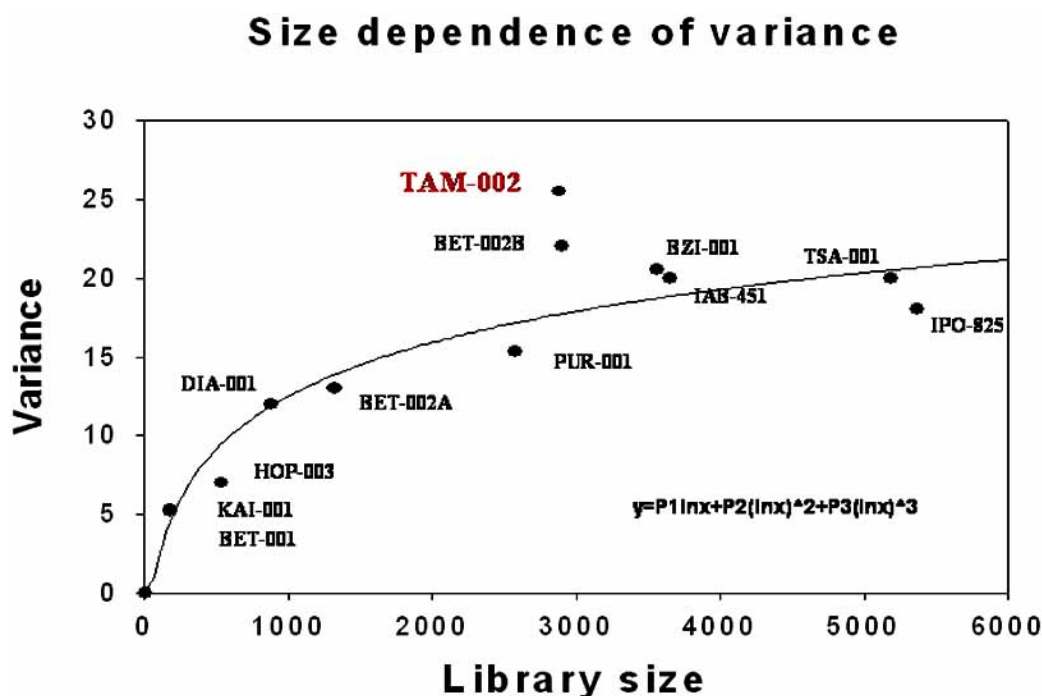


Fig. (1). Analysis of the size dependence of the theoretical chemical diversity of 12 NIBR combinatorial libraries, as published by Jacoby *et al.* [19]. Reproduced, with kind permission, from Ref. [19].

depend on the compound handling equipment available and also the nature of the screen itself. The knowledge gained side of the equation is even more difficult to quantify. This would not necessarily depend on the raw number of hits obtained. Other factors come into play: Are there identifiable series within the hits giving some indication of SAR? Are the compounds chemically tractable? Are there reactive functional groups or other undesirable properties? Intellectual property issues might also arise. Human judgment inevitably comes into play when evaluating the quality of knowledge gained from an HTS campaign.

There may be differences of opinion over the usefulness of the knowledge gained from an HTS campaign but this is not the same as there being no control over the probabilities of different outcomes. The composition of the collection of compounds screened will be the key determinant of these outcomes. In this paper we examine the issue of SAR acquisition. We show that, given a requirement for SAR from an HTS campaign there are optimal screening collection compositions (in terms of numbers of analogue series and size of each series) that will ensure a high probability of obtaining the desired SAR.

This paper builds on work by Nilakantan and coworkers [21] who examined the data from a set of 18 in-house HTS assays at Wyeth-Ayerst in order to evaluate the "hit-rate within a series". Each of the libraries used in a given assay was enumerated for the number of distinct chemical series present within the library and then each series was classified as active or inactive according to whether it contained any hits. Using simple probability arguments, they demonstrated that the probability,  $\mathbf{P}$ , of one or more specific analogues of a series of compounds being found to be active is related to the sample size of the series,  $n$ , and the probability,  $P_r$ , that any single compound in the series is declared a hit (Eqn. 1); this latter probability  $P_r$  was termed the hit rate within the series.

$$\mathbf{P} = 1 - (1 - P_r)^n \quad (1)$$

This equation was rearranged to solve for  $n$  so that it can be estimated for a particular probability,  $\mathbf{P}$ , based on assumed hit rates,  $P_r$ :

$$n = \ln(1 - \mathbf{P}) / \ln(1 - P_r) \quad (2)$$

One can then calculate the number of compounds required based on the hit rate ( $P_r$ ) and the probability one would like of being successful ( $\mathbf{P}$  – the "confidence" of success). Throughout the calculations in this paper we have taken  $\mathbf{P} = 0.95$ , i.e. a "95% confidence of being successful".

Examination of the screening assay results suggested an average hit rate of 4% for an active series, although it was pointed out that this was probably an overestimate for  $P_r$  since some active series might not have been detected since they may not have contained enough compounds,  $n$ , to reveal any potential activity. Nevertheless, substitution of a hit rate ( $P_r$ ) of 0.04 into equation 2 shows that, for 95% confidence ( $\mathbf{P}=0.95$ ), 73 compounds would be required to identify at least one hit from a series. Changing the hit rate to 0.03 or 0.05 alters  $n$  to 98 or 58 compounds respectively. Values of sample size for at least one hit at 95% confidence for a range of hit rates from 0.01 to 0.1 are shown in the first column of Table 1.

The confidence level approach adopted by Nilakantan and ourselves is better than designing a screening set so that the average outcome is the same as the desired outcome. For example, for a hit rate  $P_r=0.04$  one need only screen 25 compounds to get an average of 1 hit. However, if one does screen only 25 compounds then the probability of getting no hits is 0.64, so more likely than not one will get disappointing results. Using the 95% confidence method one can control the probability of disappointment, keeping it down to 0.05. Using this strategy means that the average number of hits observed per series will be higher than the desired number. For example, screening 73 compounds with a hit rate of 0.04 gives an average of 2.9 hits.

An alternative to the 'designed series' approach investigated by ourselves and Nilakantan *et al.* is to examine existing screening collections and find congeneric series within them, a 'found series' approach. This method is exemplified by Harper *et al.* [22]. In this work the series were generated using clustering to a Tanimoto similarity of 0.85. This results in much smaller and less diverse series than the combi-chem series that we have studied. A consequence of this lower within-series diversity is that the 'hit-rate' when the series is appropriate for the target is estimated to be  $P_r = 0.3$ , much larger than the values we have found, since the compounds are highly congeneric. Such 'found series' will vary in size, generally with many series of size 1 and some of a larger size depending on the history of the screening collection examined. The authors add an additional multiplier for each cluster which is the probability that a hit from that cluster can be turned into a lead. This factor is difficult to estimate but will be influenced by such factors as chemical tractability, desirable physical properties etc. Our model does not have this extra factor but in the context of designed libraries efforts will have been made to maximize attributes likely to give a lead from a hit. Like Nilakantan *et al.* these researchers only consider modeling the probability of finding  $\geq 1$  hits in a series.

It can be argued, of course, that a single hit does not indicate an active series and may represent a singleton active compound that is not amenable to further development. We have extended the probability model (see Appendix II) to give estimates of sample sizes required for other numbers of hits ( $k$ ) and also for a greater certainty ( $\mathbf{P}=0.99$  or "99% confidence") of obtaining those hits as shown in Table 1.

## PREDICTING THE THEORETICAL SIZE OF A SCREENING COLLECTION

A more realistic scenario is where a number of chemotype-series are screened as part of a screening campaign. Each series will have a different hit rate  $P_r$ . For some series  $P_r$  will be very small or zero, indicating that this chemotype is probably inappropriate for the target under investigation, whereas others will show good hit rates and are likely to provide SAR. A simplified model for this process is developed here. An idealized screening collection, composed of a number of equally sized chemotype-series is investigated. Each series, in the model, has a hit rate of either  $P_r$  or zero. The proportion of series with a hit rate of  $P_r$  is designated  $P_s$ . There are then two probabilities: The probability of a chemotype series being appropriate or otherwise for the target ( $P_t$ ) and if the series is appropriate for the target a "within series"

**Table 1.** Extension of Nilakantan Study, Showing Library Size Needed to have 95% and 99% Confidence of Obtaining Atleast  $k$  Hits for Different Values of  $P_r$ , the First Column,  $k=1$ , Corresponds to Nilakantan's Study

Hit Rate for the Series that Demonstrates it is of Interest for the Target ( $P_r$ )	Number of Compounds to Get 95% Confidence of Atleast $k$ Hits				Number of Compounds to Get 99% Confidence of Atleast $k$ Hits			
	$k=1$	$k=2$	$k=3$	$k=5$	$k=1$	$k=2$	$k=3$	$k=5$
0.01	298	473	627	913	458	661	837	1157
0.02	148	235	313	455	228	329	417	576
0.03	98	156	208	303	151	218	277	383
0.04	73	117	155	226	113	163	207	286
0.05	58	93	124	180	90	130	165	228
0.06	48	77	103	150	74	108	137	190
0.07	41	66	88	128	63	92	117	162
0.08	36	57	77	112	55	80	102	141
0.09	32	51	68	99	49	71	90	125
0.10	28	46	61	89	44	64	81	112

hit rate  $P_r$ . This simplified model is mathematically tractable, details can be found in Appendix I.

Reality is, as always, more complex. The within series hit rate  $P_r$  will be different for each series screened. Also, a real screening collection, unless it is being constructed from scratch, is likely to have a history and be composed of differing components, some diverse selections, some combinatorial chemistry libraries of different sizes and so on. However, this highly simplified model is sufficient to give some "ball-park" figures for the optimal series and screening collection sizes for various scenarios.

We explore the concept of series size using screening results obtained using our SoftFocus<sup>®</sup> libraries in 17 assays, drawn from campaigns conducted both in-house and in the hands of customers. The libraries considered in these studies incorporate a total of 85 molecular scaffolds, or cores, each of which has been designed to selectively target particular protein family classes, such as G-protein coupled receptors (GPCRs), protein kinases (PKs) and voltage-gated ion channels (VGICs) [23-27]. From the results, we will identify the theoretical minimum sizes for SoftFocus<sup>®</sup> arrays that should still retain a useful SAR generation capability. Finally, we will extend the series size concept to explore its logical impact on the size of screening collections. In summary, our aim is to identify, across a range of targets, how many chemotype arrays an ideal screening collection should contain in order to find at least one active series.

## METHODS

### Definitions

In the paper, the following definitions have been used:

- $k$  the minimum number of hits needed from a series for SAR.
- $n$  the number of compounds in each series.
- $P_s$  probability of a series having 'hit potential'.
- $P_r$  hit rate of a series that has hit potential (i.e., a "within series" hit rate).
- $N$  the number of series in the screening collection.

## DEFINITION OF SoftFocus<sup>™</sup> LIBRARY CHEMOTYPES

As in the work of Nilakantan and coworkers [21], the designation of each chemotype series was based on the concept of a set of compounds that share a common core. In the case of the SoftFocus<sup>®</sup> libraries, this was initially set *via* extraction of the Pharmascape<sup>™</sup> code from each compound identifier. This code relates to an in-house proprietary database of chemistries developed to describe ring scaffolds (RS). For example, in SoftFocus<sup>®</sup> library SFK01, compound identifiers take the form, 032\_0021\_0004. The first 3-digit prefix, the Pharmascape<sup>™</sup> code, identifies the core chemistry that was utilised. By extraction of this code from each library compound identifier, it was possible to rapidly assign each compound to a series class (chemotype code), without recourse to clustering or sub-structure searching methods.

Following the initial assignment, all RS cores used in the SoftFocus<sup>®</sup> libraries were subjected to a detailed comparison. This was to allow for the fact that an RS could be utilized in more than one library, depending on the subsequent chemistries applied. For example, in an SFG library, which is designed to target GPCRs, the same core could be used in libraries that contained either a primary amine attached to the core (therefore providing a primary target focus towards basic amine-recognising GPCRs, such as the serotonin or dopamine receptors) or in which this amine was neutralized by conversion into an amide. In such instances, all relevant SoftFocus<sup>®</sup> libraries were considered to contain an equivalent core and thus all compounds were given the same chemotype code even though the Pharmascape<sup>®</sup> code might differ. This approach more closely mirrors the chemotype definitions used by Nilakantan and coworkers.

## SoftFocus<sup>™</sup> BIOLOGICAL SCREENING DATA

Biological assay data from the screening of SoftFocus<sup>®</sup> libraries was drawn from high throughput screening (HTS) campaigns conducted either in-house on behalf of customers or from results provided by customers using our libraries. All of these assays used percent inhibition as the measure of activity. For functional assays, where the mode of action

(either agonism, antagonism or inverse agonism) might vary, only the occurrence of activity was counted.

As in the studies reported by Nilakantan and coworkers, protocols, screening concentrations and activity cut-offs were observed to vary from assay to assay. Consequently, in this study we used the same basic criteria to nominate a compound as active:

1. The compound must have been tested more than once.
2. The compound must have displayed a percent inhibition (or activation) value of  $\geq 50\%$  each time it was tested.

## RESULTS

Table 2 shows the screening results obtained in-house in 7 assays for 20 chemotypes contained in SoftFocus<sup>®</sup> kinase targeted libraries. The highest hit rates obtained for any series are around 36% with the lowest, of course, being zero. The average assay hit rate varies from assay to assay with the lowest being 0.3% (only 2 chemotypes active at 1% each) and the highest around 14% with ~ 10,000 compounds or more tested in each assay.

A conspicuous observation in the results presented in Table 2 is that multiple active chemotypes (where activity  $\geq 1\%$ ) are readily observed in all of the assays. Indeed, active chemotypes with viable SAR are even observed for assays B and F, for which the kinases screened, were regarded as "hard targets". Of course, it should be recognized that this

high chemotype activity rate is in part a direct result of the design methodology utilized (sic. SoftFocus<sup>®</sup> approach [24, 28]) in their production and that such a hit rate could not necessarily be expected if screened against a target that was outside the design scope. Table 3 contains screening results for 9 in house assays for 15 chemotypes contained in SoftFocus<sup>®</sup> GPCR targeted libraries. The highest hit rate obtained for any series was 35% and the average assay hit rates varied from 0 (1 assay) to ~9%, although this latter figure was boosted by the fact that 2 series gave very high hit rates, 35 and 22% respectively.

In comparison to the kinase screening results, the GPCR results presented in Table 3 generally shows a lower active chemotype rate per assay. However, this is predominantly a product of the selection of GPCRs tested and not an indication of a weaker design rationale. Many of the GPCRs screened should be regarded as "hard targets", since there was typically no small molecule structure activity data in the public domain. Instead, if any ligand data was available, it was limited and typically peptidic. Despite the relatively high proportion of hard target receptors, it is rewarding to see that active chemotypes could still be observed for 8 out of 9 of the assays.

In Table 4, a comparison of the active chemotype rates from across all the assays are presented, where the active chemotype rate ( $P_s$ ) is determined for specific series hit rates ( $P_r$ ) of  $\geq 1\%$ ,  $\geq 2\%$  and  $\geq 5\%$ , respectively. In effect, this table illustrates potential  $P_s$  values for each  $P_r$  value consid-

**Table 2. Screening Results Against Kinase Enzymes Obtained by BioFocus DPI, Using SoftFocus<sup>®</sup> Kinase Libraries. Percentage of Chemotype Identified as Active from %inh Screen. Blank Cells Indicate No Compounds Screened**

Chemotype	Assay						
	A	B	C	D	E	F	G
C07	3% (1236)	0% (1236)					
C09			0% (968)	1% (968)	0% (968)	0% (968)	0% (968)
C12			0% (968)	4% (968)	0% (968)	0% (968)	0% (968)
C14	3% (1331)	0% (1331)	1% (1357)	1% (1357)	1% (1357)	0% (1357)	0% (1357)
C16	1% (99)	0% (99)					
C21	1% (1522)	0% (1522)					
C30	3% (1180)	3% (1180)					
C34	1% (917)	0% (917)	37% (920)	36% (920)	6% (920)	0% (920)	1% (920)
C41	2% (381)	1% (188)	3% (381)	6% (381)	4% (381)	0% (381)	4% (381)
C44	2% (887)	0% (887)	10% (903)	37% (903)	0% (903)	0% (903)	0% (903)
C45	3% (536)	1% (536)					
C46	1% (1038)	0% (1038)					
C47	1% (1311)	0% (1311)					
C55	2% (485)	1% (485)	10% (485)	27% (485)	9% (485)	1% (485)	15% (485)
C57	2% (1238)	1% (1238)	2% (1238)	7% (1238)	2% (1238)	1% (1238)	4% (1238)
C58	1% (92)	0% (92)					
C59	2% (879)	0% (879)	23% (879)	19% (879)	3% (879)	0% (879)	2% (879)
C60	3% (1614)	0% (1614)	10% (1242)	13% (1242)	13% (1242)	0% (1242)	4% (1242)
C61	2% (531)	0% (372)	2% (531)	4% (531)	1% (531)	0% (531)	4% (531)
C62	4% (894)	0% (608)					
<b>Average</b>	<b>2.1% (16171)</b>	<b>0.5% (15533)</b>	<b>8.9% (9872)</b>	<b>14.1% (9872)</b>	<b>3.5% (9872)</b>	<b>0.3% (9872)</b>	<b>3.2% (9872)</b>

ered. In the kinase assays, it can be seen that even for a hit rate of  $\geq 5\%$ , that over half of the assays yield at least one active chemotype. Inspection of the percentage of active chemotypes in these assays illustrates that the lowest active rate ( $P_s$  value) was 6%. For the GPCR screens, the lower success rate in the assays suggests that a hit rate ( $P_r$  value) of either  $\geq 1\%$  or at most  $\geq 2\%$  is recommended. If we err on the side of caution and opt for a  $P_r$  value of  $\geq 1\%$  then the associated lowest active rate ( $P_s$  value) was 11%.

## DISCUSSION

The SoftFocus<sup>®</sup> libraries have been designed to contain ligands for particular classes of target, e.g. kinase, GPCR. Some targets, of course, are easier to find hits for than others. To give an idea of the range of possible values for the parameters  $P_s$  and  $P_r$  and the consequences for screening collection design we have considered both an “easy target” and a “hard target” case. Rough estimates for  $P_s$  and  $P_r$  for these two cases were generated from the data.

“Easy” target: For this case we can obtain a rough estimate of  $P_r$  from Tables 2 and 3 where SoftFocus libraries have been screened against their receptor class (though some of these targets could be defined as “hard”). Taking the average  $P_r$  across Tables 2 and 3 excluding those cases where there were no hits we get an estimate of  $P_r$  of about 5%. This can be taken as a typical  $P_r$  where a focused library is screened against its receptor class. Sometimes the hit rate is much higher. However  $P_s$  must also be taken into account. Using the 5% hit rate column of Table 4 we can see that out of 156 series-screening events 22 had hit rates  $\geq 5\%$  giving an estimate of  $P_s$  of 22/156 or about 14%. This suggests that even for an easy target one should plan to screen several chemotype libraries. For our modeling work we have taken the “easy” target parameters as  $P_r = 0.05$  and  $P_s = 0.15$ .

“Hard” target: Examining Table 4 we can see that there were 3 assays (F, I, N) where all chemotype hit rates were below 2%. The most difficult target with hits was assay N where only 11% of chemotypes screened showed hits. For assay O there were no hits which suggests there may be “harder” (maybe impossible) targets. This data cannot be used to give an estimate for  $P_r$ . We shall take the parameters  $P_r = 0.01$  and  $P_s = 0.1$  as practical “hard target” estimates.

With these estimates of  $P_r$  and  $P_s$  using the simple model of biological screening as described in appendix I it is possible to estimate the optimal library size and screening collection size for these scenarios. Tables 5 and 6 show total screening collection sizes required for a number of library sizes for the “easy target” and “hard target” cases. It can be seen that for  $k = 1$  the optimal library size is 1, i.e. a fully diverse collection. Hence if one is absolutely certain that one requires no confirmatory SAR than this is the optimal strategy and this is true for both “easy” and “hard” target cases. If one requires SAR than there are clear optimal library sizes (highlighted). Fig. 2 shows this information in graphical form. The asymptotes in the graphs for  $k = 2$ ,  $k = 3$  and  $k = 5$  show that composing a screening collection out of chemotype series that are too small (e.g.  $< 50$  compounds) in the hope of gaining SAR is a sub-optimal strategy. The screening collection size required to get the SAR becomes huge. This is because the probability of the under-represented series showing the required SAR becomes vanishingly small. In contrast, using chemotype series that are sub-optimal in the sense of being too large brings us to the other side of the minima where the size of the screening collection needed to obtain the desired SAR increases relatively slowly. It therefore makes sense to err in favour of series that are too large rather than too small when designing a screening collection. As expected, both the optimal library size and screening

**Table 3. Screening Results Against GPCR Receptors Obtained by BioFocus DPI, Using SoftFocus<sup>®</sup> GPCR Libraries. Percentage of Chemotype Identified as Active from %inh Screen. Blank Cells Indicate No Compounds Screened**

Chemotype	Assay								
	H	I	J	K	L	M	N	O	P
C02					0% (5)				
C04		0% (106)		0% (106)	0% (106)	0% (106)	0% (106)	0% (243)	2% (243)
C15			6% (501)					0% (502)	0% (501)
C17		0% (1000)		0% (1000)	0% (1000)				
C22					0% (3)	0% (40)			
C23		0% (75)		0% (75)		0% (75)	0% (75)		
C27						2% (1000)			
C28			0% (1)					0% (1)	0% (1)
C32		1% (146)		0% (146)	6% (146)	0% (146)	0% (146)	0% (1050)	35% (1050)
C36		1% (95)		1% (95)	3% (96)	0% (96)	0% (96)	0% (207)	22% (207)
C49			8% (38)					0% (38)	0% (38)
C56					0% (1000)				
C63		0% (652)	5% (983)	1% (652)	4% (652)	0% (652)	1% (652)	0% (983)	2% (983)
C65		0% (1229)		0% (1229)	1% (1040)	0% (1229)	0% (1229)		
C66	4% (1039)	0% (1040)		0% (1040)	1% (1040)	0% (1040)	0% (1040)		
Average	4.0% (1039)	0.3% (4343)	4.9% (1523)	0.3% (4343)	1.6% (5088)	0.3% (4384)	0.1% (3344)	0.0% (3024)	8.7% (3023)

**Table 4.** Percentage of Active Chemotypes Series ( $P_s$ ) that Displayed a Hit Rate ( $P_r$ ) Greater Than 1%, 2% and 5%, Respectively. Bold Italic Rows are Assays where All Chemotype Hit Rates were Below 2% (Referenced in the Text)

			Proportion of 'Active' Chemotypes					
			Active $\geq$ 1% Hit Rate ( $P_r$ )		Active $\geq$ 2% Hit Rate ( $P_r$ )		Active $\geq$ 5% Hit Rate ( $P_r$ )	
Target Focus	Assay	No. of Chemotypes Screened	Percentage ( $P_s$ )	Count	Percentage ( $P_s$ )	Count	Percentage ( $P_s$ )	Count
Kinase	A	18	100%	18	61%	11	0%	0
	B	18	28%	5	6%	1	0%	0
	C	11	82%	9	73%	8	45%	5
	D	11	100%	11	82%	9	64%	7
	E	11	73%	8	55%	6	27%	3
	<b><i>F</i></b>	<b><i>11</i></b>	<b><i>18%</i></b>	<b><i>2</i></b>	<b><i>0%</i></b>	<b><i>0</i></b>	<b><i>0%</i></b>	<b><i>0</i></b>
	G	11	64%	7	55%	6	9%	1
GPCR	H	1	100%	1	100%	1	0%	0
	<b><i>I</i></b>	<b><i>8</i></b>	<b><i>25%</i></b>	<b><i>2</i></b>	<b><i>0%</i></b>	<b><i>0</i></b>	<b><i>0%</i></b>	<b><i>0</i></b>
	J	4	75%	3	75%	3	75%	3
	K	8	25%	2	0%	0	0%	0
	L	10	50%	5	30%	3	10%	1
	M	9	11%	1	11%	1	0%	0
	<b><i>N</i></b>	<b><i>9</i></b>	<b><i>11%</i></b>	<b><i>1</i></b>	<b><i>0%</i></b>	<b><i>0</i></b>	<b><i>0%</i></b>	<b><i>0</i></b>
	O	9	0%	0	0%	0	0%	0
P	7	57%	4	57%	4	29%	2	

collection size increase as we move from an easy target to a hard target.

### THE COST OF SAR

Screening for SAR is likely to involve additional costs over screening fully diverse compounds for singleton hits. Examining the optimal strategies given by the minima in Fig. 2 we can see a steady climb in the number of compounds that need to be screened as we move from singleton to  $k = 2, 3$  and 5. However, this is not a fair comparison since the total number of hits garnered is different for the different cases.

Interestingly, under this model, the total number of hits obtained depends only on the total screening collection size and the product  $P_s P_r$  and not on the size of analogue blocks used. To see this, consider extracting a completely random compound from the screening collection. This compound will have a probability  $P_s$  of being within an active analogue series and if it is in an active series a probability  $P_r$  of being active giving an overall probability of  $P_s P_r$  of being active. Hence accounting for the total number of hits obtained gives no differentiation between the different screening collection design strategies. When screening for SAR, hit series containing less than the desired number of compounds will be inevitably obtained but these series are not counted as "successes". This is why more compounds must be screened when SAR is the aim.

A useful comparison is to compare the costs of screening for SAR (i.e. considering series of hits of size less than  $k$  of little value) compared with accepting singleton hits to be of equal value to hits that form part of a series. Table 7 shows the details. For example, taking the easy target case and 5 hits as the requirement we can see that to get a 95% confi-

dence of at least 5 singleton hits we need to screen 1218 diverse compounds. Conversely, if we are only interested in analogue series of 5 or more compounds than we must screen 3115 compounds composed of 24 analogue blocks of about 129 compounds each, Table 5. This is about 2.6 times the number of compounds needed to get the singleton hits and this inflation represents the cost of acquiring the SAR. Therefore, if one believes that a hit series of 5 analogues is worth more than 2.6 times that of 5 singleton hits then the strategy of screening analogue blocks of compounds should be chosen over that of diversity.

### SUMMARY

The value of a screening campaign is the information that is garnered from the data. This information lies primarily in the resulting hits. Crucially, the quality of these hits depends on the compounds that went into the screen. Many factors influencing compound quality such as appropriate properties, lack of reactive functional groups, chemical tractability and others have been extensively discussed elsewhere. In this paper we have concentrated on the issue of SAR generation. In order to generate SAR the screening collection needs to be composed of blocks of compounds which are structural analogues. We have shown that the size of these blocks must not be too small since the total size of the screening collection required to achieve the desired SAR then becomes very large. Taking the "hard target" example as the one to base our recommendations on (the "easy targets" can take care of themselves) we can see that to get the bare minimum confirmatory SAR of a series of at least 2 compounds of the analogue series blocks should be about 200 compounds each ( $\geq 186$ ; Table 6). This can be considered the minimum recommended size. To get the more substantial SAR that exists in a congeneric hit series of at least 5 compounds the

**Table 5.**  $P_s = 0.15$ ;  $P_r = 0.05$ . Sizes of Hypothetical Screening Collections to Get 95% Confidence of SAR for  $k=1,2,3$  and 5 for the “Easy Target” Case. The Smallest Screening Collections that will Give the Required SAR are Underlined. Figures in the Table have been Rounded to the Nearest Integer. This Means that Library Size Times Number of Libraries Required Does Not Exactly Equal the Figures in the Total Number of Compounds Columns in the Table

Library Size	$k=1$		$k=2$		$k=3$		$k=5$	
	No of Libraries Required	Total No. of Compounds	No of Libraries Required	Total No. of Compounds	No of Libraries Required	Total No. of Compounds	No of Libraries Required	Total No. of Compounds
<i>1</i>	<b>398</b>	<b>398</b>	N/A	N/A	N/A	N/A	N/A	N/A
2	203	407	7987	15974	N/A	N/A	N/A	N/A
3	139	416	2753	8260	159771	479313	N/A	N/A
4	106	425	1423	5693	41498	165991	N/A	N/A
5	87	434	882	4412	17243	86216	63908954	319544768
10	48	483	230	2304	1735	17346	313574	3135737
25	26	653	54	1358	156	3891	2786	69647
<b>38</b>	22	827	<b>33</b>	<b>1267</b>	66	2511	501	19047
50	20	1005	26	1309	42	2098	191	9562
<b>70</b>	19	1330	21	1498	<b>28</b>	<b>1930</b>	72	5033
75	19	1415	21	1560	26	1936	61	4551
100	19	1855	19	1920	21	2112	34	3389
<b>129</b>	18	2381	19	2405	19	2488	<b>24</b>	<b>3115</b>
150	18	2766	19	2777	19	2820	21	3196
200	18	3687	18	3688	18	3696	19	3795
250	18	4608	18	4608	18	4610	19	4631
300	18	5530	18	5530	18	5530	18	5534
350	18	6452	18	6452	18	6452	18	6452
400	18	7373	18	7373	18	7373	18	7373
450	18	8295	18	8295	18	8295	18	8295
500	18	9217	18	9217	18	9217	18	9217
550	18	10138	18	10138	18	10138	18	10138
600	18	11060	18	11060	18	11060	18	11060
650	18	11982	18	11982	18	11982	18	11982

recommended block size is about 650 compounds ( $\geq 643$ ; Table 6). Of course, detecting SAR at the primary screening level involves assaying more compounds and so costs must be higher. We have shown that the consequent inflation of the number of compounds required over a conventional diverse screen is a factor of 2 for the minimal SAR of a two compound cluster to 2.6 for an SAR cluster of at least 5 compounds. These numbers are not onerous given the capabilities of modern screening technologies. We suggest that the inflation in overall costs will probably not scale linearly with these numbers due to bulk savings. This is particularly the case for compound collections consisting of analogues synthesized in parallel. The robust and flexible chemistry generally used in such campaigns is easily extrapolated to allow efficient follow up of the initial SAR clusters. This contrasts with the difficulty and synthetic costs of following up individual “one-off” compounds as hits. It is interesting to note that the inflation in the number of compounds re-

quired for an SAR cluster of at least 5 compounds is not that much greater than for an SAR cluster of at least 2 compounds. Indeed, from a medicinal chemistry perspective the larger sets of analogue hits are likely to provide substantially more information about the probable success of the series in a subsequent hit-to-lead phase.

In this study we have used data derived from screening focused combinatorial libraries against their designed target family. However, the arguments about designing screening collections using blocks of analogue series also apply in a general screening setting. Without the design input the hit rates are expected to be lower. This is particularly true of the “easy target” case where we have found the number of compounds needed to get hits to be quite small. Such small numbers would not be recommended for general screening situations where compound numbers similar to those from the “hard target” study would be more appropriate.

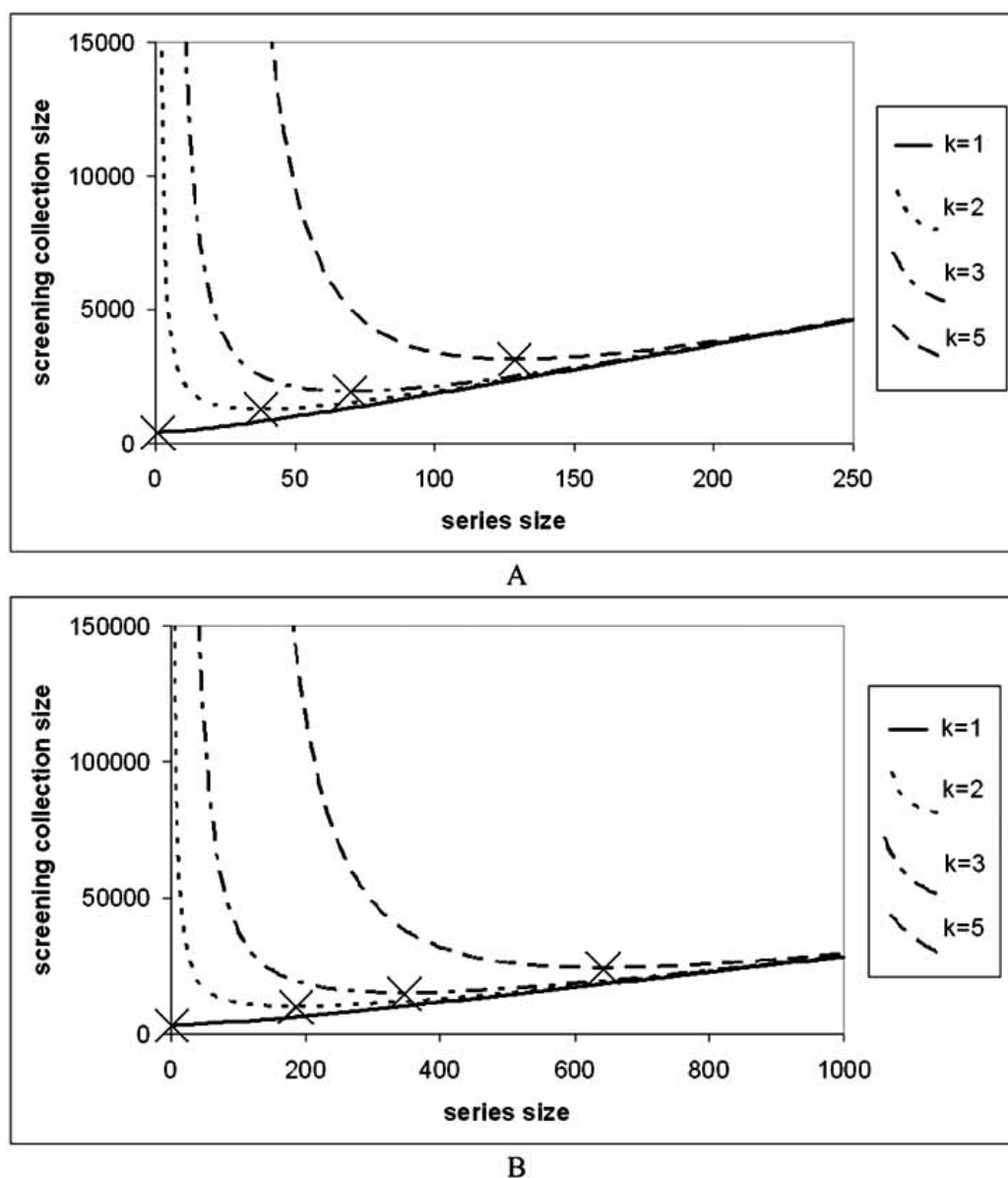
**Table 6.**  $P_s = 0.10$ ;  $P_r = 0.01$  Sizes of Hypothetical Screening Collections to Get 95% Confidence of SAR for  $k=1,2,3$  and 5 for the “Hard Target” Case. The Smallest Screening Collections that will Give the Required SAR are Underlined. Figures in the Table have been Rounded to the Nearest Integer. This Means that Library Size Times Number of Libraries Required does not Exactly Equal the Figures in the Total Number of Compounds Columns in the Table

Library Size	$k=1$		$k=2$		$k=3$		$k=5$	
	No. of Libraries Required	Total No. of Compounds	No. of Libraries Required	Total No. of Compounds	No. of Libraries Required	Total No. of Compounds	No. of Libraries Required	Total No. of Compounds
<i>1</i>	<b>2994</b>	<b>2994</b>	N/A	N/A	N/A	N/A	N/A	N/A
2	1504	3008	299572	599143	N/A	N/A	N/A	N/A
3	1007	3021	100526	301579	29957321	89871964	N/A	N/A
4	759	3035	50600	202398	7545924	30183694	N/A	N/A
5	610	3049	30563	152813	3041166	15205829	$\sim 3.0 \times 10^{11}$	$\sim 1.5 \times 10^{12}$
10	312	3118	7021	70205	263130	2631303	1239553002	12395530023
25	133	3333	1161	29037	15356	383898	6663166	166579141
50	74	3717	333	16673	2167	108330	205623	10281164
75	55	4131	172	12881	754	56526	31021	2326588
100	46	4574	112	11187	376	37592	8727	872651
150	37	5544	66	9917	156	23361	1664	249593
<b>186</b>	34	6305	<b>52</b>	<b>9740</b>	103	19250	746	138778
200	33	6614	49	9761	91	18230	577	115486
250	31	7769	40	10107	64	16017	276	69091
300	30	8994	36	10745	50	15097	161	48422
<b>347</b>	29	10194	33	11528	<b>43</b>	<b>14876</b>	110	38208
350	29	10272	33	11583	43	14876	108	37729
400	29	11592	31	12566	38	15090	79	31682
450	29	12943	30	13660	35	15601	62	28103
500	29	14316	30	14837	33	16327	52	25973
550	29	15704	29	16079	31	17215	45	24765
600	29	17103	29	17371	30	18230	40	24184
<b>643</b>	28	18313	29	18511	30	19182	<b>37</b>	<b>24047</b>
650	28	18510	29	18698	30	19343	37	24050

An additional factor to consider is that HTS is, by its nature, an intrinsically noisy process which gives results which are valid statistically but not necessarily at the individual compound level. Where an initial HTS hit compound has no analogues in the screening collection it might therefore be missed by the HTS process and incorrectly classified as inactive. This loss of this false negative will never be known. Where several analogues of the “missed-active” compound are also in the screening collection, it is unlikely that all of them will be classified inactive.

Discussions about compound acquisition often centre around “diversity” and the related idea of “covering chemical space”. However, these concepts are generally rather vague. For example, any particular screening collection can only cover an infinitesimal part of all possible chemical space and so it is difficult to evaluate the merit of trying to cover as much as possible. The granularity of biological activity with respect to small molecule structure is quite fine, i.e. small changes in the ligand can result in large changes in

binding affinity. Martin *et al.* [29] have shown that only about 30% of compounds with a Tanimoto  $>0.85$  of an active are active themselves. Therefore inputs of two diverse screening collections with one more “diverse” than the other are both likely to have the same overall hit rate and produce similar numbers of singleton hits. So diversity on its own is not a useful driver for acquisition decisions. We have instead focused on outcomes from the HTS campaign and the usefulness of these outcomes for the medicinal chemistry that must follow. There may be some differences of opinion amongst medicinal chemists as to the relative value of different initial outcomes but there will be general consensus about the goal. In this work, we have shown that analysis of a number of real-life hit-to-lead projects can provide useful information that relates the probability of success to the quality of the HTS inputs that went into the project. We expect that a similar analysis of wider datasets from HTS-to-lead campaigns will inform and increase the usefulness of the HTS process as it matures.



**Fig. (2).** Comparison between size of screening collection and optimal size of library. (A) “easy target”:  $P_s = 0.15$  and  $P_r = 0.05$  and (B) “hard target”:  $P_s = 0.10$  and  $P_r = 0.01$  for 95% probability of obtaining the desired outcome. The crosses show the optima in the sense that this is the smallest screening collection size that will give 95% confidence of obtaining the required hit series.

**Table 7.** Numbers of Compounds that Need to be Screened to Get 1, 2, 3 or 5 Hits as Singletons by Screening Diverse Compounds or as Series by Screening a Collection Composed of Analogue Blocks. In Each Case the Optimal Screening Collection Composition is Assumed (Highlighted Rows in Tables 5 and 6. The Total Numbers of Hits Obtained by Screening the Analogue Blocks will be Greater than Shown Here but these Extra Hits will Fall in Series that are Below the SAR Criteria

No. of Hits	Number of Compounds that Need to be Screened					
	Easy Target			Hard Target		
	As Singletons, by Screening Diverse Compounds	As Series, by Screening Analogue Blocks	Inflation	As Singletons, by Screening Diverse Compounds	As Series, by Screening Analogue Blocks	Inflation
1	398	N/A	N/A	2994	N/A	N/A
2	631	1267	2.0	4742	9740	2.1
3	837	1930	2.3	6294	14867	2.4
5	1218	3115	2.6	9151	24047	2.6

## APPENDIX 1

Calculating the size of a screening collection when it is composed of combinatorial chemical arrays and there is a requirement for at least one array to yield a hit series of at least  $k$  compounds.

Suppose a series has 'hit potential' i.e. the hit rate is greater than some minimum  $P_r$ . A conservative estimate for the number of hits can be represented by a Binomial distribution with probability  $P_r$ .

$$X \sim B(n, P_r)$$

where  $n$  is the number of compounds in the series screened.

The probability of  $k$  or more hits is

$$P_k = P(X \geq k)$$

which can be calculated as

$$P_k = 1 - \sum_{i=0}^{k-1} P(X = i) = 1 - \sum_{i=0}^{k-1} \binom{n}{i} P_r^i (1 - P_r)^{n-i}$$

We want to find the probability of success when screening  $N$  series. Each series has a probability  $P_s$  of having 'hit potential' and if it has hit potential a probability  $P_k$  of returning a 'success' (of  $k$  or more hits). Hence the probability of a success when screening a single library is  $P_s P_k$ . so the number of successes obtained in the screening campaign is Binomially distributed

$$Y \sim B(N, P_s P_k)$$

In particular, the probability of at least one success is:

$$\mathbf{P} = 1 - (1 - P_s P_k)^N$$

From this we can calculate the number of chemotype arrays that need to be screened to get a confidence  $\mathbf{P}$  of at least one success:

$$N = \frac{\ln(1 - \mathbf{P})}{\ln(1 - P_s P_k)}$$

## APPENDIX II

Extending Nilakantan's  $k = 1$  series size model. Calculating the size of library required to have a confidence  $P_k$  of  $k$  or more hits for a given hit rate, when  $k > 1$ .

Suppose the hit rate of a library is  $P_r$  and the series size is  $n$ . The number of hits obtained will be binomially-distributed.

$$X \sim B(n, P_r)$$

An exact calculation for  $\mathbf{P}_k$  (the probability, or confidence, of  $k$  or more hits) is given by equation (3). We need to solve (3) for  $n$ . Statistical software packages that have been investigated do not have a function for this (in general the number of trials  $n$  is an input parameter to statistical routines). Equation (4) shows an iterative scheme that was found to converge rapidly to the required value for  $j = 0, 1, 2, \dots$  where  $n_0$  is an initial guess. The iteration gives a fractional value for  $n$  which is rounded to the nearest integer.

$$\mathbf{P}_k = P(X \geq k) = 1 - \sum_{i=0}^{k-1} P(X = i) = 1 - \sum_{i=0}^{k-1} \binom{n}{i} P_r^i (1 - P_r)^{n-i} \quad (3)$$

$$n_{j+1} = \frac{1}{\log(1 - P_r)} \left[ \log(1 - \mathbf{P}_k) - \log \left[ \sum_{i=0}^{k-1} \left( \frac{\prod_{s=0}^{i-1} (n_j - s)}{i!} \right) P_r^i (1 - P_r)^{k-1-i} \right] \right] + k - 1 \quad (4)$$

## REFERENCES

- [1] Gordon, E.M.; Barrett, R.W.; Dower, W.J.; Fodor, S.P.A.; Gallop, M.A. *J. Med. Chem.*, **1994**, *37*, 1385.
- [2] Terrett, N.K.; Gardner, M.; Gordon, D.W.; Kobylecki, R.J.; Steele, J. *Tetrahedron*, **1995**, *51*, 8135.
- [3] Fox, S.; Farr-Jones, S.; Yund, M.A. *J. Biomol. Screen.*, **1999**, *4*, 183-186.
- [4] Furka, A. *Drug Disc. Today*, **2002**, *7*, 1-4.
- [5] Wölcke, J.; Ullmann, D. *Drug Disc. Today*, **2001**, *6*, 637-641.
- [6] Ertl, P. *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, 374.
- [7] Leeson, P.D.; Davis, A.M.; Steele, J. *Drug Disc. Today*, **2004**, *1*(3), 189.
- [8] Rose, S. *Drug Discov. Today*, **2002**, *7*, 133.
- [9] Lowrie, J.F.; Delisle, R.K.; Hobbs, D.W.; Diller, D.J. *Comb. Chem. High Throughput Screen.*, **2004**, *7*, 495.
- [10] Lipinski C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. *Adv. Drug Delivery Rev.*, **1997**, *23*, 2-25
- [11] Lipinski, C.A. *J. Pharmacol. Toxicol. Methods*, **2001**, *44*, 235-249
- [12] Ajay, Bemis, G.W.; Murcko, M.A. *J. Med. Chem.*, **1999**, *42*, 4942.
- [13] Frimurer, T.M.; Bywater, R.; Nærum, L.; Nørskov Lauritsen, L.; Brunak, S. *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 1315.
- [14] Clark R.D.; Pickett S.D. *Drug Disc. Today*, **2000**, *5*, 49-58.
- [15] Walters W.P.; Murko M.A. *Adv. Drug Del. Rev.*, **2002**, *54*, 255-271.
- [16] Lee, M.S.; Nakanishi, H.; Kahn, M. *Curr. Opin. Drug Disc. Dev.*, **1999**, *2*, 332.
- [17] Hann, M.; Green, D. *Curr. Opinion Chem. Biol.*, **1999**, *3*, 379-383.
- [18] Chen, G.; Zheng, S.; Luo, X.; Shen, J.; Zhu, W.; Liu, H.; Gui, C.; Zhang, J.; Zheng, M.; Puah, C.M.; Chen, K.; Jiang, H. *J. Comb. Chem.*, **2005**, *7*, 398.
- [19] Jacoby, E.; Schuffenhauer, A.; Popov, M.; Azzaoui, K.; Havill, B.; Schopfer, U.; Engeloeh, C.; Stanek, J.; Acklin, P.; Rigollier, P.; Stoll, F.; Koch, G.; Meier, P.; Orain, D.; Giger, R.; Hinrichs, J.; Malagu, K.; Zimmermann, J.; Roth, H.-J. *Curr. Top. Med. Chem.*, **2005**, *5*, 397.
- [20] Dolle, R.E. *J. Comb. Chem.* **2003**, *5*, 693.
- [21] Nilakantan, R.; Immermann, F.; Haraki, K. *Comb. Chem. High Throughput Screen.*, **2002**, *5*, 105.
- [22] Harper, G.; Pickett, S.D.; Green, D.V.S. *Comb. Chem. High Throughput Screen.*, **2004**, *7*, 63-70
- [23] Birault, V.; Harris, C.J.; Le, J.; Lipkin, M.; Nerella, R.; Stevens, A.P. In *Proceedings of the 15<sup>th</sup> European Symposium on Structure-Activity Relationships (QSAR) and Molecular Modelling*, Aki (Şener) E. and Yalçın I. Eds., Computer Aided Drug Design & Development Society, Turkey. **2004**, pp. 324-326.
- [24] Birault, V.; Harris, C.J.; Le, J.; Lipkin, M.; Nerella, R.; Stevens, A.P. *Curr. Med. Chem.*, **2006**, *13*(15), 1735-1748.
- [25] Crossley, R.; Rose, V.S.; Stevens, A.P. Construction of libraries of compounds focused towards receptors and other biological targets for screening and design of drugs or agrochemicals. BioFocus plc. PCT Int. Appl. WO 03/004147 A2. **2003**.
- [26] Crossley R. *Curr. Top. Med. Chem.*, **2004**, *4*, 581-588.
- [27] Crossley R. and Slater M.J. In *CHEMOGENOMICS: Knowledge-based Approaches to Drug Discovery*, Jacoby, E. Ed., World Scientific Publishing. **2004**, pp. 85-108
- [28] Harris, C.J.; Stevens, A.P. *Drug Disc. Today*, **2006**, *11*, 880-888.
- [29] Martin, Y.C.; Kofron, J.L.; Traphagen, L.M. *J. Med. Chem.*, **2002**, *45*, 4350-4358.

Received: August 9, 2007

Revised: February 19, 2008

Accepted: March 8, 2008