

PD3 052

Synthetic combinatorial libraries: Views on techniques and their application

Dave Madden, Viktor Krchňák and Michal Lebl

Selectide Corporation, 1580 E. Hanley Boulevard, Tucson, AZ 85737, U.S.A.

Received 4 January 1995

Accepted 22 February 1995

Key words: Molecular diversity; Library techniques; One-bead-one-compound technique; Screening; Structure elucidation

SUMMARY

This paper reviews synthetic combinatorial libraries. Particular emphasis is placed on one-bead-one-compound libraries, although issues relating to all synthetic techniques are touched upon. The discussion is focused on questions relating to synthetic diversity, drug discovery assays and structure determination techniques.

INTRODUCTION

By now there is little argument over whether combinatorial libraries are a useful tool in the discovery of drug candidates. Indeed, there may even be no argument over whether combinatorial libraries are an *essential* part of any drug discovery effort. Instead, the discussion has become more one of differences in emphasis and implementation. Ultimately, the debates will be settled by empirical evidence; but in the meantime, there is ample room for philosophical differences.

There are two principal types of synthetic combinatorial libraries (for a review of library techniques see e.g. Refs. 1-5): Lam's one-bead-one-compound ('Selectide') approach [6] and Houghten's SCL (synthetic combinatorial libraries) and Geysen's iterative approach [7,8]. The one-bead-one-compound lead discovery process is characterized by three main steps: (i) synthesis of a library using the split-mix method [6,7,9]; (ii) screening the library either using an on-bead assay [10] or releasing the compound from the bead and testing the released compounds in solution [11,12]; and (iii) determination of the structure of compounds of interest. The main feature of this process is that each bead in the library contains only one chemical entity, and the identity of compounds is not tracked during the synthesis, but is determined only after the compound has been shown to possess desired properties. On the other hand, Houghten's SCL are synthesized according to a predetermined synthetic strategy, where each sublibrary bears a

characteristic feature (e.g., a certain amino acid at a certain position) and once the sublibrary containing compound(s) with desired properties is identified, a new set of sublibraries containing this feature is synthesized and screened until all building blocks are determined. Other modifications of this technique have been described (see, for example, Refs. 13 and 14).

This review attempts to expose differing viewpoints on the three key components of a combinatorial chemistry program – synthetic diversity, drug discovery assays and structure identification.

SYNTHETIC DIVERSITY

When the synthesis of combinatorial libraries was limited to oligomeric compounds, such as peptides and oligonucleotides, there was little debate over what constitutes diversity. When using a finite number of building blocks for the construction of libraries, the answer was simple – bigger is better, at least in order to increase the probability of discovering a hit. With the advent of non-oligomeric libraries, the question of diversity comes front and center to the debate. What constitutes diversity? Is it measurable by any known objective standard? Is it really that important relative to other considerations? The question of diversity relates to the probability of success in discovering a lead compound. However, in the design of new libraries one must consider other factors relating to the probability of discovering a viable drug candidate, for which factors such as pharmacokinetic properties, toxicity and cost are all relevant additional factors. The principal challenge facing those employing combinatorial chemistry for drug discovery is the solution to a simultaneous equation, the result of which will be the optimal allocation of resources between library synthesis and traditional medicinal chemistry to be performed on discovered leads. One's perspective on this is likely to depend on a host of factors unrelated to absolutes, such as where a research group's expertise lies and the quantity and character of total resources available.

Complexity and diversity in combinatorial libraries Each combinatorial library is characterized by its complexity and diversity. Complexity is represented by the number of possible compounds and is easily quantified. For example, a pentapeptide library, synthesized using the 20 natural amino acids, can contain 20^5 individual peptides. Diversity reflects the dissimilarity of compounds within a given library, and there is no simple quantitative measure of this property. Peptides represent rather limited diversity. The only changing feature of peptides is a side chain attached to the α -carbon of a repetitive backbone. Increasing the number of α -amino acids used in library synthesis does not greatly increase its diversity; it simply maps the accessible conformational space more densely. As an example, no appreciable increase in diversity results from addition of one methylene group to the side chain. If the methylene group is instead inserted into the backbone, shifting the amino group from the α - to the β -position, this has a dramatic effect on diversity, since it influences the spacing of all side chains separated by this insert (Fig. 1). From this point of view, a library built with the same chemistry as peptide libraries, but combining α , β , γ , etc. amino acids, creates a library with appreciable diversity.

A key aspect in the generation of diversity is the selection of building blocks. This selection may depend on the available information regarding structural requirements for a small ligand or whether no structure–activity information is known. The selection is often accomplished based on the intuition of the investigator; however, several groups have described computer-assisted selection, such as Chiron's flower plots [15]. Regardless of the process, selection criteria should

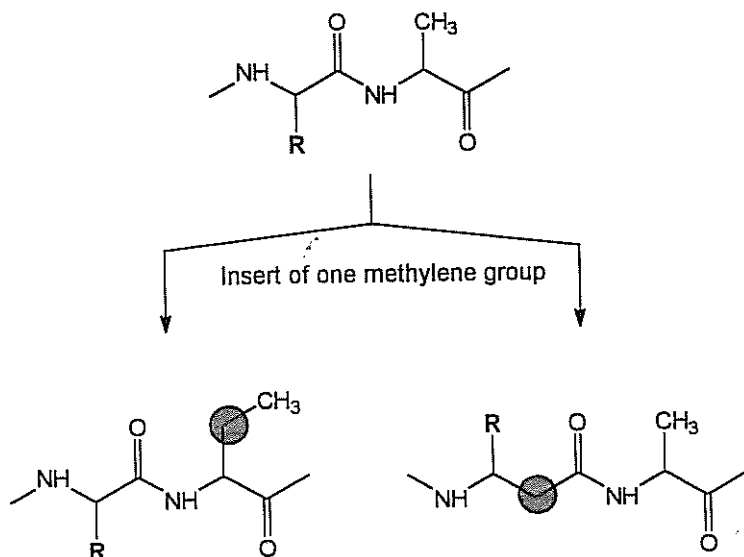


Fig. 1. Different effects of insertion of one methylene group into a peptide backbone or an amino acid side chain.

be based on the possibility for different types of interaction. All interactions known to play a critical role in the binding of two molecules should be considered. These include hydrophobic, aliphatic, aromatic and charged interactions (i.e., building blocks should contain positively and negatively charged groups). In addition, hydrogen bond formation (i.e., the presence of both a hydrogen donor and acceptor) and chelating groups should also be available in the library.

The different strategies in selection of building blocks can be illustrated by Fig. 2. Figure 2A shows structures of amines used for peptoid library synthesis [16]. However, this library was synthesized especially for screening against the α -adrenergic receptor. The selection of carboxylic acids for a scaffold-based library for trypsin inhibitors is illustrated in Fig. 2B [17]. The selection of carboxylic acids for a generic library, based on a computer algorithm favoring the most dissimilar structures, is depicted in Fig. 2C [18]. Intuitive selection of carboxylic acids for a generic library is illustrated in Fig. 2D. This selection is similar to that used in a small model library of alkylated and acylated amino acids, in which ligands for streptavidin were identified [19].

The enormous wave of effort now concentrated on the development of non-peptide combinatorial libraries* belies the question of whether one can reasonably expect to identify active lead compounds from a screen of random organic chemicals, in some cases bearing no relation to what might be found in nature. In this respect, the biotechnology and pharmaceutical industries have come full circle – almost. Most early drugs were derived from the screening of natural products, and the large (small by combinatorial standards) libraries of chemical compounds now harbored by the large pharmaceutical companies resulted from the synthesis of numerous analogs

*Non-peptidic libraries can be classified into two groups [20]: (i) scaffold-based libraries, in which a small scaffold (e.g. cyclopentane, cyclohexane, benzene) contains three to four functional groups to which different sets of building blocks can be attached. This approach is based on Hirschman's scaffolded mimics of somatostatin [21,22]; and (ii) libraries based on a combination of sequential chemical reactions, which result in a randomized molecule, which may or may not resemble a scaffold-based compound [23,24].

of these initially discovered compounds. As a result, these libraries are anything but random and may be anything but diverse. Is that good or bad? The answer to this question will no doubt be found over the next five to ten years, as one measures the discovery efficiency of combinatorial chemistry against the historical rate of the drug industry. In the meantime, though, it is prudent not to completely discard traditional techniques, and to combine these with combinatorial chemistry, at least until the results are in. The pharmaceutical industry is replete with examples of ideology overcoming pragmatism, with disastrous effects. For instance, the boom in biotechnology was founded in part on the hypothesis, now in many cases disproved, that by employing natural proteins as drugs one could obtain a higher success rate in drug development. Now disillusioned with the early promise of biotechnology drugs, many biotechnology companies are returning to the process of screening. Much can be said also about the crow now routinely eaten by the advocates of rational drug design as a stand-alone drug discovery tool. We, of course, are sold on the enormous value to be reaped with combinatorial chemistry, but not to the exclusion of all else.

The foregoing preamble leads to the question of what one can reasonably expect to obtain through the screening of combinatorial libraries – drugs or early leads? In our view, primary library screening will yield a number of different hits, in some cases with numerous different structural characteristics. Based on the results of primary screens and in combination with

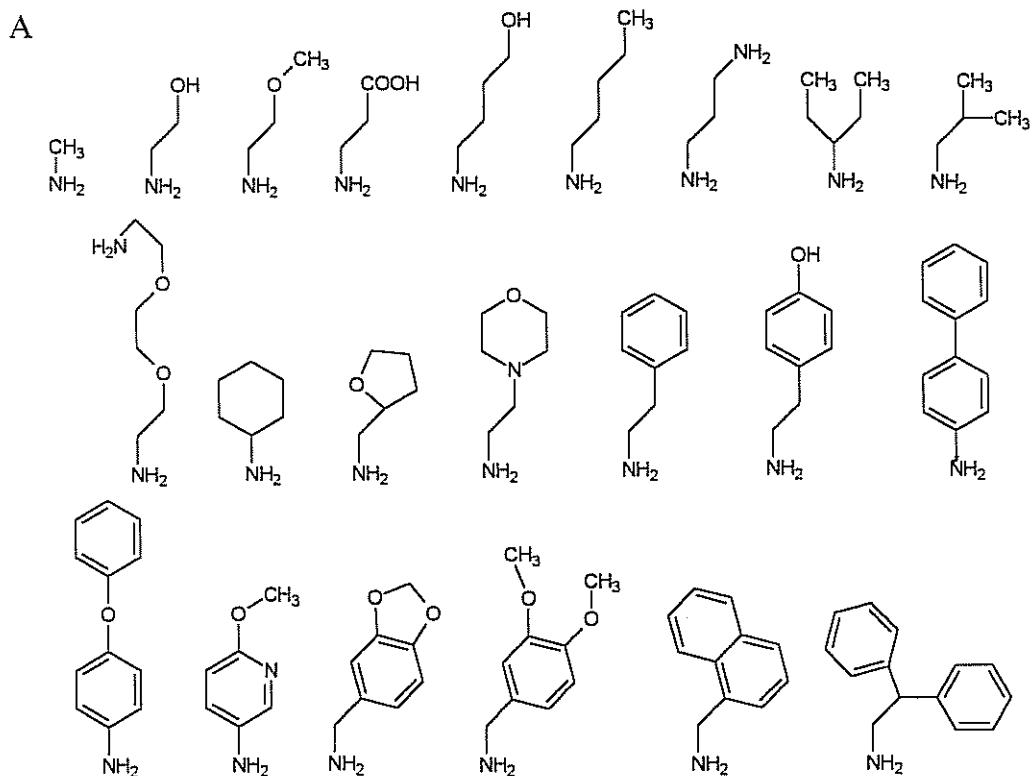


Fig. 2. (A) Building blocks used for diversity generation: Amines used in the building of a library dedicated to screening of α -adrenergic receptor antagonist [16].

computational approximation methods, one can design a number of secondary libraries casting a finer net across the sea of energy minima, with the expectation that one could identify one or several compounds with greater activity by several orders of magnitude. In this effort we have been successful on at least one project (see Table 1). At some point, however, the limitations of the solid phase, and the diminishing ability to produce libraries of increasingly finer variation, cause one to resort to traditional one-at-a-time synthesis – although at this point one would presumably be closer to a drug than if starting from a hit in a natural product screen.

B

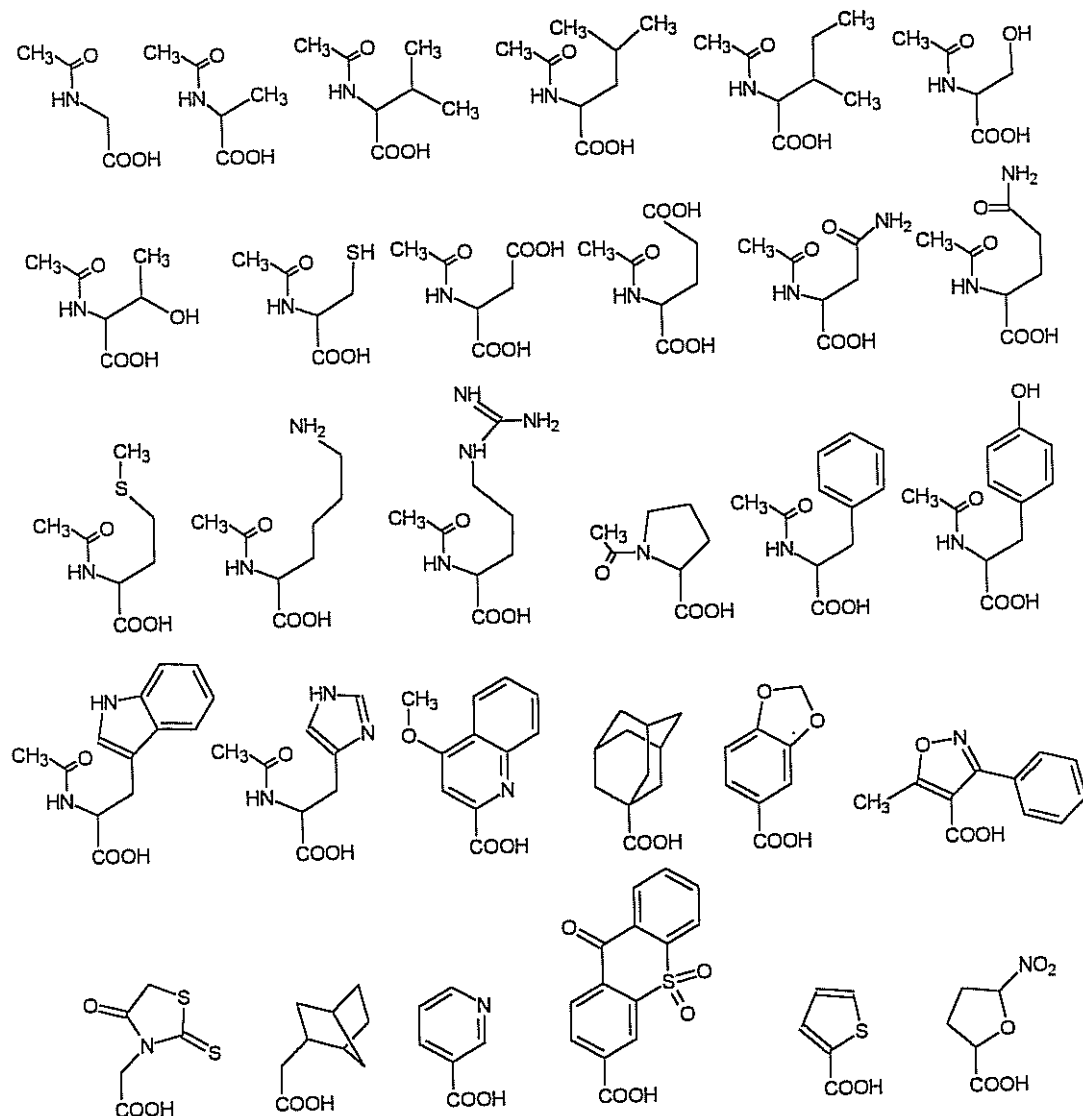


Fig. 2. (B) Building blocks used for diversity generation: Carboxylic acids used in the construction of a template library in which chymotrypsin inhibitors were identified [17].

This stepwise approach to combinatorial drug discovery is at odds with the theory of some who advocate the screening of many millions of compounds at a time. This advocacy generally stems from the expectation, more often desire, to select a drug candidate from a library. In our view, however, one is then looking for a very low probability event. Low enough, in fact, to easily dwarf a library of many millions of compounds. One might instead consider the screening of successively finer grids, first locating the general area of activity, and further refining the structure through a series of combinatorial and traditional techniques. Taken this way, one can initially screen smaller libraries, thus shortening library development and synthetic time and simplifying structure elucidation.

A related question arises when one considers the philosophy of library design. What portion of a library must one screen in order to be reasonably confident that, if a hit exists within the library's grid, it is susceptible to identification in a screening assay? In our view, the answer lies not only in statistics, but also in the question of the similarity of library building blocks and the number of critical residues or minimal common surface required for binding to a given target.

The practical limitation on the size of a library prepared using the one-bead-one-compound

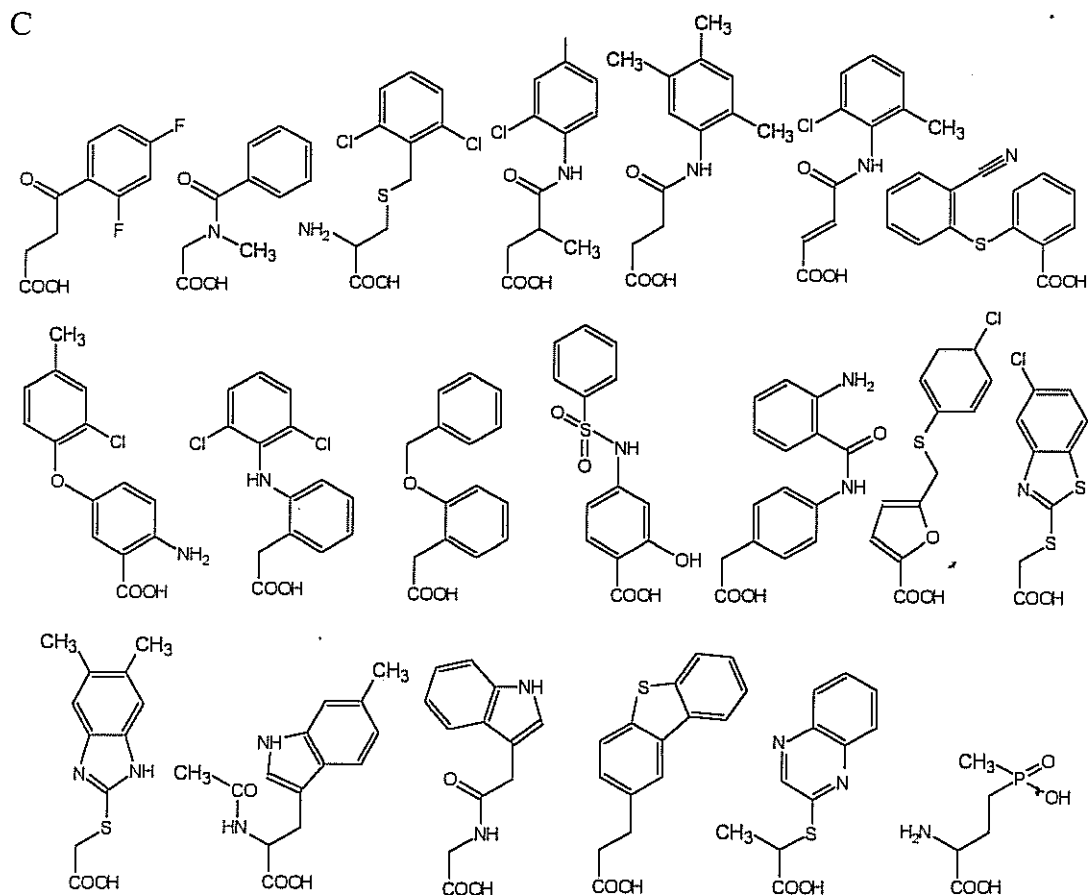


Fig. 2. (C) Building blocks used for diversity generation: Diverse building blocks identified by a computer algorithm [18].

technology is several million beads, and somewhat fewer different compounds. This limitation results from the ability to manipulate solid-phase supports. For example, a library of hexapeptides will contain 64 million different peptides, if only 20 natural amino acids are used for randomization. Increasing the number of building blocks (amino acids) increases the number of compounds exponentially. To synthesize a library in which 95% of possible species are represented with 99% of confidence, one needs more than threefold excess of beads over the number of possible species if the total number of species is greater than 10^5 [25]. For smaller libraries, the required redundancy is greater. To synthesize a decapeptide library of the format described above and to satisfy the confidence level outlined, one would need more than 10 000 kg of resin.

This is hardly a limitation in our view, since one need not screen a complete library to adequately map the relevant conformational space. It is more realistic to expect a number of hits for the given macromolecular acceptor which share similar or analogous structural features. In other words, the motif required for binding or any other biological function is more important than the individual molecule, since not all building blocks are equally important and do not contribute

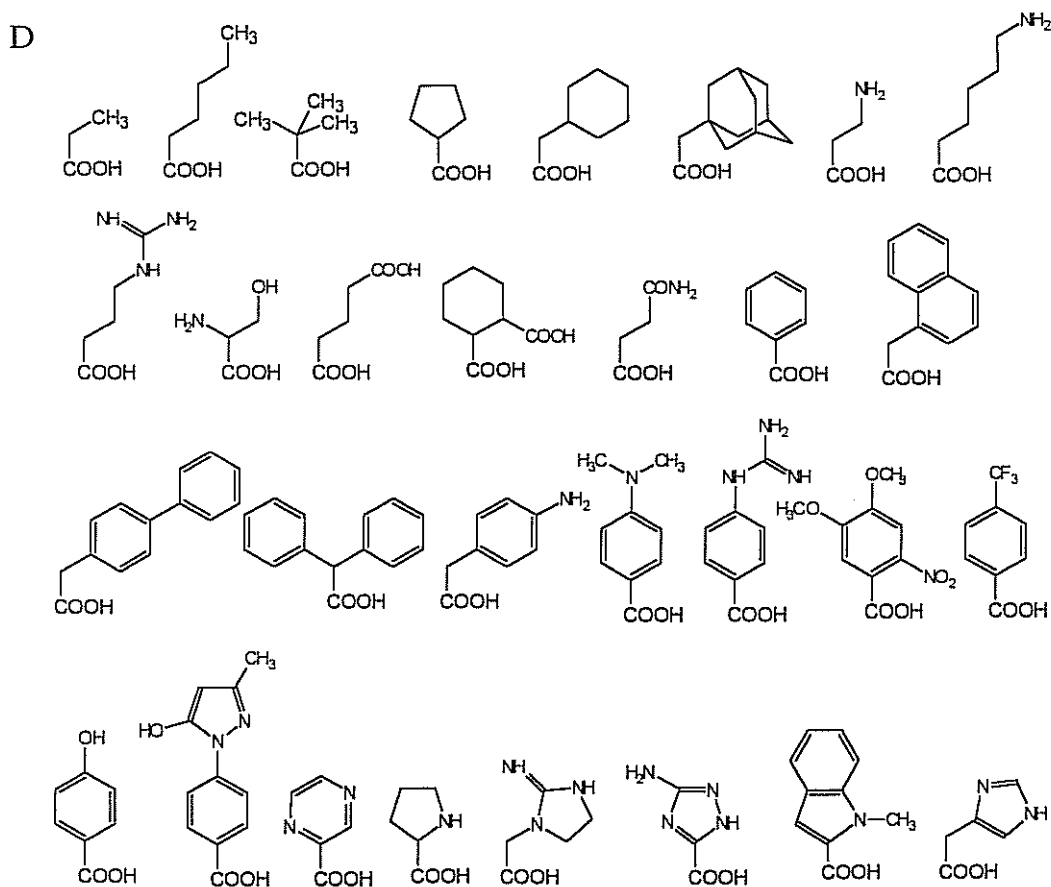


Fig. 2. (D) Building blocks used for diversity generation: A minimal set of carboxylic acids used for generic libraries at Selectide.

TABLE 1
 EXAMPLES OF THE RESULTS OBTAINED IN SCREENING AND LEAD OPTIMIZATION IN SEVERAL PROJECTS [5]

Project	Primary lead (μM)	Optimized lead
gpIIb/IIIa antagonist	1–320	NA
Thrombin inhibitor	0.8–700	25 nM
Factor Xa inhibitor	15	300 pM
HER-2 ligand	40	70 nM
IL-8 ligand	5	NA
NADPH oxidase inhibitor	5	NA
HIV-1 RNase inhibitor	0.3–200	150 nM

NA = not available.

equally to binding. When viewed in this way, one should not exclude building blocks to enable the synthesis of a complete library, since a compound need not be optimally composed to identify a primary lead; however, the most significant features must be present. In this respect, we favor the inclusion of a larger number of building blocks, even though this may reduce the representation of compounds in a library when viewed against the theoretical possibilities.

For example, in a peptide molecule we can identify the critical residues – those which cannot be replaced without significant loss of activity – and the noncritical residues, which serve primarily as structural units displaying the residues critical for interaction. The number of expected hits from the peptide library depends on the number of critical residues in the binding structure, the number of beads screened, and the number of amino acids randomized in each position, but is independent of the length of the library constituents or the total number of permutations in the library. It can be calculated by the formula:

$$n = x P_f [S / (A_n)^{n_{\text{crit}}}]$$

In this equation n is the number of expected positive hits, x is the number of different binding motifs, P_f is the 'placement' factor, i.e., the number of possible placements of each motif in the peptidic chain, S is the number of beads screened, A_n is the number of amino acids (subunits) used for randomization and n_{crit} is the number of critical residues.

To take advantage of this observation, we have developed a method for synthesizing a library of motifs. In this method, the split synthesis is combined with the coupling of mixtures of amino acids using the algorithm described in Fig. 3. The resulting beads carry the structures in which a given number of positions within the sequence are defined, representing a motif (in the example shown, a motif composed of three amino acids), and the remaining positions are occupied by a mixture of amino acids. Each bead thus contains AA^n (AA = number of amino acids in the mixture used for coupling, n = number of positions in which the mixture was coupled) different peptides with one defined motif. For example, a tripeptide motif can be arranged in 20 different ways in the molecule of a hexapeptide (e.g., ABCXXX, XABCXX, XAXXBC, ..., where A, B and C represent defined amino acid residues and X represents a mixture of amino acids). Since there are 8000 (20^3) tripeptide motifs composed of 20 amino acids, the full representation of a library of tripeptide motifs in hexapeptide libraries using 20 amino acids for randomization and 20 amino acids in coupling mixtures ($(6_{20}(3_{20})\text{LofL}$) will be composed of 160 000 motifs, each

represented on a bead, compared to a complete library of hexapeptides, which would be composed of 64 000 000 compounds. Results from this library format have recently been published [26,27].

ASSAYS

Turning to the debate over assay techniques, there is more evidence to suggest a path forward. What is the best screening technique: on bead, in solution, or some variation of both? We have taken the pragmatic approach at Selectide, having reduced both methods to practice successfully and having created a hybrid approach which utilizes most of the good characteristics from both techniques.

On-bead screening

The arguments in favor of the on-bead screening technique relate primarily to the speed with which one can assay a large library of compounds. Using the on-bead histochemical approach or an automated approach utilizing a fluorescently labelled probe and a fluorescence-activated cell sorter (FACS), one can assay millions of beads in a period of a few hours. The second advantage relates to the ability to use the library beads themselves as a technique to separate active compounds from inactive ones. In maintaining the relationship between one compound and

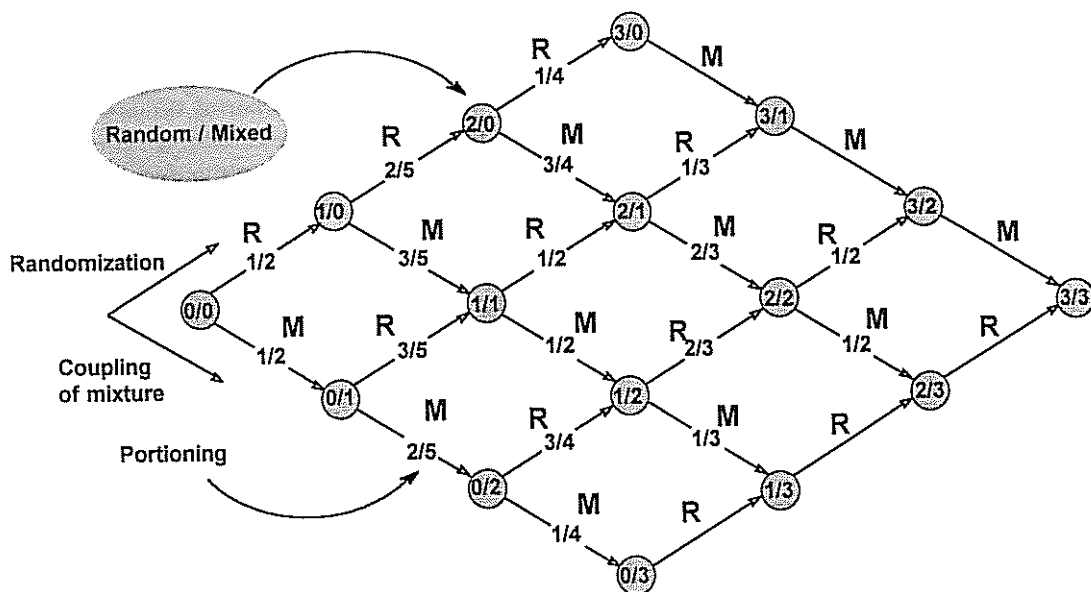


Fig. 3. Scheme of the synthesis of a 'library of libraries' with a tripeptide motif in a hexapeptide frame [27]. The polymeric carrier is split in the ratio $n_{Rr} : n_{Mr}$ (numbers located on the lines connecting circles denote the fraction of carrier undergoing the specified operation; n_{Rr} = number of remaining randomization steps; n_{Mr} = number of remaining steps in which a mixture of amino acids is to be coupled), based on the status n_r/n_m (number located in the shaded circles; n_r = number of randomizations performed on the carrier; n_m = number of mixtures to be coupled). At the beginning of the synthesis $n_r/n_m = 0/0$, at the end $n_r/n_m = 3/3$. The lines pointing up designate a randomization step, the lines pointing down designate a mixture step. This scheme can be applied to any length of peptide and motif.

its corresponding bead, one can use a host of techniques for structure identification, including Edman degradation and the polymerase chain reaction (PCR) for peptides and oligonucleotides, or techniques in which the bead can be tagged with a compound from which the synthetic history of a non-oligomeric bead can be discerned. A final advantage relates to economics. Using the on-bead approach, one need only synthesize small quantities of each compound on the bead. Since the concentration of compounds visible to the target of interest is quite high, one can still identify low-affinity ligands. The second component of the economic advantage relates to the ability to reuse libraries. Once a library has been screened, it can easily be treated and reused, lacking only those compounds selected in the previous assays. In our experience, libraries can be successfully used up to 20 times.

Lam's histochemical procedure [6] has been successfully applied by many laboratories. Bead-bound peptide libraries were screened against acceptor molecules (e.g., a natural or artificial receptor, enzyme, antibody, or even a small molecule) using an ELISA-type assay [6,10,28,29]. This approach was subsequently expanded to include a fluorescence-based assay using, for example, FACS [30], fluorescence microscopy [31,32], measurement of fluorescence in solution [33], radio ligand binding [34,35] or magnetic bead binding [36]. For all its advantages, the very simple and fast on-bead approach has its limitations, including the possibility of nonspecific interaction (highly charged peptides, very hydrophobic molecules, etc.) and, given the high off-rate of some receptors, the inability to observe binding. To eliminate unwanted interactions, each bead identified as positive in the first screen must undergo specificity testing. An example of the importance of specificity testing is illustrated in Fig. 4. One million beads from a pentapeptide library were incubated with the gpIIb/IIIa receptor and 7480 positive beads were detected. These beads were stripped of the coloration and reincubated in the presence of a high concentration of the specific ligand for this receptor, G4120 [37]. Nonstained beads in this step are specific, but they were stripped again and reincubated in the presence of a low concentration of the competitor to define the highest affinity binders. Only eight beads were selected in this way and five of them contained the expected sequence Arg-Gly-Asp [5].

Based simply on empirical evidence, the on-bead technique seems to work better in our hands for some proteins (thrombin, factor Xa) than for others (factor VII, IL-1). The failure to identify hits when the library contains those compounds known to interact may relate to some form of steric hindrance created by the compound's attachment to the bead. It may also relate to the orientation of the compound on the bead – even though we and others have developed techniques for varying this orientation [38–40]. Other effects, such as bead surface characteristics, may mask positive compounds and disable them from interacting with the target protein. A variety of effects may also create false positives. One can imagine proximity effects that would create dimers and trimers or other intermolecular effects that would cause one's assay to identify interactions where none exists in monomeric compounds. There is also the possibility of nonspecific interaction between the target and the bead itself, even though the typical bead's healthy constituent of polyethylene glycol would tend to discourage such interactions.

In-solution screening

Many pharmacologists argue that the only effective approach to high-throughput screening is in solution. Given the technical ability to achieve equal throughput, ease of handling and a comparable cost, one can hardly argue with this statement. In our experience, however, screening

in solution requires trade-offs, just as on-bead screening does. However, based on some advances in the basic technology, these trade-offs are becoming less burdensome, facilitating the in-solution screening of combinatorial libraries (see below). Perhaps the most obvious advantage of in-solution screening is the ability to use well-established assays in a primary screen. Unlike on-bead screening, which requires a soluble protein to perform a primary assay, in-solution screening is more flexible, enabling one to screen with membrane-bound proteins, whole cells, or a host of other functional or signal transduction assays. One can even cascade these assays, so that any selected compound can undergo a battery of confirmatory assays, even before the structure of the compound has been identified and before one has invested in the resynthesis of the library compound. Unless one uses an iterative approach to the synthesis and screening of libraries, the principal limitation in this process has been the quantity of compound yielded from a single synthetic bead. Using 80 μm tentagel beads, one can expect the reliable production of tens of picomoles of compound on a single bead. This is hardly enough to perform a single screening assay at low concentration (100 nM), let alone for duplication and confirmation. Technical advances in the synthetic beads have overcome this problem and we have had routine success in yielding nanomoles of compound from a single bead, increasing assay flexibility significantly. The second technical hurdle to be approached for in-solution screening is the design and synthesis of linkers. These must be compatible with the synthetic process as well as cleavable under conditions

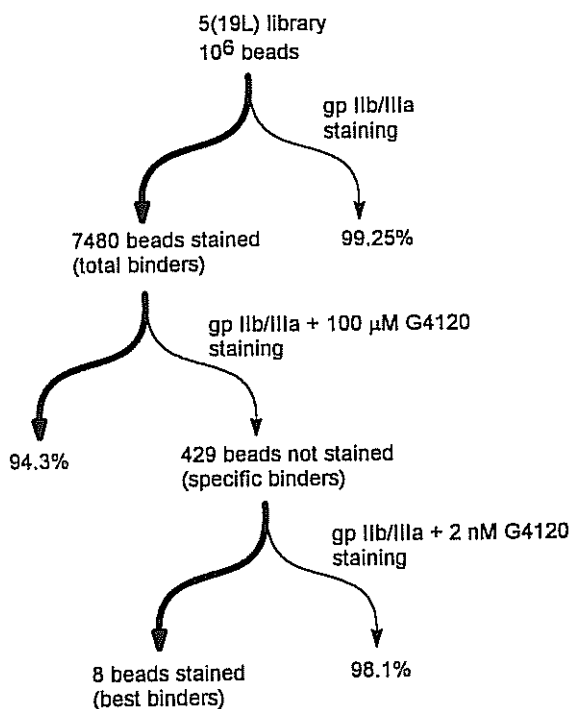


Fig. 4. Scheme of screening with elimination of nonspecific and weak ligands (this example was taken from gpIIb/IIIa receptor ligands screening). First step: incubation of the library with labelled receptor; second step: reincubation of the regenerated (stripped) beads selected in the first step with a specific inhibitor to eliminate nonspecific interactions – negative beads are selected for the last step; third step: reincubation of stripped beads selected in the second step with a lower concentration of inhibitor to select the most active ligands. Thick lines denote staining.

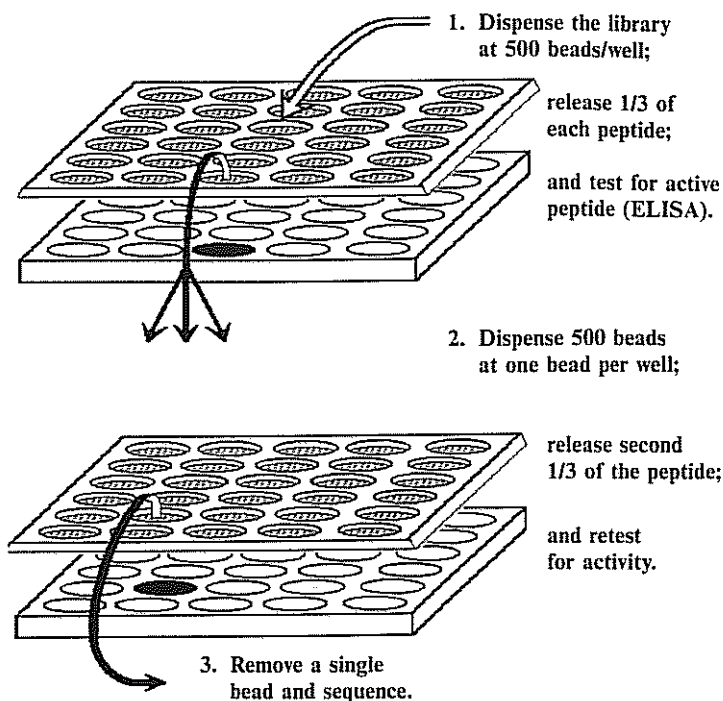


Fig. 5. Scheme illustrating library screening using a double cleavable linker [11].

tolerated by the biological assay to be carried out. We have had to perform significant development to produce linkers away from the ester-based linkers we had earlier published. Despite these improvements, in-solution screening remains slow relative to on-bead screening. One can reasonably expect to screen several hundred thousand compounds in a few days. These assays are easily susceptible to automation, however, and one can anticipate increasing gains in area.

To accommodate one-bead-one-compound libraries for screening in solution, one needs a link between the released compound and its corresponding parent bead, which contains information regarding the structure of a released compound. Since it would be impractical to release and test these compounds one by one, we devised a simple method to test a number of compounds simultaneously. In our original releasable one-bead-one-compound peptide libraries, three copies of each peptide were synthesized on beads via specific linkers [11,41,42]. Two of these linkers contained an ester bond which allowed the release of the peptide into solution under different conditions. The first linker closed a six-membered dioxipiperazine ring at neutral pH and released the peptide into the solution; the second ester bond was cleaved at elevated pH. The third linker was an amide bond that held the 'informative' peptide for sequencing. The practical performance of a solution assay using one-bead-one-compound libraries is illustrated in Fig. 5.

Kinetically controlled two-stage release from one type of UV-sensitive linker was later used, for example, by Pharmacoepia [43]. An acid-sensitive linker, cleaved by exposure to gaseous trifluoroacetic acid, was used for kinetically controlled release of peptides from benzhydrylamine resin [12].

In general, two-stage release can be accomplished by several different means. When the compounds are connected to beads by two different linkers, both releases are unambiguous; however, this represents more demanding chemistry. If a safety catch mechanism is applied in linker design, e.g., thermodynamically favored ring closure by the attack of an internal nucleophile that is masked during library synthesis, the same mechanism can be used for both releases: to trigger the release one needs to remove the protecting group. Two different protecting groups will accommodate two independent releases (Fig. 6). In this arrangement there is a need for cleavage of the second protecting group after the first release. Kinetic release in this type of linker can be accomplished either after total cleavage of the protecting group (in this case there is only one type of protecting group) or partial kinetically controlled cleavage of the protecting group, followed by total release from this part of the linker (Fig. 7).

Hybrid screening

In order to take advantage of the attributes of both techniques, we have developed a third alternative, called hybrid screening. In this case, we initially screen a library using the on-bead approach. After selecting positive beads in the primary screen, the test compound is cleaved from the bead and tested in duplicate in solution, either in a binding assay or in a functional assay. In this fashion one can estimate and confirm the activity of the compounds before identification of the structure and resynthesis of the compound.

STRUCTURE IDENTIFICATION

With the sea change in combinatorial chemistry away from oligomeric compounds for which the structure is easily determined using only minute quantities, structure identification has perhaps become the most important enabling aspect of combinatorial chemistry. Techniques developed and in development for structure identification necessarily take the form of recognition, rather than true structure determination, since techniques such as mass spectrometry and NMR spectroscopy generally require far larger quantities than one could routinely expect from a combinatorial synthesis. As a result, investigators have in general focused on techniques by which the synthetic history of a compound can be recorded contemporaneously with library synthesis. This can be achieved by coding (see below). These tags have taken the form of oligonucleotides, peptides and halogenated aromatic compounds. As different as these tagging molecules may

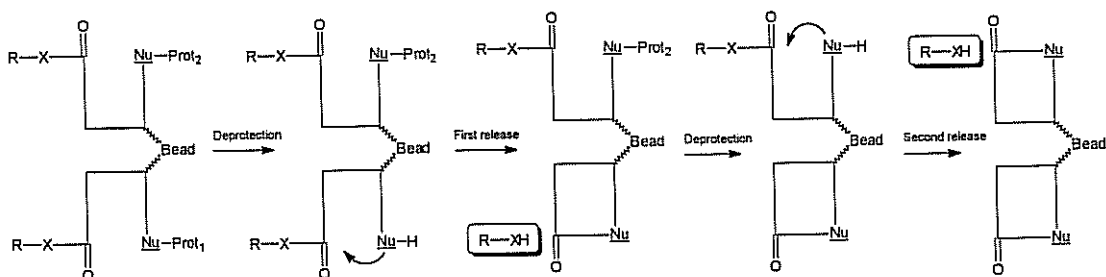


Fig. 6. Schematic representation of a double releasable linker, based on the principle of intramolecular participation of an orthogonally protected nucleophile.

appear, they share the same principal limitations, i.e., the requirement for orthogonal protecting groups and the need to survive a wide variety of (often harsh) conditions needed to perform the synthesis of the test compound. In our exploration of non-oligomeric reactions on the solid phase, this limitation has proved considerable and has caused us to explore a series of direct determination methods using mass spectrometry, often in combination with algorithms for minimizing the propensity for creating libraries in which there are many compounds having the same molecular weight, and by attempting to document consistent similarities in fragmentation patterns to provide an additional dimension in structure determination.

The concept of coding is derived from nature. Amino acid sequences of proteins and peptides are encoded by nucleic acids and nucleic acids serve as a source of information for the synthesis of the corresponding protein. The same principle of coding is used in biological (e.g. phage) libraries (for reviews see Refs. 4 and 44). The extension of the phage display concept, in which a peptide can be encoded by a nucleic acid, led to a publication suggesting this technique for synthetic libraries [45], which was later implemented by several groups [30,46]. The first synthetic peptide libraries did not require coding, since the peptide itself, in minute quantities (tens of picomoles), was sufficient for structure determination. Edman degradation, having been fine-tuned for years, serves this purpose well. This technique was then adapted to enable peptides to code for structures that are not sequenceable [47,48].

Coding for each building block by one amino acid and assembling the code in a linear manner, thus forming a peptide, has one principal disadvantage. One needs to run as many cycles of Ed-

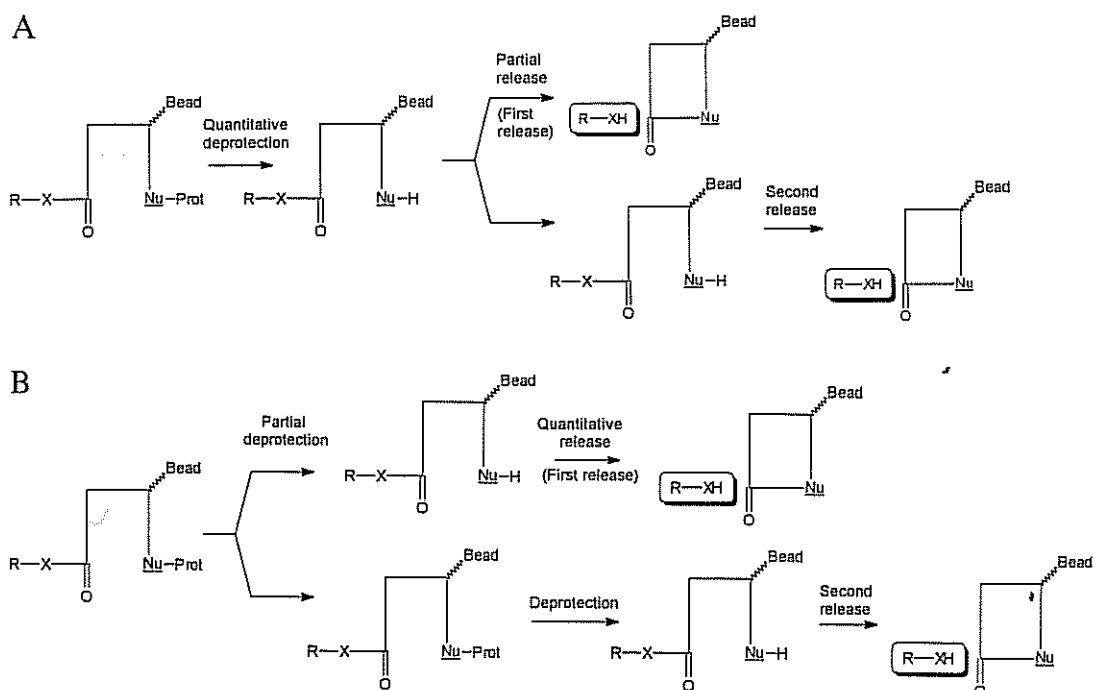


Fig. 7. Schematic representation of a double release assay based on kinetic release. (A) Quantitative deprotection, followed by partial release; (B) partial deprotection, followed by complete release.

man degradation as there are randomization steps performed in the library. One cycle takes close to 1 h, and yields little information. We and others have developed 'bar' coding techniques to speed decoding. In our technique, the code is assembled in such a way that one cycle of Edman degradation cleaves all coding amino acids. However, in this type of arrangement each randomization step requires a different set of coding amino acids, since the amino acids must represent the building block as well as the randomization cycle. For a library constructed with 20 building blocks in each of five randomizations it would require 100 different coding amino acids. However, when each building block is coded by a doublet of amino acids, n amino acids can code for $n*(n-1)/2$ building blocks. For the example described above, 15 coding amino acids are sufficient. We have made available a large number of coding amino acids by taking a diamino acid (e.g. lysine, ornithine) and acylating its side-chain amino group by a variety of acids. Once completed, one can select those that have a satisfactory retention time of their phenylhydantoins [5].

Another digital coding technique was described by Still and co-workers [29,49]. In this method each bead is stamped with a code in which reactive species attack the aromatic moieties of the bead's core structure, thus forming a covalently bound coding structure. The code is subsequently read using gas chromatography.

The difficulty in structure determination due to the small quantities of compound generally produced in a library format carries with it another question for which the answer is far less than clear. If the only way one can determine the structure of a compound is through indirect means, how can one determine that which has actually been screened? Since the chemistries now under development in many laboratories are far less efficient than those of peptides or oligonucleotides, one is almost certainly synthesizing more than one compound on the bead, and may be creating several stereoisomers if the intended compound has any chiral centers. For the foregoing reasons it is our view that, while indirect tagging techniques are in some cases useful, they are a far from universal solution to the non-peptide problem.

These difficulties, arising from techniques in which the identity of the active compound remains a mystery until selected and decoded, raise the question of the relative merits of such techniques versus methods for deconvoluting leads through successive assays. While these iterative approaches avoid the cumbersome tagging process, in the absence of a corroborating direct structure identification technique one is still faced with the question of how to know what compound has been synthesized and, consequently, screened assuming less than perfectly efficient chemistries. In this respect, one trades the synthetic difficulty of tagging approaches for a cumbersome and biased iterative screening algorithm.

Nevertheless, the structure determination question is not without hope for a near-term solution, since advances in screening techniques serve not only to speed and improve screening, they also enable the use of alternative means to determine the structure of active compounds. Since each compound selected is more likely to be a true active compound, one can afford to spend more time and resources on structure identification.

A LOOK FORWARD

What then is the future of combinatorial chemistry? In our view, combinatorial chemistry will continue to evolve in three obvious directions: increased diversity in library synthesis, assay innovation principally to increase throughput, and a movement toward direct structure determi-

nation techniques and alternative indirect techniques. Perhaps a more important development will be the increasing integration of combinatorial chemistry with other chemical techniques now largely in isolated use. In our view, the integration of combinatorial chemistry and computational chemistry – or rational drug design – will grow rapidly, and may provide the next big breakthrough in drug discovery, provided one can find a way to effectively manage the mass of data one can rapidly generate. As investigators look for direction in choosing libraries to synthesize and screen, they will increasingly look not only to intuition to select the next library to screen, or reach into the refrigerator to pull out that which is most readily available, but may access a database of possible structures and the results of earlier assays to perform computer-based speculations of possible libraries prior to their synthesis and screening.

ACKNOWLEDGEMENTS

The authors are indebted to all scientists at Selectide who dedicated their ideas and skills to prove Lam's idea in practice and to expand the library technique further.

REFERENCES

- Gallop, M.A., Barrett, R.W., Dower, W.J., Fodor, S.P.A. and Gordon, E.M., *J. Med. Chem.*, 37 (1994) 1233.
- Gordon, E.M., Barrett, R.W., Dower, W.J., Fodor, S.P.A. and Gallop, M.A., *J. Med. Chem.*, 37 (1994) 1385.
- Desai, M.C., Zuckermann, R.N. and Moos, W.H., *Drug Dev. Res.*, 33 (1994) 174.
- Lebl, J.K. and Craig, L., *Curr. Opin. Biotechnol.*, 5 (1994) 40.
- Lebl, M., Krchňák, V., Sepetov, N.F., Seligmann, B., Štrop, P., Felder, S. and Lam, K.S., *Biopolymers (Pept. Sci.)*, 37 (1995) 177.
- Lam, K.S., Salmon, S.E., Hersh, E.M., Hruby, V.J., Kazmierski, W.M. and Knapp, R.J., *Nature*, 354 (1991) 82.
- Houghten, R.A., Pinilla, C., Blondelle, S.E., Appel, J.R., Dooley, C.T. and Cuervo, J.H., *Nature*, 354 (1991) 84.
- Geysen, H.M., Rodda, S.J. and Mason, T.J., *Mol. Immunol.*, 23 (1986) 709.
- Furka, A., Sebastyén, F., Asgedom, M. and Dibó, G., *Int. J. Pept. Protein Res.*, 37 (1991) 487.
- Lam, K.S. and Lebl, M., *Immunomethods*, 1 (1992) 11.
- Salmon, S.E., Lam, K.S., Lebl, M., Kandola, A., Khattri, P.S., Wade, S., Pátek, M., Kočíš, P., Krchňák, V., Thorpe, D. and Felder, S., *Proc. Natl. Acad. Sci. USA*, 90 (1993) 11708.
- Jayawickreme, C.K., Graminski, G.F., Quillan, J.M. and Lerner, M.R., *Proc. Natl. Acad. Sci. USA*, 91 (1994) 1614.
- Carell, T., Wintner, E.A., Bashir-Hashemi, A. and Rebek, J.J., *Angew. Chem., Int. Ed. Engl.*, 33 (1994) 2059.
- Erb, E., Janda, K.D. and Brenner, S., *Proc. Natl. Acad. Sci. USA*, 91 (1994) 11422.
- Simon, R.J., Martin, E.J., Miller, S.M., Zuckermann, R.N., Blaney, J.M. and Moos, W.H., In Crabb, J.W. (Ed.) *Techniques in Protein Chemistry*, Vol. V, Academic Press, San Diego, CA, 1994, pp. 533–539.
- Zuckermann, R.N., Martin, E.J., Spellmeyer, D.C., Stauber, G.B., Shoemaker, K.R., Kerr, J.M., Figliozzi, G.M., Goff, D.A., Siani, M.A., Simon, R.J., Banville, S.C., Brown, E.G., Wang, L., Richter, L.S. and Moos, W.H., *J. Med. Chem.*, 37 (1994) 2678.
- Eichler, J., Lucka, A.W. and Houghten, R.A., *Pept. Res.*, 7 (1994) 300.
- Chapman, D., Poster presented at the symposium 'Chemical and Biomolecular Diversity', San Diego, CA, December 14–16, 1994.
- Staňková, M., Issakova, O., Sepetov, N.F., Krchňák, V., Lam, K.S. and Lebl, M., *Drug Dev. Res.*, 33 (1994) 146.
- Lebl, M., Krchňák, V., Šafář, P., Stierandová, A., Sepetov, N.F., Kočíš, P. and Lam, K.S., In Crabb, J.W. (Ed.) *Techniques in Protein Chemistry*, Vol. V, Academic Press, San Diego, CA, 1994, pp. 541–548.
- Hirschmann, R., Nicolaou, K.C., Pietranico, S., Salvino, J., Leahy, E.M., Sprengler, P.A., Furst, G., Smith III, A.B., Strader, C.D., Cascieri, M.A., Candelore, M.R., Donaldson, C., Vale, W. and Maechler, L., *J. Am. Chem. Soc.*, 114 (1992) 9217.

- 22 Hirschmann, R., Nicolaou, K.C., Pietranico, S., Leahy, E.M., Salvino, J., Arison, B., Cichy, M.A., Spoons, P.G., Shakespeare, W.C., Sprengeler, P.A., Hamley, P., Smith III, A.B., Reisine, T., Raynor, K., Maechler, L., Donaldson, C., Vale, W., Freidinger, R.M., Cascieri, M.R. and Strader, C.D., *J. Am. Chem. Soc.*, 115 (1993) 12550.
- 23 Bunin, B.A. and Ellman, J.A., *J. Am. Chem. Soc.*, 114 (1992) 10997.
- 24 Bunin, B.A., Plunkett, M.J. and Ellman, J.A., *Proc. Natl. Acad. Sci. USA*, 91 (1994) 4708.
- 25 Burgess, K., Liaw, A.I. and Wang, N., *J. Med. Chem.*, 37 (1994) 2985.
- 26 Sepetov, N.F., Krchňák, V., Staňková, M., Lam, K.S., Wade, S. and Lebl, M., In Maia, H.L.S. (Ed.) *Peptides 1994* (Proceedings of the 23rd European Peptide Symposium), ESCOM, Leiden, 1995, pp. 467–468.
- 27 Sepetov, N.F., Krchňák, V., Staňková, M., Wade, S., Lam, K.S. and Lebl, M., *Proc. Natl. Acad. Sci. USA*, (1995) in press.
- 28 Lam, K.S., Hruby, V.J., Lebl, M., Knapp, R.J., Kazmierski, W.M., Hersh, E.M. and Salmon, S.E., *Bioorg. Med. Chem. Lett.*, 3 (1993) 419.
- 29 Ohlmeyer, M.H.J., Swanson, R.N., Dillard, L.W., Reader, J.C., Asouline, G., Kobayashi, R., Wigler, M. and Still, W.C., *Proc. Natl. Acad. Sci. USA*, 90 (1993) 10922.
- 30 Needels, M.C., Jones, D.G., Tate, E.H., Heinkel, G.L., Kochersperger, L.M., Dower, W.J., Barrett, R.W. and Gallop, M.A., *Proc. Natl. Acad. Sci. USA*, 90 (1993) 10700.
- 31 Chen, J.K., Lane, W.S., Brauer, A.W., Tanaka, A. and Schreiber, S.L., *J. Am. Chem. Soc.*, 115 (1993) 12591.
- 32 Meldal, M., Svendsen, I., Breddam, K. and Auzanneau, F.I., *Proc. Natl. Acad. Sci. USA*, 91 (1994) 3314.
- 33 Ator, M.A., Beigel, S., Dankanich, T.C., Echols, M., Gainor, J.A., Gilliam, C.L., Gordon, T.D., Koch, D., Koch, J.F., Kruse, L.I., Morgan, B.A., Krupinski-Olsen, R., Siahaan, T.J., Singh, J. and Whipple, D.A., In Hodges, R.S. and Smith, J.A. (Eds.) *Peptides: Chemistry and Biology* (Proceedings of the 13th American Peptide Symposium), ESCOM, Leiden, 1994, pp. 1012–1015.
- 34 Kassarijian, A., Schellenberger, V. and Turck, C.W., *Pept. Res.*, 6 (1993) 129.
- 35 Wu, J., Ma, Q.N. and Lam, K.S., *Biochemistry*, 33 (1994) 14825.
- 36 Fassina, G., Bellitti, M.R. and Cassani, G., *Protein Pept. Lett.*, 1 (1994) 15.
- 37 Barker, P.L., Bullens, S., Bunting, S., Burdick, D.J., Chan, K.S., Deisher, T., Eigenbrot, C., Gadek, T.R., Gantzios, R., Lipari, M.T., Muir, C.D., Napier, M.A., Pitti, R.M., Padua, A., Quan, C., Stanley, M., Struble, M., Tom, J.Y.K. and Burnier, J.P., *J. Med. Chem.*, 35 (1992) 2040.
- 38 Lebl, M., Krchňák, V., Sepetov, N.F., Nikolaev, V., Stierandová, A., Šafář, P., Seligmann, B., Štrop, P., Thorpe, D., Felder, S., Lake, D.F., Lam, K.S. and Salmon, S.E., In Epton, R. (Ed.) *Innovation and Perspectives in Solid Phase Synthesis*, Mayflower Worldwide Ltd, Birmingham, 1994, pp. 233–238.
- 39 Holmes, C.P. and Rybak, C.M., In Hodges, R.S. and Smith, J.A. (Eds.) *Peptides: Chemistry and Biology* (Proceedings of the 13th American Peptide Symposium), ESCOM, Leiden, 1994, pp. 992–994.
- 40 Kania, R.S., Zuckermann, R.N. and Marlowe, C.K., *J. Am. Chem. Soc.*, 116 (1994) 8835.
- 41 Lebl, M., Pátek, M., Kočíš, P., Krchňák, V., Hruby, V.J., Salmon, S.E. and Lam, K.S., *Int. J. Pept. Protein Res.*, 41 (1993) 201.
- 42 Kočíš, P., Krchňák, V. and Lebl, M., *Tetrahedron Lett.*, 34 (1993) 7251.
- 43 Baldwin, J.J., Lecture presented at the symposium 'Chemical and Biomolecular Diversity', San Diego, CA, 1994.
- 44 Winter, J., *Drug Dev. Res.*, 33 (1994) 71.
- 45 Brenner, S. and Lerner, R.A., *Proc. Natl. Acad. Sci. USA*, 89 (1992) 5381.
- 46 Nielsen, J., Brenner, S. and Janda, K.D., *J. Am. Chem. Soc.*, 115 (1993) 9812.
- 47 Kerr, J.M., Banville, S.C. and Zuckermann, R.N., *J. Am. Chem. Soc.*, 115 (1993) 2529.
- 48 Nikolaev, V., Stierandová, A., Krchňák, V., Seligmann, B., Lam, K.S., Salmon, S.E. and Lebl, M., *Pept. Res.*, 6 (1993) 161.
- 49 Nestler, H.P., Bartlett, P.A. and Still, W.C., *J. Org. Chem.*, 59 (1994) 4723.