COMBINATORIAL COMPOUND LIBRARIES FOR DRUG DISCOVERY: AN ONGOING CHALLENGE

H. Mario Geysen, Frank Schoenen, David Wagner and Richard Wagner

Almost 20 years of combinatorial chemistry have emphasized the power of numbers, a key issue for drug discovery in the current genomic era, in which it has been estimated that there might be more than 10,000 potential targets for which it would be desirable to have small-molecule modulators. Combinatorial chemistry is best described as the industrialization of chemistry; the chemistry has not changed, just the way in which it is now carried out, which is principally by exploiting instrumentation and robotics coupled to the extensive use of computers to efficiently control the process and analyse the vast amounts of resulting data. Many researchers have contributed to the general concepts as well as to the technologies in present use. However, some interesting challenges still remain to be solved, and these are discussed here in the context of the application of combinatorial chemistry to drug discovery.

A GUIDE TO DRUG DISCOVERY

SOLID-PHASE CHEMISTRY
In solid-phase synthesis, the compounds being made are attached (usually by a linker group) to insoluble, functionalized, polymeric material (usually beads), allowing them to be readily separated (by filtration, centrifugation, and so on) from excess reagents, soluble reaction by-products or solvents.

Department of Chemistry, University of Virginia, McCormick Road, P.O. Box 400319, Charlottesville, Virginia, 22904-4319, USA.
Correspondence to H. M. G.
e-mail: geysen@virgina.edu
doi: 10.1038/nrd1035

In the decade since combinatorial chemistry was adopted by the pharmaceutical industry, it has undergone significant developments as accumulated individual experiences led to best practices. Initially, the focus was on the number of compounds produced, with little regard for their quality, in the belief that once hits were found, subsequent smaller rounds of more controlled synthesis would suffice to identify leads. This practice gave rise to many accounts of 'failures', where activity observed in assays was not reproducible. More problematic was the finding that the putative active compound — synthesized and characterized by conventional means — was completely inactive. As a result, numbers gave way to parallel synthesis of discrete compounds, often followed by purification to improve the reliability of the outcome.

So, in the last few years, the pharmaceutical industry has been sourcing an increasing proportion of compounds for screening by highly automated traditional solution-phase chemistry followed by high-pressure liquid chromatography/mass spectroscopy (HPLC/MS) purification. In an industry where research budgets are huge, the actual cost of a given unit operation, such as compound synthesis, is not of primary concern when compared with the need to obtain leads for the various targets under development. This movement away from solid-phase, numerically large, combinatorial library synthesis results from the current industry view, which can be summarized by the following statements: first, the quantitation of SOLID-PHASE CHEMISTRY, especially at the single-bead level, is very difficult and not generally applicable. Second, solid-phase chemistry is limiting with respect to the diversity of chemistries that can be carried out, and most of the successful solid-phase chemistries are very dependent on amide bond formation. The non-amide-bond chemistries that have been successfully enabled on solid-phase are few and so well explored that intellectual property issues now compromise their use by others. Third, ENCODING strategies associated with library synthesis are unreliable with respect to linking compound identity with assay outcome. MIX-AND-SPLIT SYNTHESIS STRATEGIES are precluded for lack of an acceptable encoding methodology, and require single-bead assay methods that are compromised by the

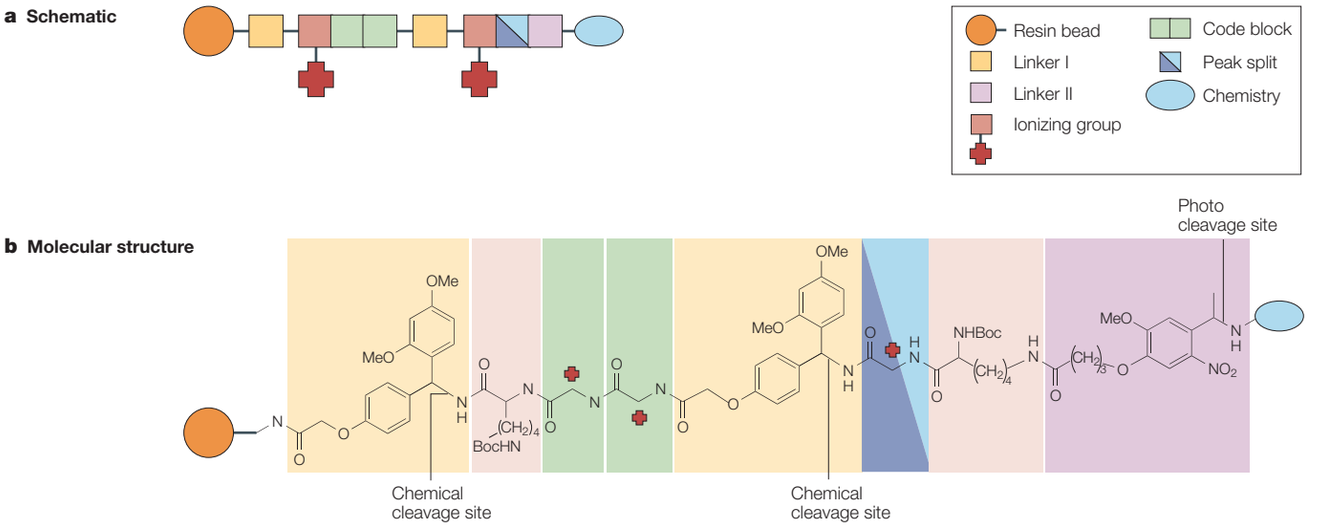

Figure 1 | **Analytical constructs for combinatorial chemistry.** One of the more commonly used constructs is shown. Assembly of the construct uses simple amide-bond formation chemistry with a coupling efficiency of close to 100% at incorporation. **a** | A block schematic of the construct shows that it comprises a linear format with three linker groups, two of which are chemically cleavable (yellow) and one that is photo-cleavable (lilac). In addition, there are two ionizable groups (red), a code block (green) and a peak-split signature element (light/dark blue). Both the code and the peak-split blocks are incorporated into the construct using isotopically (stable) modified reagents, which are described more fully in the main text. The positions indicated by the stars correspond to the locations where isotopically modified amino acids are incorporated, as described more fully in the main text. The ionizable group can be either positively or negatively charged, but is typically positive, as this charge mode usually results in a higher sensitivity when analysed by mass spectroscopy. In the most common implementation of the analytical construct, with acid-cleavable Rink linkers, the final charged group is an amine that is butoxycarbonyl (Boc) protected during the synthesis. This ensures that after chemical cleavage, with concurrent removal of the Boc protecting group, each of the two solution-phase fragments are positively charged and will therefore be observed in a mass spectrum. **b** | The molecular structure of the construct, with the cleavable bonds indicated.

quality of solid-phase synthesis. Finally, solid-phase chemistry should be restricted to the synthesis of discrete compounds (parallel synthesis) and include purification of each compound, with a target yield of 10 mg final purified compound.

In this article, each of the above issues will be addressed in terms of technological developments that might offer a potential solution.

Quantitation at the single-bead level

Rapid quantitation and identification (that is, determination of the molecular mass) of all the products resulting from chemistry carried out using resin-based solid-phase methodologies can be achieved using analytical constructs[1] (FIG. 1), which allow the complete monitoring of a library synthesis at each step of the procedure, and also provide a facile method of bead encoding. In addition to enabling systematic quality assessment throughout the library synthesis process, this technology revolutionized the way in which the optimization of solid-phase chemistry protocols and monomer scanning is carried out (see below).

Quantitation is achieved by using a simple single-quadrapole mass spectrometer, which typically has a throughput of about 20–30 samples per hour. This results in the preparation of high-quality characterized bead-based libraries, from which any synthesis failures have been removed prior to assay. At the completion of the synthesis of the analytical construct, the desired library-specific chemistry is carried out on the amino group of the photo-cleavable link, as shown in FIG. 2. Two possible orthogonal cleavage protocols are now possible. In the first case (FIG. 2), a chemical cleavage generates two charged species: the 'outer chemistry' or ligand-containing moiety, and the 'between linkers' component. The ratio between the two species, after being corrected for ionization efficiency, is indicative of the yield of product. In the second case (FIG. 3), a photo-cleavage releases only the ligand, which is then suitable for assay.

Diversity of solid-phase chemistries

It is widely accepted that the 'translation' of a solution-phase reaction sequence to a solid-phase protocol suitable for the production of a quality compound library is usually time-consuming and not always successful. The more robust of the available synthesizers have limitations in the range of temperatures in which they can operate and restrictions on the solvent usage, which can necessitate a significant investment in experimentation to re-define a suitable reaction protocol. Despite this perceived barrier, there is now a significant body of published reaction sequences for solid-phase library production. The rapid increase is well illustrated by the fact that in 1992–1995, there were 168 reported solid-phase chemistries, and in 1996 alone there were 270

ENCODING
Strategy to identify members of a combinatorial bead-based library. A surrogate analyte is associated with each member of the combinatorial library. This is often achieved by the use of tags attached to the beads on which the library members are assembled, which allows the reaction history of each bead to be determined.

MIX-AND-SPLIT SYNTHESIS STRATEGY
The solid support (for example, beads) is divided into portions, each of which is subjected to reaction with a single monomer. Combining these portions results in a single batch of solid support bearing a mixture of components. Repetition of the divide, couple, recombine processes results in a library in which each discrete particle of solid support carries a single library member.

reported solid-phase chemistries, which had further increased to 420 by 1997 (REFS 2–4). After the report for year 1997, no further summaries were published, although on the basis of the steady increase in publications describing solid-phase reactions[5–9], it is likely that the number for 1998 would have been higher than that for the previous year. However, many of these protocols are either unsuitable because the overall yields of the products are low and could lead to ambiguity in the interpretation of the eventual assay results, or because the range of compatible monomer usage is restricted to such an extent that the diversity of the possible compounds of the resulting library is too limited.

In the same way that a combinatorial library explores all the possible combinations of monomers compatible with the required chemical protocol, the rapid optimization or discovery of the chemical protocol itself can be carried out in a combinatorial way. The dimensions of the problem are now those variables such as solvent, temperature, time, reagent and monomer characteristics (for example, whether it is aliphatic or aromatic). Using an analytical-construct approach, a combinatorial search through a significant number of the possible combinations of these variables becomes a relatively simple exercise to carry out, and quickly results in an optimized chemistry that can then be used to synthesize a compound library.

The form of this construct is very similar to that described in FIG. 1; it differs in that the orthogonal link is replaced by a non-cleavable spacer chosen to provide comparatively unrestricted access to the chemistry site. All other features are retained to allow encoding and quantitation at the single-bead level. A typical approach for rapidly exploring many permutations of the parameters that affect the outcome of a candidate reaction sequence is illustrated in FIG. 4.

Carrying out the optimization (reaction screen) in the same equipment, and using a similar encoded construct as the one that will eventually be used for library generation guarantees the relevance of the determined reaction protocol. It should be noted here that criteria for selecting a practical library synthesis protocol might not equate to the combination of reaction conditions giving the maximum yield at each individual step. For example, as a general principle, changes of solvent between the individual steps of a library-producing reaction sequence should be avoided, as the inter-step washing required is wasteful of solvent (in terms of cost and disposal) and can lead to resin breakdown as a result of the physical size changes (that is, swelling) that occur as the solvent is varied. Successful solid-phase reaction sequences represent many compromises that are dictated by the choice of equipment and the resin type used.

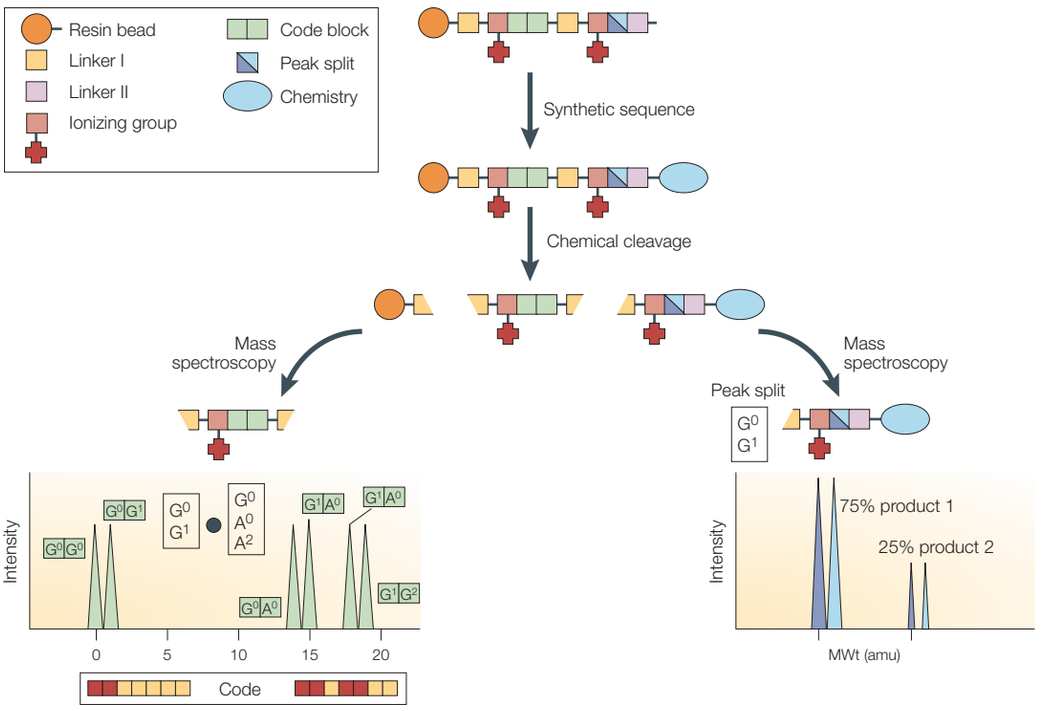

Figure 2 | **Product analysis using an analytical construct.** The desired library-specific chemistry is carried out on the amino terminal of the analytical construct. Chemical cleavage generates two charged species: the 'outer chemistry' or ligand-containing moiety, and the 'between linkers' component. In the bottom left is shown the region of the mass spectrum containing the 'between linkers' components referenced to the lowest molecular weight. More specifically is shown the code derived from the two-step synthesis using the indicated mixtures of isotopically labelled amino acids (either glycine or alanine). The principles of coding with this strategy are explained in the main text and in FIG. 7. The bottom right of the figure shows the region of the spectrum containing the two products of the chemistry easily identified by the signature peak-split introduced using an equimolar mixture of isotopically different glycines.

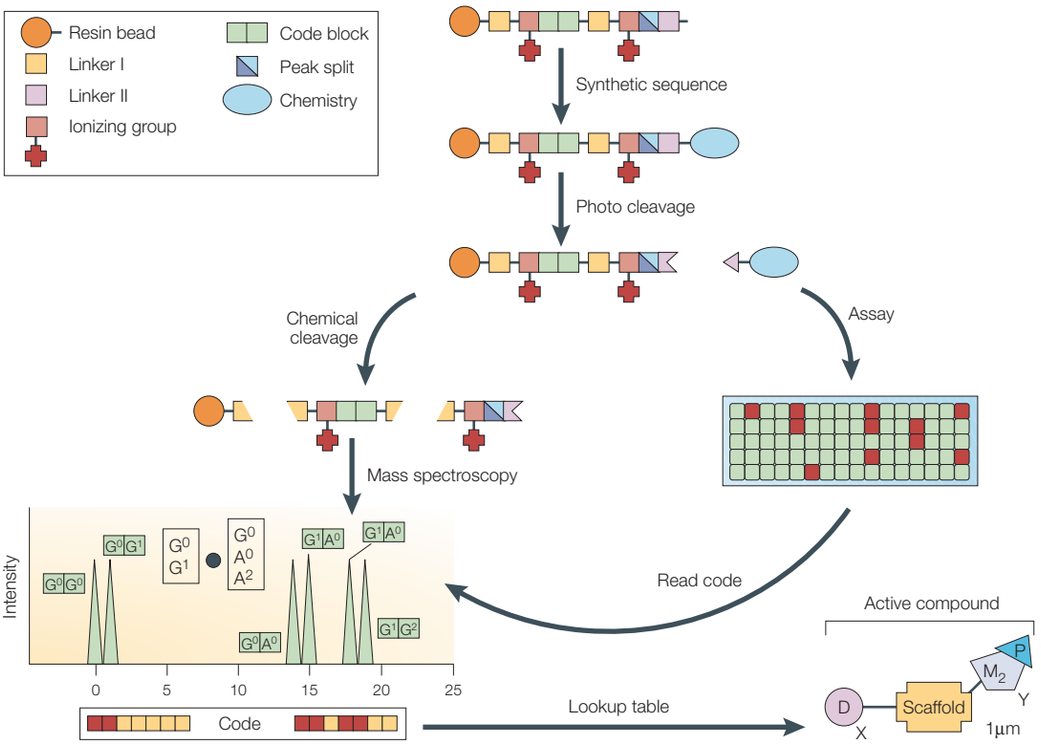

Figure 3 | **Assay operation using an analytical construct.** Photo-cleavage of the construct shown will release only the ligand produced by the synthetic sequence, which can be used in an assay. If a ligand tests positive in the assay, its identity can be determined by chemical cleavage of the remainder of the construct from which it originated. This gives the 'between linkers' component containing the code block, the mass spectra of which can be decoded to give the ligand identity. For an explanation of the mass spectrum, see the legend to FIG. 2.

In the past few years at GlaxoSmithKline, rapid-reaction optimization (screening) has been used to reduce several desired chemical protocols to the level whereby they were suitable for library generation. Reaction screens tested approximately 10,000–72,000 variations of the candidate reaction sequence, and typically about 5,000 beads were analysed to obtain the optimized chemistry. It should be appreciated that whole sections of the reaction space generated in the screen can be quickly eliminated on the basis of a few sampled beads; for example, sampling from a given solvent might indicate that no reaction occurred at all, therefore obviating the need to continue to sample from this area of reaction space. The experimental component of a reaction optimization or screen, including the sampling and data analysis, takes of the order of two weeks to carry out.

As an example of the complete process that includes the initial experimental design, acquisition of reagents, setting up of the robotic synthesizer and so on, a schematic showing the complete time line for the last screen completed at GlaxoSmithKline and subsequent library synthesis is shown in FIG. 5.

Encoding strategies

Numerically large libraries are by necessity associated with an encoding strategy. The earliest demonstrated encoding, which was applied to peptide libraries, required iterative decoding by several rounds of synthesis to 'arrive' at the identity of the binding peptide and is only applicable where the coupling rate is approximately monomer independent[10]. Four encoding methods applicable to mix-and-split solid-phase combinatorial library synthesis for lead discovery have been used as the basis of a significant commercial endeavour[11]. Of these, the first three use a chemical encoding process, and the fourth a physical labelling (ID radio chip inserted with the solid-phase used for synthesis), which allows a directed sort to be carried out between synthesis steps. At the completion of each chemistry step, the cassette containing the resin and transponder is 'interrogated' to determine its synthetic history. Based on the protocol, each cassette is then individually placed in the appropriate reactor for the next chemistry step. FIG. 6 shows a comparison in terms of the number of steps between a discrete synthesis strategy, a tag-encoding procedure and an encoding method incorporated with the analytical construct described above. When taken together with the ability to quantitate the outcome of the chemistry at each step of the synthesis, the spatial, mix-and-split procedure is well suited to the production of numerically large libraries with few restrictions as to the chemistry protocol implemented.

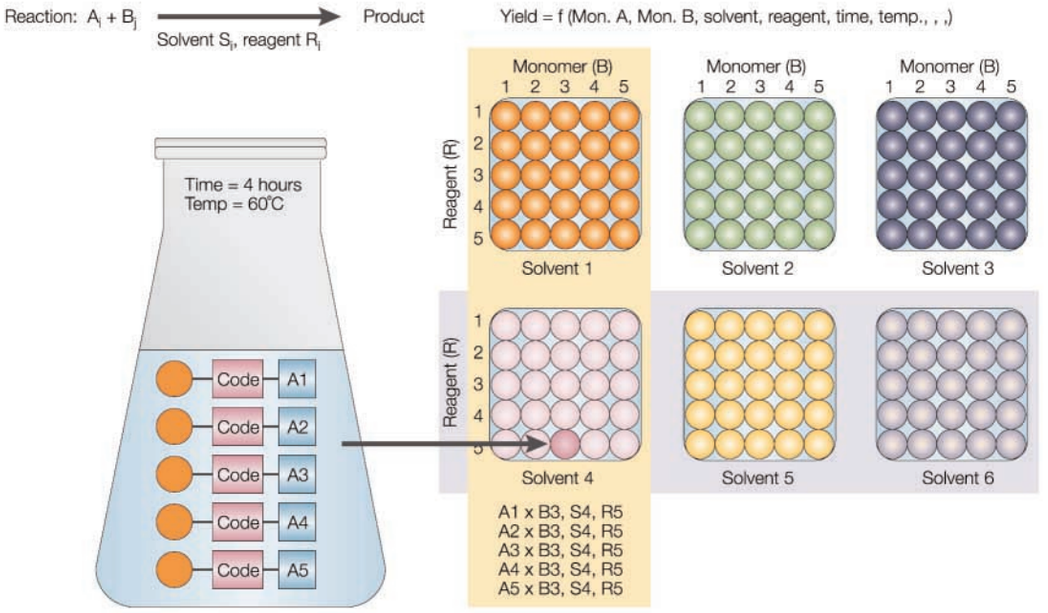

Figure 4 | **Optimization of reaction protocols using an analytical-construct approach.** This example illustrates the optimization of the reaction monomer type A with monomer type B to give a product, in a process that evaluates six different solvents and five different reagents. In all, five different actual A monomers are permutated against five different B monomers, for a total of 750 reactions (5 x 5 x 5 x 6 = 750). The coding requirements for this process are usually significantly more modest than those for encoding a chemical library, and in this instance only five different codes are needed. At the conclusion of the reaction, resin beads are sampled from each well and analysed by mass spectroscopy after chemical cleavage of the construct. Sampling from the well marked with a star would give the reaction outcome for all variations of monomer A with monomer B_3, in solvent 4, and using reagent R_5, after a time of 4 hours at 60 °C.

Binary code assembly and visualization. Of the several schemes demonstrated by our group[1], a binary encoding protocol was found to be the most useful for general use with bead-based libraries. This results in a code that can be read by mass spectroscopy with a throughput of greater than 50 samples per hour. This code format is based on a two-cycle incorporation of mixtures of glycine and alanine, in which the individual components differ in the number and composition of stable isotopes of carbon (^{13}C), hydrogen (^{2}H) and nitrogen (^{15}N).

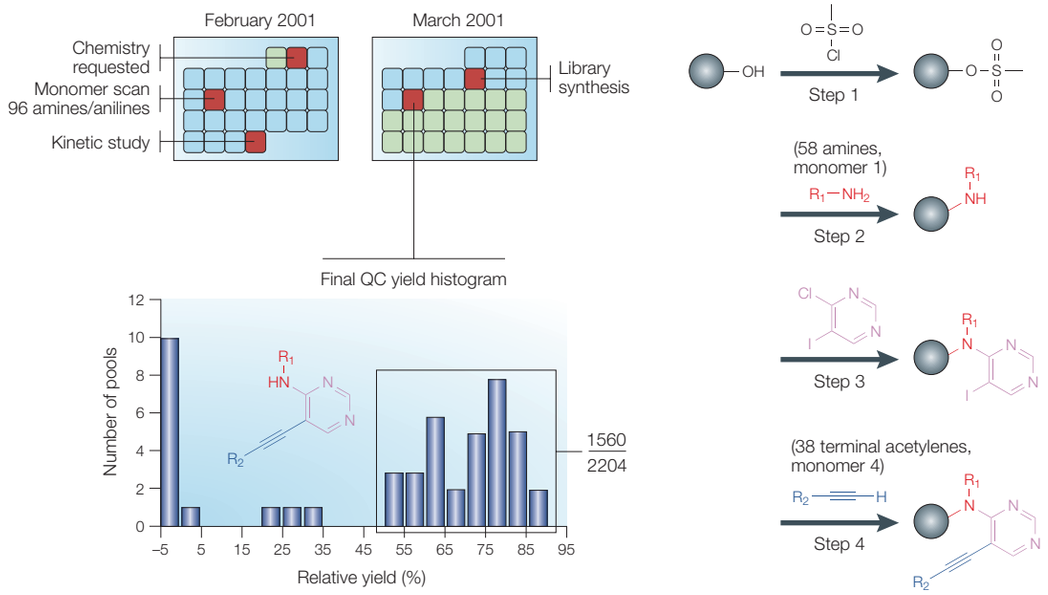

Figure 5 | **An example of a complete process using the analytical-construct approach.** The chemistry protocol is shown on the right, and on the bottom left is the yield distribution for the complete library, which targeted a total of 2,204 compounds.

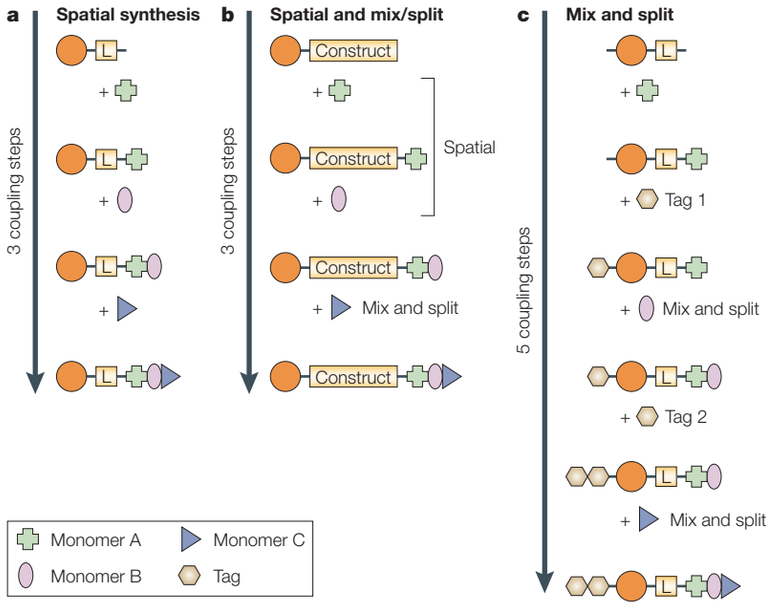

Figure 6 | **Encoding strategies.** The figure shows a comparison in terms of the number of steps between **a** | a discrete synthesis strategy (in which each product is synthesized in its own vessel, and its identity is therefore encoded by its position), **b** | an encoding method incorporated with the analytical construct, which uses both the coding in the construct and spatial encoding, and **c** | a full mix-and-split procedure, in which each reaction step up to the last needs to be chemically encoded in the tag.

As a short-hand nomenclature, glycine is shown as G^x, where the value of x indicates the number of substituted atoms, and similarly for alanine the symbol is A^x. In FIG. 7, the progressive assembly of the code block is shown along with a representation of the mass spectrum in the region corresponding to the total molecular mass of the released fragment. It is usual to only use a subset of all the possible codes such that the generated pattern contains either four or six peaks. This, in effect, represents EVEN PARITY and allows for a built-in check that a product peak has not been superimposed on one of the code peaks, or that a product of the chemistry is present in the code region. Individual beads are sampled, chemically cleaved and the resulting supernatant analysed by direct infusion into an electrospray quadrapole mass spectrometer, typically with a 1-minute data-acquisition cycle. Using an auto sampler, >500 samples are processed per instrument per 24 hours. Code reading and checking is accomplished automatically by software that takes into account the natural abundances of isotopes of the comprising atoms of the complete code block.

Solution-phase assay. Encoded bead-based libraries represent a unique problem for assay when separation of the ligand from the corresponding code is required, as is the case for a solution-phase assay. The overall procedure (FIG. 8) can be broken down into three discrete steps. First is the distribution of the beads, cleaving the compounds from the beads, and separating the solution containing the ligand from the bead but retaining the association should later decoding be required. Second is the distribution of the ligand solution into the requisite number of sub-plates and completing the assay. The final step is analysing the assay result for positives, and reading the codes from the corresponding beads and completing the identification of the candidate active compound(s).

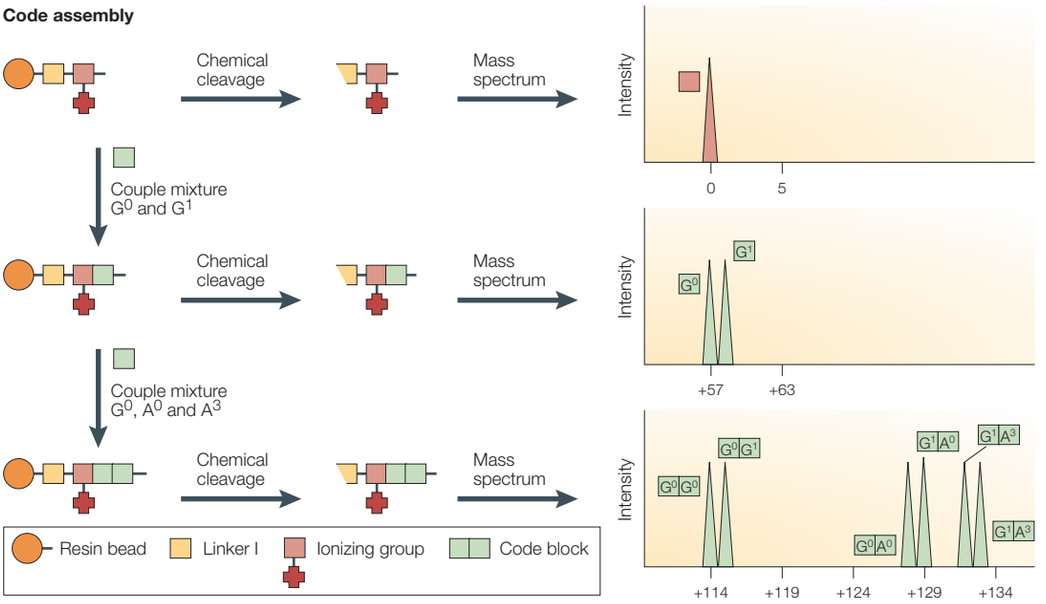

Figure 7 | **Assembly of the code block.** The progressive assembly of the code block is shown along with a representation of the mass spectrum in the region corresponding to the total molecular mass of the released fragment. At the completion of the first code-block incorporation — a mixture of G^0 and G^1 — two peaks 1 atomic mass unit apart are seen, and after the second coupling with a mixture of G^0, A^0 and A^3, six peaks are produced corresponding to all of the combinations of the binary and tertiary mixtures, and comprises the complete code.

EVEN PARITY
The knowledge that an even number of code peaks (usually four or six) indicates whether or not a reaction product has overwritten one of the peaks. Should this occur, it is still possible to read the code by inference, providing only one overwriting event has occurred.

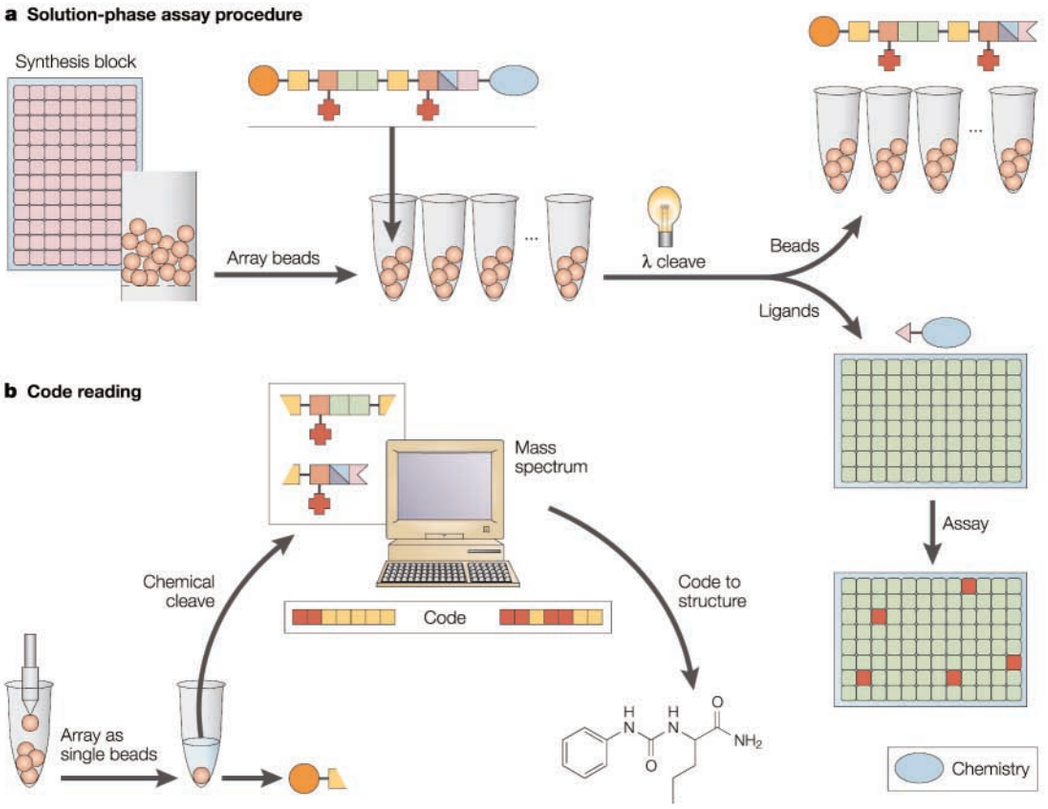

Figure 8 | **Procedure for solution-phase assays with encoded bead-based libraries. a** | First, the beads are distributed into vessels, and the ligands released into solution by photo-cleavage (the associations with the beads need to be maintained). The ligand solutions are then used in assays. **b** | After analysis of the assay results, the beads corresponding to active compounds are decoded by mass spectrometry to reveal the identity of the compounds.

Statistical decoding methodology. Using a statistical approach essentially eliminates the false-positive problem that results from either an actual false positive occurring in the assay, or of a positive resulting from an undetermined outcome during synthesis. Solutions comprising an average of about five compounds (simultaneous cleavage of five beads) are assayed with an average representation of about five for each compound to be tested. As the arraying of the five beads for photo-cleavage is random, for each active compound an average of five wells will 'light up'. Decoding all of the beads associated with a positive assay result generates a table of codes whereby the common code equating to the active compound occurs with an average frequency of five. Finding a code with a frequency of greater than three is a clear indication that the corresponding compound is active in the assay, as this frequency is significantly higher than could be accounted for by chance alone. In this way, false positives, either associated with the assay itself or from some unknown factor occurring during any of the synthesis or cleavage procedures, have been eliminated. The sequential processing of the data and the results obtained are illustrated in FIG. 9. Also shown is the validation of the identified active compound by discrete synthesis and purification.

Scale of synthesis

In addressing the first three of the technical issues of the synthesis and testing of numerically large combinatorial libraries using a bead-based approach, the scale of the synthesis is dictated by the capacity of the individual beads used, which is about 1 nmole for the typical 160 μm beads. Ideally, a library is made and then tested against the various target screens available at that time, avoiding the need for long-term storage. The fact that the identity of any given compound of an encoded library is not known until after it is assayed, and the code read, does not allow a confirmatory test of those compounds assessed to be hits in a given screen. In our experience, however, the statistical decoding described above would seem to address this need. A clear advantage of the bead-based scale of synthesis is, then, a sparing use of the monomers required for the library synthesis, the ready supply of which can present problems. The disadvantage of bead-based libraries is the multi-step procedures of bead arraying, cleavage of the compound from the beads, retaining the relationship between the code (which is still attached to the bead) and the compounds during the assay process, and the decoding. The synthesis of compounds at the multi-milligram scale, on the other hand, is a significant investment in the compounds, a very large proportion of which will never test

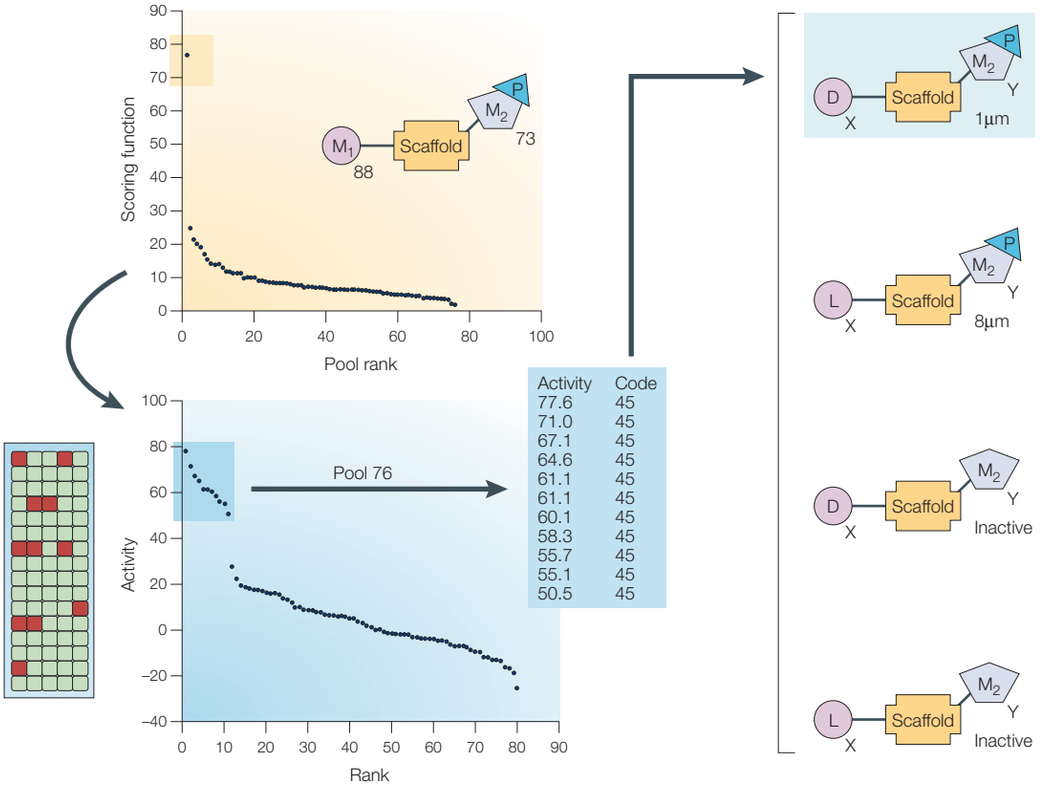

Figure 9 | **Decoding data using a statistical approach.** Synthesis of a library of compounds based on the attachment of two sets of diverse monomers to a scaffold as shown in the schematic, top left, resulted in 73 pools of beads with a common monomer 2 (M_2), of 88 different compounds as defined by the identity of monomer 1 (M_1). The assay results of an average of five beads per well were analysed and ranked according to the likelihood of containing an active compound. One pool (76) stood out from the rest, and the individual assay values were again ranked, showing 11 activity values significantly above the remaining wells. Beads from the corresponding wells were arrayed as 'singles', chemically cleaved, and codes read by mass spectrometry. Tabulation of all the codes showed a single common value — code 45 — to be present in each (lower left). Resynthesis of the designated compound with and without the protecting group on a basic functionality, and with both optical isomers in the monomer 1 position, confirmed the finding that the active compound was as designated by the original code obtained.

positive in any screen. The investment comprises the resources associated with the synthesis, purification, and a facility for storage and retrieval of compound for assay. This latter requirement has its limitations, in that experience has shown that even at lower temperatures, storage in a solvent (usually dimethylsulphoxide) leads to degradation and/or precipitation of a proportion of the compounds, which in turn leads to a false outcome in the assay. Without a doubt, however, the most important aspect of using a larger scale of synthesis is the smaller number of compounds made overall and therefore the lower probability of finding a hit or lead.

The numbers issue

Current estimates put the size of the human genome at about 35,000 genes and that of a typical bacterial genome at 4,000 genes. Considering only the human genome, if even 10% of the genes represent good candidates for small-molecule interactions, either as a medically important target or as a tool in understanding function, this is still a significant number. Including the various human, animal and plant pathogens, it is not unlikely that the total number of targets for which small-molecule ligands are required is >10,000. Small-molecule space itself has been estimated to be huge — 10^{14} to more than 10^{30} molecules — depending on the criteria under which the calculation is made. If we assume for the moment that the average numerical size of a class of compounds that can be produced by a common synthetic strategy (a requirement for library synthesis) is 10^5–10^6, then the lower boundary above still translates to between 10^8–10^9 different libraries. This number dwarfs the combined global output of chemical libraries synthesized so far.

By contrast, the successes — for example, identifying and defining the characteristics of epitopes as recognized by antibodies[12], and the mechanism of antigen–antibody interactions[13,14], to name a few — achieved in the 1980s and early 1990s with peptide libraries, both chemically derived and those constructed using molecular biology (for example, by phage display), reflects the very different scale of the problem itself. Only one type of monomer (that is, amino acids)

was involved, and, on the basis of the genetic code, the number of different monomers was only 20. This allowed a more systematic and complete search to be made for peptide sequences with ligand properties. For example, most linear epitopes recognized by antibodies comprise six or fewer amino acids, therefore requiring a library of only 20^6 or possibly 20^7 members for completeness. This number of compounds is readily prepared by either of the two methods above, and, in principle at least, systematically addresses many of the potential protein–protein interactions.

Returning to the small-molecule problem, it is not conceivable that a systematic approach using synthesis and screening is going to be feasible. It is more likely that as computational approaches able to assess the likelihood that a given chemical class comprises members with utility as ligands develop further, the number of libraries targeted for synthesis can be decreased markedly. This 'virtual' screening ideally considers not only the inherent ability of the molecules to interact with biological targets of interest, but also their likely toxicity, bioavailability and clearance as it pertains to therapeutic utility.

Conclusions

The greatest potential of combinatorial chemistry is in the numbers. Synthesis technology has advanced to the stage where numerically large, quality libraries, at a practical scale of synthesis, can be produced. Encoding strategies eliminating the problems of indeterminate assay results are also available. Most, if not all, of the negative aspects associated with solid-phase synthetic methodologies have been overcome. Rapid optimization of candidate-chemical synthesis protocols, as well as discovery of novel synthetic routes to chemical classes, has been demonstrated. The challenge is now clearly focused on selecting the best compound libraries for synthesis and testing. The magnitude of small-molecule space is so vast as to preclude a systematic search in the foreseeable future. It would be a great pity to limit the use of available library technologies to the synthesis of relatively small libraries, either comprising analogues around a hit or lead, or for primary screening purposes. The frequently expressed opinion that only a few hundred compounds from a given chemical class is all that is required to determine if a ligand from that class exists at all is not supported by experiment (H. M. Geysen, unpublished data). Conclusions such as this last statement, as well as the belief that present computational techniques are sufficient to predict activity, need to be carefully validated by experiment, which is something that is very difficult given the magnitude of the problem itself.

Past experiences from the pharmaceutical industry suggests that a 'blind' approach to the synthesis of compound libraries on solid-phase, without either an extensive program to validate the applicable chemistry or an ability to quantitate the outcome, often leads to uninterpretable assay results and is not cost effective. This, coupled with the perceived limitations in the richness of chemistries which can be carried out on, or translated to solid-phase, has resulted in a move to supply compounds for first-stage discovery screening by highly automated solution-phase synthesis equipment followed by high-throughput purification techniques. This is both expensive in terms of the setup costs as well as the ongoing operating expenses. We should not lose sight of the very concept of solid-phase synthesis, namely the ability to drive reactions to completion without incurring a significant purification penalty. Neither peptide or nucleotide chemistry would be practical without this innovation, the invention of which served as the basis for the Nobel Prize awarded to Bruce Merrifield. The general translation of classical organic chemistry to solid-phase does pose some unique challenges, but the inherent gains realizable in the production of numerically large compound libraries make solid-phase library production a necessity rather than a choice.

1. Geysen, H. M. *et al.* Isotope or mass encoding of combinatorial libraries. *Chem. Biol.* **3**, 679–688 (1996).
2. Hermkens, P. H. H., Ottenheijm, H. C. J. & Rees, D. C. Solid-phase organic reactions: a review of the recent literature. *Tetrahedron* **52**, 4527–4554 (1996)
3. Hermkens, P. H. H., Ottenheijm, H. C. J. & Rees, D. C. Solid-phase organic reactions. Part 2: A review of the literature November 1995-November 1996. *Tetrahedron* **53**, 5643–5678 (1997).
4. Booth, S., Hermkens, P. H. H., Ottenheijm, H. C. J., Rees, D. C. Solid-phase organic reactions III: A review of the literature November 96-December 97. *Tetrahedron* **54**, 15385–15443 (1998).
5. Dolle, R. E. Comprehensive survey of chemical libraries yielding enzyme inhibitors, receptor agonists and antagonists, and other biologically active agents: 1992 through 1997. *Mol. Divers*. **3**, 199–233 (1998).
6. Dolle, R. E. & Nelson, K. H. Comprehensive survey of combinatorial library synthesis: 1998. *J. Comb. Chem.* **1**, 235–282 (1999).
7. Franzen, R. G. Recent advances in the preparation of heterocycles on solid support: A review of the literature. *J. Comb. Chem*. **2**, 195–214 (2000).
8. Hall, S. E. Recent advances in solid phase synthesis. *Mol. Divers.* **4**, 131–142 (1998).
9. Andres, C. J., Denhart, D. J., Deshpande, M. S. & Gillman, K. W. Recent advances in the solid phase synthesis of drug-like heterocyclic small molecules. *Comb. Chem. High Throughput Screen.* **2** , 191–210 (1999).
10. Geysen, H. M., Meloen R. N. & Barlelin, S. J. Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid. *Proc. Natl Acad. Sci. USA* **81**, 3998–4002 (1984).
11. Czarnik, A. W. Encoding strategies for combinatorial chemistry. *Curr. Opin. Chem. Biol.* **1**, 60–66 (1997).
12. Geysen, H. M., Mason, T. J. & Rodda, S. J. Cognitive features of continuous antigenic determinants. *J. Mol. Recognit.* **1,** 32–41 (1988).
13. Geysen, H. M. *et al.* Chemistry of antibody-binding to a protein. *Science* **235**, 1184–1190 (1987).
14. Getzoff, E. D. *et al*. Mechanisms of antibody-binding to a protein. *Science* **235**, 1191–1196 (1987).

Acknowledgements

The authors would like to thank the members of the Diversity Sciences Department at GlaxoSmithKline for the fine work carried out on this technology over the period 1993–2001.

Online links

FURTHER INFORMATION

Geysen laboratory:

http://www.virginia.edu/chem/people/faculty/geysen/

Access to this interactive links box is free online.