URL: http://www.iupac.org/symposia/proceedings/phuket97/lowe.html

Combinatorial Libraries for Studying Molecular Recognition

Gordon Lowe, Susanna Leon and Rachel Quarrell

Dyson Perrins Laboratory, Oxford University, South Parks Road, Oxford, OX1 3QY, UK

Abstract: Combinatorial libraries have great potential for studying biomolecular recognition. They consist of large numbers of structurally diverse compounds which can be generated rapidly. The key to their use is in the development of selection assays which allow the most active compounds to be identified. The most commonly used selection procedures are based on binding, the compound(s) with the highest affinity being isolated and identified. In this report some fundamental questions concerning the assaying of libraries on solid supports by enzyme catalysis are addressed.

INTRODUCTION

The synthesis of diverse compounds with related molecular structure can now be achieved rapidly by the techniques of combinatorial chemistry (ref. 1). Such sets of compounds, referred to as libraries, have most commonly been screened in the search for lead compounds for a therapeutic target or for the optimisation of lead compounds in the development of drug candidates. It is for this reason that the technology has been rapidly adopted and developed by the pharmaceutical industry. Combinatorial chemistry also provides the opportunity to address fundamental questions in the area of biomolecular recognition.

Combinatorial Library Format

The preferred format of a combinatorial library depends on the way in which it is to be used and consequently it is imperative to have an established assay before the library format is determined. A combinatorial library may be synthesised in solution or on a solid support. If prepared in solution, members of the library are generally synthesised simultaneously in parallel using semi- or fully automated equipment. The advantage of this method is that it allows individual compounds to be purified and characterised by traditional methods. A disadvantage is that the number of compounds that can be prepared simultaneously is relatively limited unless there is investment in expensive automated equipment. Synthesis on solid supports such as resins allows very large combinatorial libraries to be generated, which may be assayed as individuals or mixtures, on the solid phase or after release into solution. Where the split-and-mix methodology is used (Fig. 1),

---

*Invited lecture presented at the International Conference on Biodiversity and Bioresources: Conservation and Utilization, 23–27 November 1997, Phuket, Thailand. Other presentations are published in *Pure Appl. Chem.*, Vol. 70, No. 11, 1998.

each resin particle carries one member of the library, and a solid-phase assay is very effective (ref. 2).

Combinatorial Chemistry

The solid phase method of synthesis was introduced by Merrifield for the synthesis of peptides (ref. 3) and later adapted for the synthesis of oligonucleotides. Not surprisingly, therefore, combinatorial libraries of these oligomers were the first to be developed. Combinatorial libraries of several million compounds can be prepared by the split-and-mix method of synthesis, each bead behaving as a micro-reaction vessel and carrying a single library member. Even larger libraries of oligonucleotides are commonly prepared by using a mixture of nucleoside phosphoramidites in the synthesis at the library sites.

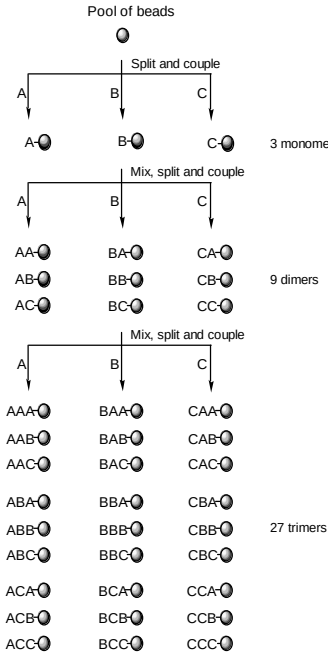

Fig. 1. The split-and-mix method of synthesis of a combinatorial library using building blocks A, B and C.

One of the current challenges in combinatorial chemistry is to develop highly efficient methods of synthetic organic chemistry on the solid phase. A major benefit of the solid phase method of synthesis is that an excess of reagents can be used to drive reactions to completion and then washed from the resin before the next synthetic step is performed.

The Selection Assay

Biological assays are generally performed on single compounds and this is still possible if the combinatorial library is prepared by parallel synthesis. This is nevertheless a time consuming process, even with automated high throughput screening methods. Time can be saved by assaying mixtures of compounds, but there has been much debate about this approach since compounds

could have either synergistic or antagonistic effects and thus give false positive or negative results. However, as a compromise, relatively small mixtures of compounds (usually about 10) are now being used and the components of the most highly effective mixtures then assayed individually.

Assaying compounds while still attached to resin raises many questions. The resin must be compatible with synthetic methods which involve the use of many organic solvents and reagents. The resin must also be compatible with the biological assay which will be performed in aqueous solution. Moreover, it is essential that the biological macromolecule used in the assay is compatible with the resin and can penetrate its interstices. One of the most commonly used supports is TentaGel, a grafted copolymer of polystyrene and polyethylene glycol chains of approximately 3kDa (ref. 4). The mobility of molecules covalently attached to the polyethylene glycol chains of TentaGel resin beads has been shown by NMR spectroscopy to be similar to that in free solution (ref. 5). Fluoresceinated enzymes have been shown by confocal fluorescence microscopy to penetrate to the core of such beads (ref. 5), but it has also been shown that the proteolytic enzyme chymotrypsin 'shaves' only the surface molecules of a substrate attached to such beads (ref. 6). PEGA beads, however, which are made of co-grafted polyacrylamide and polyethylene glycol, not only allow large macromolecules to penetrate the interstices (ref. 7), but also allow enzyme catalysis to proceed (ref. 8). The main disadvantage is that they are less robust than TentaGel and more difficult to handle. There is certainly scope for the development of robust resins of regular size and loading which are compatible with both the solvents and reagents required for synthesis and the aqueous environment required for biological assays.

The most commonly used biological assays involve the selection of compounds from combinatorial libraries with the highest affinity for the target macromolecule such as a receptor or enzyme. Rarely has the catalytic activity of an enzyme been exploited to effect the selection. However, before the specificity of proteases can be addressed using peptide libraries attached to resin beads, it is important to establish that the resin itself does not perturb the specificity of the enzyme.

THE SPECIFICITY OF PROTEASES

Substrate recognition by proteases

The high degree of specificity frequently manifested by proteases stems from their ability to recognise seven or more amino acid residues flanking the scissile bond of the peptide or protein substrate (Fig. 2) (ref. 9). This postulate was verified by the X-ray crystal structure of several protein protease inhibitors complexed with their cognate protease, e.g. in the bovine pancreatic trypsin inhibitor (BPTI) / trypsin complex, seven amino acid residues in the reactive site loop of BPTI make van der Waals contact with the trypsin molecule (ref. 10). Although the amino acid sequence of a natural substrate for a given protease may be known and the site of cleavage identified, this does not necessarily represent the sequence most effectively cleaved by the enzyme, as the recent investigation of the amino acid sequence around the autoproteolysis sites of herpes simplex virus type 1 protease illustrates (ref. 11). In order to design highly selective and potent inhibitors of particular proteases, a general and convenient strategy for determining specificity is required.

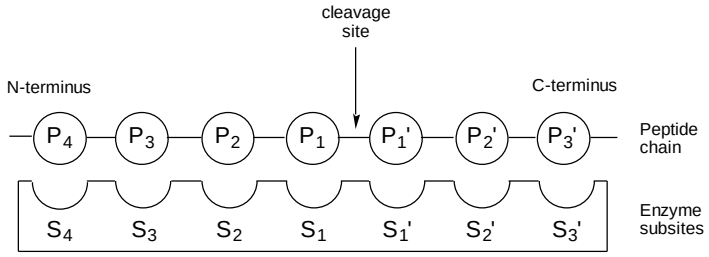

Fig. 2. A schematic diagram of the active site of a protease, each S subsite being capable of binding an amino acid residue P. The catalytic site of the protease is between S_1 and S_1' and the scissile bond is indicated by the arrow.

Library size

If a protease has seven binding subsites each capable of recognising only one of the proteinogenic amino acids, then the enzyme would be capable of selecting and cleaving one heptapeptide out of a total of 20^7 or 1,280,000,000. In order to determine the specificity of such enzymes an entirely new approach is needed which has become possible with the advent of combinatorial chemistry. If the library is prepared by the split-and-mix method of synthesis and retained on beads, a practical limit to the size of the library is dictated by the need for sufficient loading on a single bead to allow structural analysis of the selected product. Beads of 90 μm diameter will generally carry sufficient peptide for analysis. A gram of the PEGA type resin, which is a convenient amount to handle, will consist of several million such beads. To ensure that there is a high probability that all the peptide sequences are represented, between 5 and 10 times as many beads are required as there are components in the library (ref. 12). Thus a library of 160,000 tetrapeptides based on the 20 proteinogenic amino acids is generally the practical limit to the size of a complete combinatorial peptide library.

Assay for protease specificity

Since the subsites which bind the amino acids of the substrate close to the scissile bond generally provide the major contribution to the specificity of the enzyme, our strategy is to use a tetrapeptide library attached to PEGA beads through several glycine residues and a nucleophile-labile HMBA linker (ref. 13) The peptide library is then flanked at its N-terminus by further glycine residues and capped with the N-dansyl fluorophore (Fig. 3). Incubation of the dansylated peptide library with the target protease for a limited period of time followed by thorough washing of the beads to remove all the cleaved dansyl-peptides provides a set of beads with varying degrees of intact dansylated peptide. Those beads carrying peptides with the highest susceptibility should be most extensively cleaved and thus appear dimmest under a fluorescence microscope.

An advantage of this strategy is that the site of cleavage of a tetrapeptide is not predetermined. Consequently the tetrapeptide library explores up to four subsites on each side of the scissile bond and so reveals the subsites with the greatest influence on the protease specificity. Release of the cleaved and uncleaved peptide from individual beads can be achieved with propylamine and the peptide sequence and site of cleavage determined by mass spectrometry.

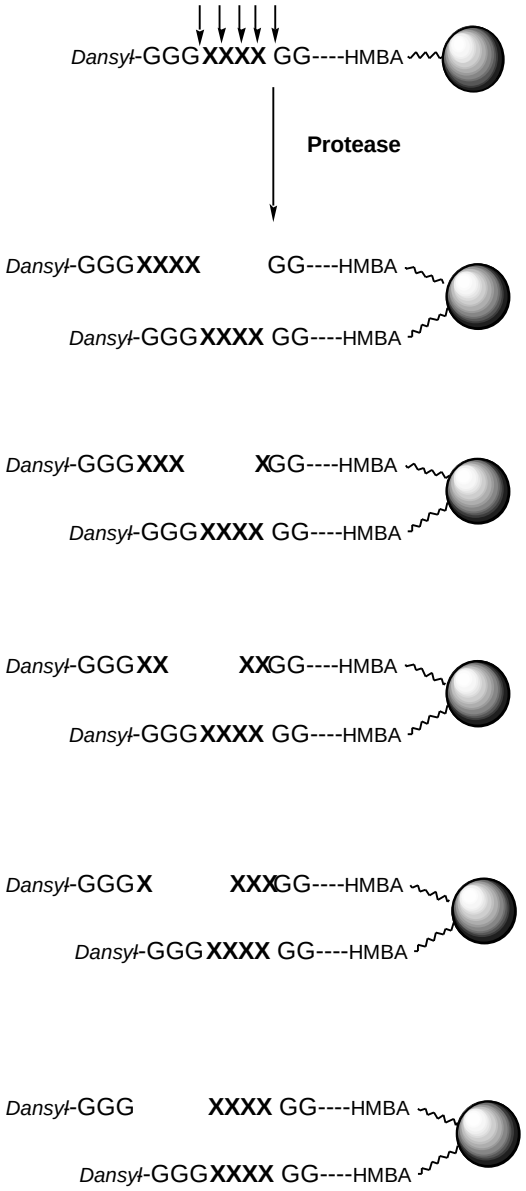

Fig. 3. The probable sites of cleavage of a combinatorial peptide library by a protease are indicated by the arrows. Beads carrying cleaved peptides are isolated, and the sequence of the cleaved and uncleaved peptides determined.

If a dipeptide library is used then the analysis is simpler, as the individual members of the library can be synthesised in parallel in a 20 x 20 array. The beads carrying the most susceptible peptide sequence are identified by release of the fluorescent label into solution. The position of the well within the array immediately identifies the dipeptide sequence and the site of cleavage can

be determined by releasing the cleaved and uncleaved peptides from the beads and identifying them by mass spectrometry.

Papain is a cysteine protease which was shown by Schechter and Berger to have seven binding subsites, four on the N-terminal and three on the C-terminal side of the scissile bond (ref. 9). Moreover, they showed that papain has a strong predilection for L-phenylalanine at the P_2 subsite (ref. 14). A dansylated-peptide attached to PEGA beads through an HMBA linker was prepared containing one L-phenylalanine residue, the remaining amino acids being glycine (Fig. 4). Papain cleaves this peptide between the first and second glycine residues on the C-terminal side of the phenylalanine residue, as expected from the known specificity of the enzyme in solution (ref. 14). The same PEGA-bound peptide is hydrolysed by α-chymotrypsin between the phenylalanine and the adjacent glycine on its C-terminal side. This also is in accord with the specificity of this enzyme in solution which is known to prefer large hydrophobic, particularly aromatic, residues at the P_1 subsite. These observations suggest that the specificity of proteases is not perturbed when peptide substrates are bound to PEGA beads through a cleavable HMBA linker and dansylated at their N-terminus. It should therefore be possible to determine the specificity of these and other proteases by assaying combinatorial peptide libraries prepared by the split-and mix-method of synthesis on PEGA beads.

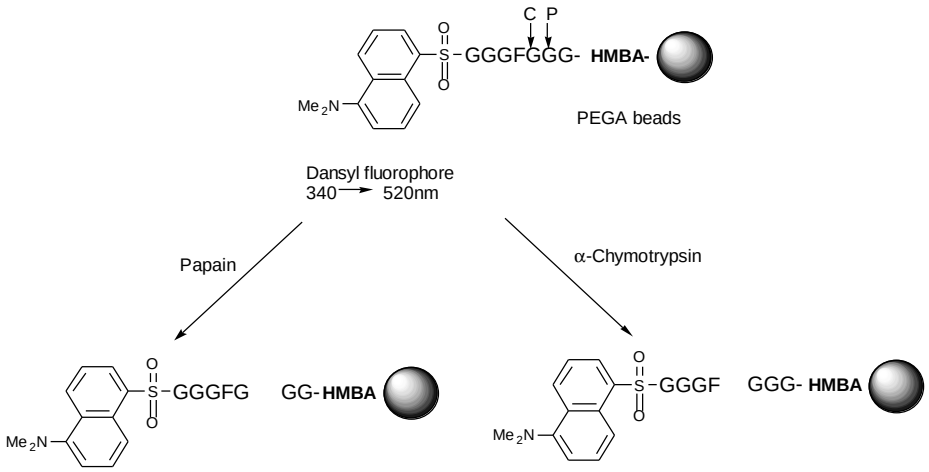

Fig. 4. The primary specificity of papain and α-chymotrypsin for peptides covalently attached to PEGA beads is the same as for those in free solution.

SUMMARY

(i) Biomacromolecules, e.g. trypsin have been shown to penetrate beads of co-grafted polyethylene glycol resin such as TentaGel.

(ii) The mobility of molecules covalently attached to the polyethylene glycol chains of TentaGel resin beads has been shown by NMR spectroscopy to be similar to that in free solution.

(iii) Susceptible peptides covalently attached to PEGA beads are cleaved by proteases.
(i) The primary specificity of the proteases, papain and α-chymotrypsin, towards substrates bound to PEGA beads is the same as in solution.

REFERENCES

1. G. Lowe, *Chem. Soc. Rev.* **24**, 309-317 (1995).
2. K. Lam, S. Salmon, E. Hersh, V. Hruby, W. Kazmierski and R. Napp, *Nature (London)*, **354**, 82-84, (1991).
3. R.B. Merrifield, *Fed. Proc. Fed. Amer. Soc. Exp. Biol.*, **21**, 412 (1962).
4. W. E. Rapp. In *Combinatorial Chemistry* (S. R. Wilson and A. W. Czarnik, ed.), Chapt. 4, John Wiley & Sons, New York, 65-93 (1997).
5. R. Quarrell, T. D. W. Claridge, G. W. Weaver and G. Lowe, *Molecular Diversity*, **1**, 223-232 (1996).
6. J. Vagner, G. Barany, K. S. Lam, V. Krchnak, N. F. Sepetov, J. A. Ostrem, P. Strop and M. Lebl, *Proc. Natl. Acad. Sci. USA* **93**, 8194-8199 (1996).
7. M. Renil and M. Meldal, *Tetrahedron Lett.* **36**, 4647-4650 (1995) and references there cited.
8. M. Meldal, I. Svendsen, K. Breddam and F.-I. Auzanneau, *Proc. Natl. Acad. Sci. USA*, **91**, 3314-3318 (1994).
9. I. Schechter and A. Berger, *Biochem. Biophys. Res. Commun.* **27**, 157-162 (1967).
10. M. Marquart, J. Walter, J. Deisenhofer, W. Bode and R. Huber, *Acta Crystallogr.* **B39**, 480-490 (1983).
11. P. J. McCann III, D. R. O'Doyle II and I. C. Deckman, *J. Virology*, **68**, 526-529 (1994).
12. K. Burgess, A. I. Liaw, and N. Wang, *J. Med. Chem.* **37**, 2985-2987 (1994).
13. E. Atherton, C.J. Logan and R.C. Sheppard, *J. Chem. Soc., Perkin Trans **1**,* 538 (1981).
14 I. Schechter and A. Berger, *Biochem. Biophys. Res. Commun.* **32**, 898-902 (1968).