Innovation and Perspectives in

SOLID PHASE SYNTHESIS

Peptides, Proteins and Nucleic Acids

Biological and Biomedical Applications

1994

Editor: Roger Epton

Collected Papers
Third International Symposium
31st August - 4th September, 1993
Oxford, England, UK

Mayflower Worldwide Limited, Birmingham, 1994

34

ONE BEAD - ONE STRUCTURE LIBRARIES

Michal Lebl[1], Viktor Krchňák[1], Nikolai F. Sepetov[1], Victor Nikolaev[1], Alena Stierandová[1], Pavel Šafář[1], Bruce Seligmann[1], Petr Štrop[1], David Thorpe[1], Stephen Felder[1], Douglas F. Lake[2], Kit S. Lam[2] and Sydney E. Salmon[2]

[1] Selectide Corporation, 1580 E. Hanley Blvd., Tucson, AZ 85737, U.S.A.
[2] Arizona Cancer Center and Department of Medicine, University of Arizona College of Medicine, Tucson, AZ 85724, U.S.A.

Recent developments in the application of the Selectide Process to the discovery of new structures are described. Several questions are addressed: (i) screening of incomplete libraries; (ii) synthesis of libraries displaying both carboxy and amino termini; (iii) synthesis of libraries for screening performed in solution; (iv) structural determination of nonpeptidic compounds in the library by the use of coding principle, and (v) the synthesis and use of nonpeptidic libraries for screening.

INTRODUCTION

The Selectide Process is one of the library techniques (1-5) which offers a powerful tool to rapidly generate and screen a diversity of chemical compounds, both peptides and non-peptides. It is based on: (i) synthesis of a multiplicity of diverse compounds on polymeric beads, with a single structure on each particle (3,6,7), (ii) screening this library in either a binding assay based on an enzyme-linked or fluorescent tag (3,6,8), or in standard solution-phase assays (after partial release of a defined portion of the compound from each bead into solution (9-12)), and (iii) determination of the structure on the "active" bead. Compound composition is determined either by sequencing peptides by Edman degradation or using a mass spectroscopic technique (13).

STRUCTURAL MULTIPLICITY

The number of structures in peptide libraries screened in one experiment ranges from 1 to 10 million. Therefore, libraries longer than pentapeptides composed of 20 natural amino acids are incomplete. If it is necessary to test longer peptides, this physical limitation enforces the synthesis of randomly representative yet incomplete libraries. Published structure-activity studies in the field of biologically active peptide molecules have shown that there is generally a limited number of two to four "critical" residues in the molecule and the rest of the structure can be replaced by various amino acid residues or non-peptide templates. The theoretical advantage of long chain libraries is that multiple interacting residues may be present in one peptide with the appropriate spatial separation to permit all of them to interact with the target and provide sufficient binding for detection. For any library the frequency at which a critical sequence motif occurs can be used for calculation of the number of critical residues within this motif:

$$n_{crit} = \frac{\log(sample)}{\log(aa)} - \frac{\log(hits)}{\log(aa)}$$

where n_{crit} is the number of "critical" residues, *sample* is the number of screened beads from a given library, *hits* is the number of positively identified beads, and *aa* is the number of amino acids (or other building blocks) used for randomization. As can be seen, the result of this calculation does not

depend on the length of the library. "Critical" for interaction refers to a certain range of binding constants (to observe a binding constant several orders of magnitude tighter, more critical residues may be necessary). The formula given above is based on the assumption that positions which are not important for binding (in given range of binding constants) can be replaced by any amino acid used in randomization without loosing experimentally observed binding. This obviously may not be true for all residues. For example, position which can be occupied by two different residues is represented by a value of 0.5 critical residues and a peptide containing four of such residues will provide the same n_{crit} as a peptide with two unreplacable residues.

Depending upon assay sensitivity, only a portion of the critical motif needs to be present in the peptide. From such leads containing "incomplete pharmacophores" dedicated libraries with a lower number of permutated positions around these "lead" pharmacophore centers can be synthesized to give more functionally relevant "complete" library information. This approach was demonstrated using the anti-insulin monoclonal antibody, where decapeptide structures which mimicked the discontinuous insulin binding epitope were selected from an incomplete library of octapeptides (9). Based on the structure of the decapeptide motifs, tetradecapeptide libraries were constructed and screened. Figure 1 illustrates the strategy in constructing the libraries and Table 1 summarizes the affinities of selected ligands.

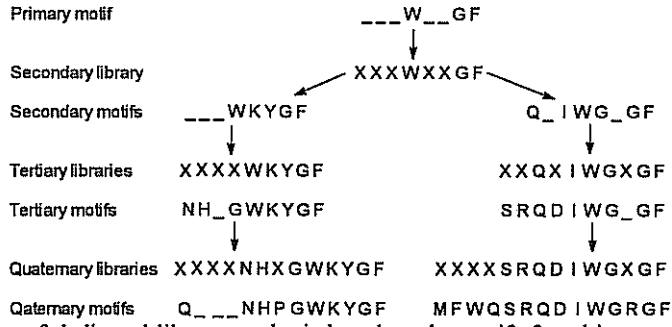


Figure 1: Scheme of dedicated library synthesis based on the motifs found in consecutive library generations.

Table 1: Binding affinities of anti-insulin MoAb specific ligands found in different libraries

Peptide	Library	IC_{50} (μM)
Insulin	natural ligand	0.014
QHIWGLGF	secondary	3.9
LNPWKYGF	secondary	2.3
NHIKWKYGF	tertiary	0.18
SKQDIWGRGF	tertiary	0.068
QSSVNHPGWKYGF	quaternary	0.035

VARIOUS LIBRARY FORMATS

Peptide libraries are usually synthesized attached to the polymeric carrier via its carboxy terminus. However, in certain cases, the free caboxy-terminus is essential for interaction with the macromolecular target. We have designed a simple way to construct peptide libraries with a free

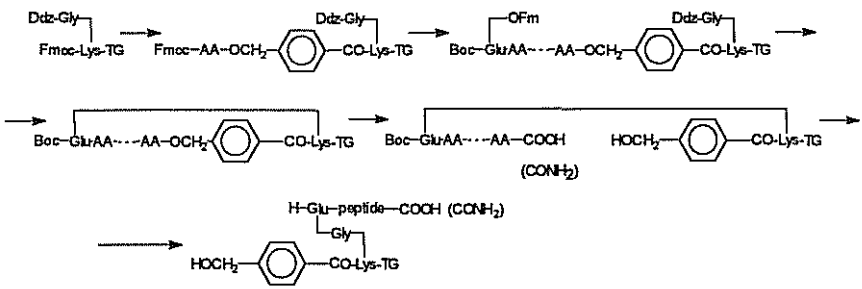


Figure 2: Synthesis of libraries with free carboxy terminus.

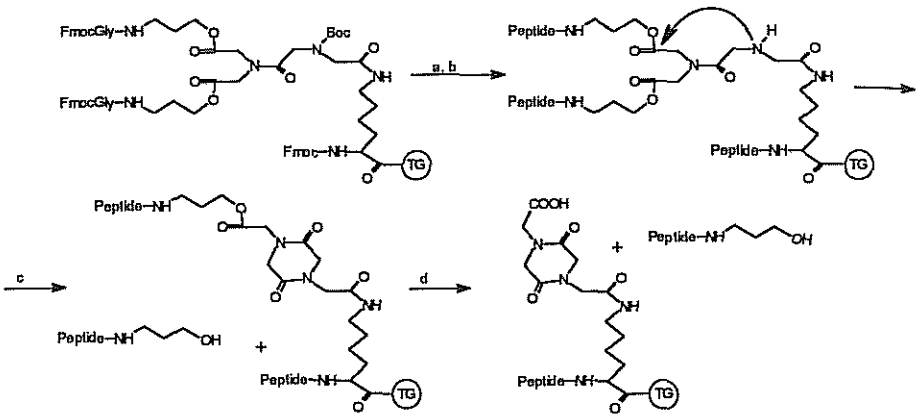
carboxy terminus, or libraries with both free amino and carboxy termini. The approach is illustrated in Figure 2. The peptide is synthesized on an alkali-labile linker. At the end of the peptide chain assembly (or in the middle, depending on the planned attachment point for the peptide to the polymeric carrier), selectively protected functional groups are unmasked, cyclization is performed, and the alkali-labile linker is cleaved. We have tested different linkers (acid sensitive, thiolytically cleavable) and chemistries for their cleavage after cyclization with the same success as with the benzyl ester linker which is given as an example in Figure 2. Since the determination of the structure is critical in the Selectide Technology, it is advantageous for the peptides to have a free amino terminus to permit Edman degradation. The chemistry depicted provides this additional benefit. A similar approach using different linker chemistry was recently presented (14).

RELEASE ASSAYS

Screening of ligands on beads utilizing binding assays are limited to screening against soluble macromolecular targets. To be able to utilize the variety of tests for screening against insoluble targets, it is necessary to release the structures generated in the library into solution. One polymeric bead contains approximately 100 pmol of peptide and therefore after its release to 100µl of solution we can obtain a 1µM solution. The affinities detectable in this type of an assay are therefore in the high nanomolar range. For screening libraries of several million structures it is impractical to release a single structure into each microtiter plate well. The synthesis of a multidetachable linker, which allows the release of the same amount of peptide in two distinct steps, allows the screening of a library in two steps. In the first step, mixtures of 500-1000 peptides are generated and tested. The beads from which "active" mixtures were released are replated in a second step, by one bead per well, and the second portion of peptide is released. The second biological test then identifies the single bead responsible for the activity and the rest of the peptide still attached to the bead is sequenced.

We have described the various chemical constructs which can be used as a multiply releasable linker (9-11), but an even simpler variant of the linker is shown in Fig. 3, where the principle of double release is given (12).

Validation of this method of screening was provided using the gpIIb/IIIa receptor as the target. A library of cyclic pentapeptides with three randomized positions (using 18 L, 18 D amino acids and glycine) within the ring formed by cystine was screened. This library of 50,653 compounds (100,000 beads were screened) afforded three "positive" beads, the structure of which was found to be Cys-Arg-Gly-Asp-Cys (2x) and Cys-Ala-Arg-Tyr-Cys. Resynthesis of all possible diastereoisomers (4 for the first peptide and 8 for the second) have shown that sequence Cys-Arg-Gly-Asp-Cys containing all L amino acids was active (IC_{50} 1µM), and that Cys-Ala-Arg-Tyr-Cys sequence was a false positive result. Using above given formula n_{crit} value calculated for this experiment is 2.99, which corresponds closely with results from structure-activity studies (Arg-Gly-Asp is the critical sequence).



a) Peptide synthesis; b) TFA; c) Buffer pH 8.5 (first release); d) 0.05% NaOH or NH_3 (gas) (second release)

Figure 3: Scheme of synthesis and two step release of peptide from library bead.

STRUCTURAL CODING

Coding for the structure of an active molecule is not itself a novel concept. The phage display technique uses the nucleic acids of the phage to code for the sequence of the interacting peptide. Sequencing of nucleic acids attached directly to peptide molecules was suggested (15) for determination of a structure of a peptide in a library. Coding for a nonsequenable structure by a peptide molecule was recently suggested by two groups (16,17).

The principle of the coding technique is illustrated in Figure 4. The "screening" structure may contain nonsequenceable amino acids, or it may be constructed by connecting various building blocks using different chemistries not compatible with Edman degradation. In any case, the sequence information from the coding sequence is sufficient to determine the structure of the screening arm. The building blocks in the screening arm must be unequivocally associated with the amino acids and positions in the coding arm. To cover a wider range of building blocks in the screening arm, more than one amino acid per building block can be used for coding. Thus, using doublets of 20 natural amino acids, 400 unnatural building blocks can be encoded.

There are several approaches to assemble the coding sequence. An optimal arrangement is the physical separation of the screening structure and coding peptide, which we describe in another article in this volume. In this case the screening structure is displayed on the surface of the bead where it is available to interact with the macromolecular target, while the coding peptide is present only inside of the bead. In the applications where the release of the screened peptide into solution is used (as

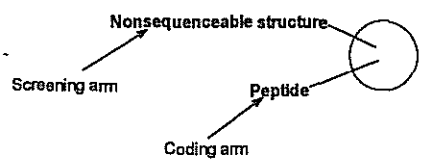


Figure 4: Construction of two structures on one bead. Coding arm can be sequenced and screening arm structure deduced.

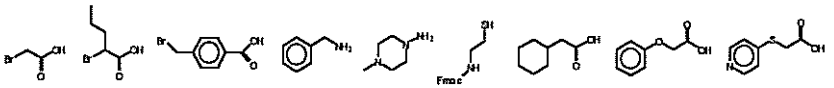


Figure 5: Building blocks used in the synthesis of model nonpeptidic coded library.

just described), the localization of screening and coding compound on the bead is of no concern, since the coding sequence is never released into the solution.

We have validated the coding principle by constructing a minilibrary of nonpeptide structures in parallel with the coding sequence. The building blocks chosen for this library are shown in Fig. 5 and representative nonpeptidic structures are shown in Fig. 6 together with their coding peptide. The structures deduced from sequencing data were confirmed by mass spectrometry, possible because all the nonpeptidic building blocks had unique molecular weights. Construction of the non peptide screening molecule involved (i) attachment of a bromo substituted carboxylic acid to the amino group on the solid carrier, (ii) alkylation of an amino or thiol group of an amine or N-protected aminomercaptan, and (iii) acylation of the newly generated amino group by a derivative of carboxylic acid. Introduction of each unnatural building block into the screening structure was followed (or preceded) by the coupling of a coding amino acid to the other arm of the molecule. We have used only glycine, alanine and leucine for coding. These amino acids therefore encoded a different structural element in every step of the randomization.

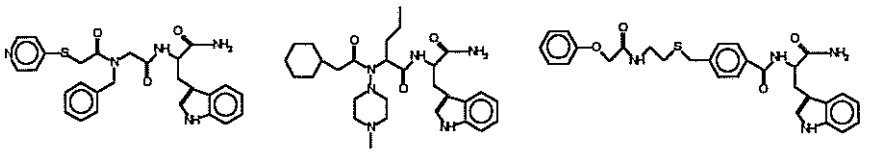


Figure 6: Representative structures from nonpeptidic library, their peptide coding, and molecular weight determined by mass spectroscopy.

NONPEPTIDE LIBRARIES

The importance of the construction and screening of non-peptide libraries for the development of specific ligand for a macromolecular target is obvious. Ideally the building blocks used should be capable of all major types of interactions (ionic, hydrogen bonding, hydrophobic, charge-transfer, chelation, aromatic, etc.), and be arranged appropriately in space. The coding principle allows the generation, screening and structure determination of hits from nonpeptidic libraries. However, the use of nonpeptidic building blocks for library generation requires optimization of chemistries used for their linking, chemistries which were not developed for solid phase applications. An example of a

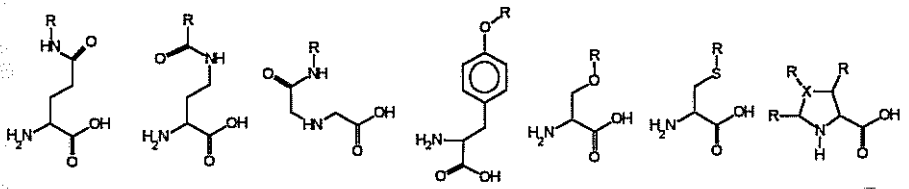


Figure 7: The simplest building stones for construction of nonpeptide libraries.

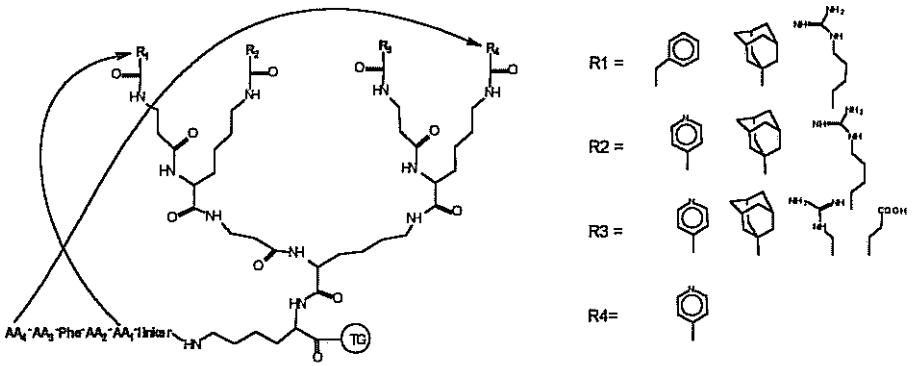


Figure 8: Structure of the library and hits identified in screening against streptavidin.

simple approach is the application of peptide chemistry to synthesize a template (scaffold, skeleton) onto which a variety of available building blocks are attached using chemistries well developed for solid-phase application. The biggest advantage of this approach is the large selection of building blocks (approximately 1500 amines and carboxylic acids can be obtained directly from commercial sources) combined with the simplicity of their introduction into the library format. The simplest building blocks are given in Figure 7. Side chains of aminodicarboxylic acids can be modified by various amines, diaminocarboxylic acids can be acylated or alkylated (18), as well as sulfhydryl or hydroxy group containing amino acids. Reactions of isocyanates, isothiocyanates, aldehydes, halides, or compounds with multiple bonds can be employed for multiplicity generation. Figure 8 shows an example of hits found from a streptavidin screen using a library randomizing carboxylic acids attached to a branched peptidic scaffold. These are significantly different from the natural ligand biotin or from the motif found in peptide libraries containing L amino acids (Leu-His-Pro-Gln-Phe (3,8)), D amino acids (D-Trp-D-Tyr-D-Gln-D-Glu-D-Ala), or their mixtures (e.g. Trp-D-Lys-Trp-D-Pro-His (19)).

REFERENCES

1. M.R. Pavia, T.K. Sawyer and W.H. Moos, *Bioorg. Med. Chem. Lett.*, 1993, **3**, 387.
2. M.H. Geysen, H.M. Rodda and T.J. Mason, *Molecular Immunology*, 1986, **23**, 709.
3. K.S. Lam, et al., *Nature*, 1991, **354**, 82.
4. R.A. Houghten, et al., *Nature*, 1991, **354**, 84.
5. S.P.A. Fodor, et al., *Science*, 1991, **251**,763.
6. K.S. Lam, et al., *Bioorg. Med. Chem. Lett.*, 1993, **3**,419.
7. A. Furka, F. Sebestyen, M. Asgedom and G. Dibo, *Int. J. Peptide Protein Res.*, 1991, **37**,487.
8. K.S. Lam and M. Lebl, *Immunomethods*, 1992, **1**, 11.
9. M. Lebl, et al., *Int. J. Peptide Protein Res.*, 1993, **41**, 201.
10. M. Lebl, et al., In *"Peptides 1992"* (Schneider, K.C. and Eberle, A.M., eds.), 1993, p. 66.
11. S.E. Salmon, et al., *Proc. Nat. Acad. Sci. U.S.A.*, 1993, in press.
12. P. Kočiš, V. Krchňák and M. Lebl, *Tetrahedron Lett.*, 1993, Submitted.
13. N.F. Sepetov, et al., *Rapid Comm. Mass Spectrom.*,1993,**7**,58.
14. C. Holmes and C.M. Rybak, *13th Amer. Peptide Symp.*, June 1993, Edmonton, poster P911.
15. S. Brenner and R.A. Lerner, *Proc. Natl. Acad. Sci. USA*, 1992, **89**,5381.
16. V. Nikolaiev, et al., *Peptide Res.*, 1993, **6**,161.
17. J.M. Kerr, S.C. Banville and R.N. Zuckermann, *J. Am. Chem. Soc.*, 1993, **115**,2529.
18. R.J. Simon, et al., *Proc. Natl. Acad. Sci. USA*, 1992, **89**,9367.
19. K.S. Lam., et al. in *Proc. 13th Amer. Pept. Symp.*, in press.