Bioorganic & Medicinal Chemistry Letters, Vol.3, No.3, pp. 419-424, 1993
Printed in Great Britain

0960-894X/93 $6.00 + .00

THE CHEMICAL SYNTHESIS OF LARGE RANDOM PEPTIDE LIBRARIES AND THEIR USE FOR THE DISCOVERY OF LIGANDS FOR MACROMOLECULAR ACCEPTORS

Kit S. Lam[+], Victor J. Hruby[*], Michal Lebl[‡], Richard J. Knapp[‡], Wieslaw M. Kazmierski[*‡]
Evan M. Hersh[+] and Sydney E. Salmon[+]

Department of Chemistry[*] and Arizona Cancer Center[+], University of Arizona,
Tucson, AZ 85721; and Selectide Corporation[‡], 10900 N. Stallard Place,
Suite 122, Tucson, AZ 85737

Abstract: A method is outlined for preparing large, diverse peptide libraries (10^5-10^7 peptides) such that there is one peptide per bead. These libraries can be used to screen for binding to macromolecular receptor (acceptor) molecules and to determine the structure of the peptides that bind.

Introduction

A central goal of modern molecular biology and drug design is the development of more rapid and rational approaches to lead discovery and development. A central and universal mechanism utilized by biological systems to control and modulate the biochemical events necessary for cellular structure and function is molecular interactions. These interactions are critical for both the conformational changes and the chemical-structural changes that are central to functional modulation in biological systems. In this regard, peptides and proteins, often modified as glyco- or lipo-conjugates, constitute the vast majority of such "molecules of life" whether they be enzymes, receptors, growth factors, hormones, neurotransmitters, antibodies, immunomodulators, or one of the many other proteins or peptides so critical to cellular function and to intercellular communication in complex living systems. Nature's choice of proteins as the major carriers of biological structure and function undoubtedly is related to the enormous diversity inherent in protein structure and function. The extent of this diversity can best be appreciated by reference to Table 1. As can be seen, even if one considers only Gly and the 19 L-amino acids generally incorporated into proteins by transcription, already at the hexapeptide level millions of separate chemical species are available. If one could incorporate other amino acids into polypeptides (say a very conservative list of 50 such amino acids), at the hexapeptide level, billions of separate chemical entities are possible. The possibility for utilizing such chemical diversity in the search for molecules (e.g. agonist and antagonist drugs) with specific biological properties provides both an opportunity and a challenge. The challenge is to develop methods to prepare very large and diverse libraries in a form suitable for evaluation of biological receptor (acceptor) molecules and systems, in sufficient quantity to allow structural determination, and at a cost in effort and chemicals acceptable to research and development organizations. Recently we have proposed and developed a methodology for preparing large statistically diverse peptide libraries (10^4-10^8) and for examining their ligand-binding activity with biologically relevant macromolecules such as antibodies, binding proteins,

Table 1. Chemical Diversity Inherent in Native and Designed Polypeptides

Dipeptide-20^2	=	400	50^2	=	2,500
Tripeptide - 20^3	=	8,000	50^3	=	125,000
Tetrapeptide - 20^4	=	160,000	50^4	=	6,250,000
Pentapeptide - 20^5	=	3,200,000	50^5	=	312,500,000
Hexapeptide - 20^6	=	64,000,000	50^6	=	1.565×10^{10}
Heptapeptide - 20^7	=	1,280,000,000	50^7	=	7.8125×10^{11}

enzymes, etc.[1] Others also have proposed different biological and chemical approaches to examine aspects of polypeptide diversity for drug discovery.[2-8] In this article we further examine the applicability of our method for the discovery of peptides that bind specifically to macromolecular acceptor proteins and discuss the implications of these findings for future developments.

Results and Discussion

The construction of a large pentapeptide library consisting of up to 2,476,099 (19^5) individual peptides which utilizes all of the 20 usual amino acid residues that are incorporated into proteins (except Cys) has been previously outlined.[1] The basic synthetic approach utilized standard methods of solid phase peptide chemistry optimized to ensure completion of the coupling of each new residue at each step. The procedure consists of dividing the resin into (in this case 19) equal and separate fractions for each coupling step, and coupling a single amino acid with each fraction to completion. Each coupling step is followed by thoroughly mixing all fractions and dividing them into 19 equal parts before the next coupling step such that a completely random, Poisson distribution of sequences is obtained. A related method for preparing multicomponent peptide mixtures of 27 tetrapeptides and 180 pentapeptides has been developed independently by Furka et al.,[9] and the method of Houghten et al.[8] also utilizes some aspects of our approach. We recognized that our methodology could provide a very large and diverse library for independent biological evaluation since each solid phase bead should contain only a single unique peptide. Of course, to be useful for biological screening and identification of the structure(s) that interact with a biological acceptor molecule, each bead must contain sufficient peptide to allow screening (binding, partial release for a bioassay, etc.) and structure determination. In considering these issues, several questions presented themselves. First, what size bead is required? Second, what is the loading necessary for both screening and structure determination? How should the peptide be constructed so that it can be retained on the bead and still be available for binding to a large macromolecular acceptor or perhaps even a cell? What sort of screening methods will be compatible with such a construction? How can one insure that each peptide in the library will be available and have an "equal opportunity" for interacting with the acceptor molecule? These and a number of other questions are being addressed in our laboratories, and already it is clear that the answers will depend on the problem posed. In the remainder of this article we further examine these questions utilizing a monoclonal

antibody (MAb) against the N-terminal of β-endorphin (clone 3-E7, Boehringer Mannheim, Indianapolis, IN) and a pentapeptide library with the 19 standard amino acids discussed above.

Theoretical evaluation of the bead requirements led us to the following tentative conclusion. If a standard solid phase resin bead of 100 to 200 μm containing a substitution level of 0.2 to 0.5 mmoles of amino acid (or peptide) per gram of resin is used (this is the size of bead and substitution level commonly employed in SPPS), substitution levels of 50-500 picomoles of peptide should be present on each bead. To determine the validity of these calculations, we examined a number of single resin beads randomly chosen from a pentapeptide library constructed using the 19 standard amino acids (except Cys) found in proteins. It is possible to readily sequence a single bead using a model 477A-01 Pulsed Liquid Automatic Peptide Sequencer (Applied Biosystems, Foster City, California). We found that each bead did contain between 60 and 300 pmoles of peptide as predicted. Since this sequencer can readily measure 2-5 picomoles of peptide for sequencing, it was possible to examine in great detail the fidelity of the synthetic peptide on each bead using modern preview analysis methods.[10] These studies have demonstrated that the pentapeptides on each bead are 97-99.9% homogeneous, and that even in the worst case, the minor byproducts appear to be primarily deletion sequences (it is possible that small amounts (less than 1%) of otherwise modified peptides are present, but thus far, none have been detected by sequence analysis). These results clearly indicate that the "standard" resins used in many applications of SPPS are suitable for synthesis of large diverse peptide libraries. Further calculations suggest that the local surface concentration on such beads may be in the mmolar to μmolar range thus increasing the chance for success in discovering even ligands that bind weakly to acceptor proteins.

Having established that "standard" resins for SPPS are suitable for construction of a large, diverse and chemically "clean" peptide library, we next turn to the examination of a peptide bead library suitable for screening. The requirements are the following: 1) optimized solid phase synthesis methods should be used; 2) side chain protecting groups on trifunctional amino acids must be removable without removing the peptide from the bead and without side reactions as a result of removing these temporary protecting groups; 3) the peptides at the surface of the bead must be accessible for binding interaction with the macromolecular acceptor, a Mab in this example. For this purpose, we have used a polyamide resin which was modified to contain various "spacer" groups prior to synthesis of the desired peptide library. In this particular case, we constructed a spacer that consisted of β-alanine, ε-aminocaproic acid and ethylenediamine. Then the 19^5 peptide library was prepared using the above methods. N^{α}-Fmoc groups were removed with 20% piperidine, and after the synthesis was completed using 5 randomized coupling steps, the side chain protecting groups were removed by 90% TFA in DMF (v/v) containing 1% anisole and 0.9% ethanedithiol. The resin was then neutralized with 10% diisopropylethylamine in DMF before use in screening.

A rapid method suitable for screening this library was developed using an enzyme-linked-immunoassay method in which the acceptor molecule, in this case, a particular MAb, was coupled to the enzyme alkaline phosphatase, which could then catalyze a colored reaction such that those beads that

Table 2. Affinity of Peptide Ligands to Anti-β-Endorphin MAb (Clone 3-E7)

Peptide	IC_{50} nM[a]			
	-OH	$-NH_2$	-β-Ala-OH	β-Ala-NH_2
YGGFL[b] (native ligand)	17.5	27.3	17.1	13.7
YGGFA	32.9	72.0	82.3	93.6
YGGFT	36.9	65.2	50.6	43.3
YGGFQ	15.0	40.1	39.4	45.4
YGGLS	726	991	916	1150
YGALQ	1980	2910	1470	2870
YGGMV	8780	14000	5140	7160

[a]IC_{50} values were obtained in solution by solution phase radioimmunoassay using [^{3}H][Leu^5]enkephalin as the labeled ligand

[b]YGGFL is leucine-enkephalin; the other ligands were those detected and structurally determined in the screening process

interacted with the MAb would become colored (in a sea of beads that were colorless), and could easily be removed from the incubation medium by a micromanipulator under a low power dissecting microscope. The methodology is outlined in our published paper.[1]

The sequences of the seven beads selected are given in Table 2. We then decided to examine how well these peptides bind compared to the native epitope for the MAb (the native N-terminal peptapeptide of β-endorphin is YGGFL - leucine-enkephalin), and whether the spacer had affected the results obtained. We, therefore, synthesized quantities of each pentapeptide selected in their carboxylate and carboxamide C-terminal forms, and in addition, synthesized the pentapeptides with β-alanine and β-alanineamide, respectively, added to the C-terminus of the pentapeptide sequence. Comprehensive binding experiments were performed for each peptide. The results are given in Table 2. The results indicate that though some slight changes in the binding constant are seen, the net effect is rather minor and the binding seen on the bead is qualitatively not affected by its method of attachment to the bead at least when using the β-Ala spacer. (The use of other spacers and their compatibility with various screening methods and different target proteins is currently under investigation in our laboratories.)

Anti-β-endorphin antibody has served as a useful model system for developing this technology. We have identified a number of sequences that interact with this antibody using slightly different libraries that were prepared to contain a variety of random sequences. Some of the results obtained with these various

Table 3. Structures of Peptides in Random Pentapeptide Libraries that Interact with Anti-β-Endorphin MAb. Sequences Marked by the Star Were Identified More Than Once.

Group I	Group II	Group III	Group IV	Group V	Group VI	Group VII
YGAFF	YGGFA	YGNFF*	YGFFQ	YGEAF*	YKGGF	LYGGF
YGAFI*	YGGFI	YGQFV	YGWFH	YGGGF	YLGGF	MYGGF
YGAFM	YGGFL*	YGVFA	YGWFN	YGHAF	YQGGF	NYGGF
YGAFQ	YGGFQ	YGVFE	YGWFQ	YGLGF*		RYGLL
YGAFT	YGGFT*	YGVFI	YGWWM	YGMGF*		
YGALQ	YGGLS	YGVFQ	YGYFF	YGPGF		
YGALT	YGGMV		YGYWQ			
YGAWD						

pentapeptide libraries are given in Table 3. We also have synthesized some of these peptides to examine their solution binding affinities. Except those of groups I and II (for example YGAPM has a binding affinity of EC_{50} = 52 nM), none of the others show affinity at even 10 μM. Several conclusions can be drawn from these findings. First, our methods allow us to detect weakly binding as well as strongly binding ligands. This lends support to our suggestion that the "concentration" of a ligand on the "surface" of a bead can simulate high concentration of the ligand in solution. Second, new insights can be obtained about the structural requirements of the antibody. Moderate to strong binding only occurred with peptides having the structural motif YG(G/A)(F/L/M/W). The affinities of peptides from Group IV to VII are much weaker, and the reasons for this are under further investigation. Peptides of group VII are examples of the fact that tyrosine does not have to be the amino terminal residue for a pentapeptide to bind to this antibody, though admittedly it is a rare case when this is so.

Table 4 illustrates several sequences found in an octapeptide library that had a limited set of different amino acids (for example, it did not contain phenylalanine). The motif found, nonetheless, was similar to that found for pentapeptides, i.e. YG(A/G)(L/W).

Discussion and Summary

A simple, highly reproducible, and synthetically clean method for producing very large, statistically diverse random peptide libraries is now available for general development. The screening methods developed also are readily used to distinguish those beads that interact with the receptors or acceptor molecules. It is interesting to observe that peptides that bind to the MAb in this case have binding affinities ranging over nearly three orders of magnitude -- from the nanomolar to the micromolar range, and other binding ligands with even lower binding constants undoubtedly would have been detected if a high concentration of the MAb were used in the screening process and/or the MAb itself could interact more strongly with its "native" epitope. These results are consistent with our initial calculations that suggested that the local concentrations

Table 4. Structures of Peptides from a Limited Octapeptide Library that Interact with Anti-β-Endorphin MAb.

YGALGWGT YGALYGIT	YGGLAHYT YGGLGLPL	YGAWAEPN	YGAWITYT

of the peptides on the bead are in the millimolar to micromolar range. Work is in progress to develop more versatile linker systems which would alter the linker arm length and their chemical properties to make them, in principle,more biocompatible for a variety of other receptors, and for cells and other biological systems.

Acknowledgements

We thank Fahad Al-Obeidi, Michael Ross and Ralph Hirschmann for their helpful suggestions and constructive criticisms, and the technical assistance of Farid Abdul Latif, Shelly Wade and Ronald Ferguson. This work was supported in part by the National Institutes of Health, The Arizona Disease Control Commission and the Selectide Corporation. Kit S. Lam is a recipient of a Leukemia Special Fellow Award (1989-1992) and a Leukemia Scholar Award (1992-1997).

References and Notes

1. Lam, K.S.; Salmon, S.E.; Hersh, E.M.; Hruby, V.J.; Kazmierski, W.M.; Knapp, R.J. *Nature* **1991**, *354*, 82
2. Parmley, S.F.; Smith, G.P. *Gene* **1988**, *73*, 305.
3. Scott, J.K.; Smith, G.P. *Science* **1990**, *249*, 386.
4. Geysen, H.M.; Melven, R.H.; Barteling, S.J. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 3998.
5. Fodor, S.P. et al. *Science* **1991**, *251*, 767.
6. Cwirla, S.E.; Peter, E.A.; Barrett, R.W.; Dower, W.J. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 6378.
7. Devlin, J.J.; Panganiban, L.C.; Devlin, P.E. *Science* **1990**, *249*, 404.
8. Houghten, R.A.; Pinilla, C.; Blondelle, S.E.; Appel, J.R.; Dooley, C.T.; Cuervo, J.H. *Nature* **1991**, *354*, 84.
9. Furka, A.; Sebestyén, F.; Asgedom, M.; Dibó, G. *Int. J. Peptide Protein Res.* **1991**, *37*, 487.
10. Niall, H.D., Tregear, G.W.; Jacobs, J. In *Chemistry and Biology of Peptides*; Meienhofer, J., Ed.,; Ann Arbor Sci. Publ.: Ann Arbor, **1972**; pp. 975-984.