Bioorganic & Medicinal Chemistry Letters, Vol.3, No.3, pp. 397-404, 1993
Printed in Great Britain

0960-894X/93 $6.00 + .00

SCREENING CHEMICALLY SYNTHESIZED PEPTIDE LIBRARIES FOR BIOLOGICALLY-RELEVANT MOLECULES

H. Mario Geysen* and Thomas J. Mason

Chiron Mimotopes Pty Ltd, P.O. Box 40, Clayton, Victoria, 3168 Australia

Practical strategies for the synthesis and testing of large peptide libraries are described. The relationship between the number of monomers (amino acids) used to generate diversity and the magnitude of the synthesis and testing required is evaluated. Peptide optimization methods using strategies of replacement by non-natural and D-optical isomers are demonstrated.

With the rapid identification of genes coding for important human receptors, there is an increased emphasis on the search for lead molecules able to bind these receptors. Traditional methods use extensive libraries of compounds collected from natural sources or synthesized relatively slowly at great cost in terms of time and labour. Advances in techniques for the parallel synthesis of peptides, either individually or as grouped mixtures, provide an alternative strategy to these libraries. Whilst peptide mimetics may or may not be attractive themselves as drug candidates, they certainly represent a useful starting point on which to base a synthetic effort leading to novel drugs. The primary objective of screening a peptide library is to identify a small peptide that shares with the natural ligand the important attributes of the interaction with its antibody or receptor. Herein is an overview of a method for identifying binding peptides and the basis for a general procedure using a library comprising mixtures of peptides.

The interactions of peptides with monoclonal antibodies (MAb) as a model for receptors has been used extensively to evaluate strategies for the synthesis and testing of large peptide libraries.[1-8] An MAb that binds to a sequential epitope is an example of the simplest case. Here, testing a complete set of linear octapeptides, which are homologous with the sequence of the protein antigen, will identify what is arguably the single "correct" peptide-ligand for that antibody.[9-10] Using current procedures, the synthesis of a library or number of peptides required for this screen is a relatively trivial exercise, and is best done by the synthesis of individual peptides, Fig. 1. For example, to test an MAb against a linear epitope of an antigen of 1000 residues (and assuming a maximum epitope length of 8 amino acids) one would need only 993 peptides; whereas, the

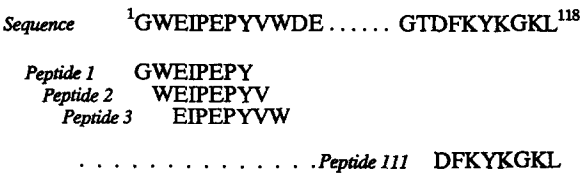

Figure 1. *Strategy for detecting a linear epitope. In this example, all 111 octapeptides from myohemerythrin (a 118 residue protein) are synthesized and tested.*

synthesis and testing of the complete set of octapeptides requires 25.6 billion (L-isomers of the common amino acids only). Fig. 2a shows a typical result for the case where the MAb binds to a well defined linear or continuous epitope, and Fig. 2b shows a comparable outcome obtained from one of the integrin receptor family members (IIb/IIIa), namely the binding of a linear epitope from a known ligand, fibrinogen.

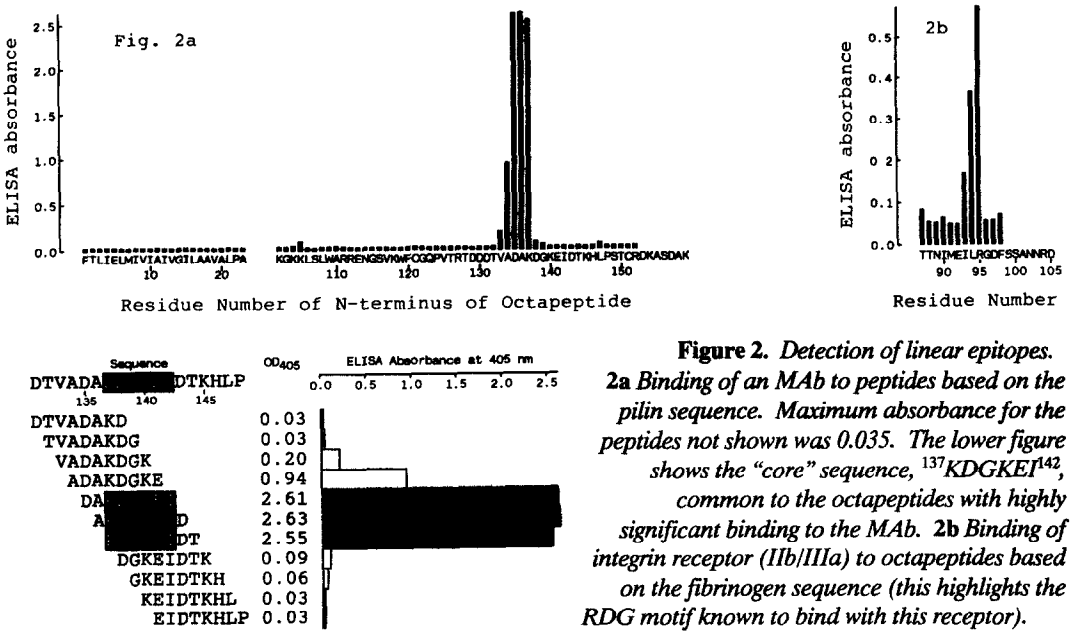

Figure 2. *Detection of linear epitopes.* **2a** *Binding of an MAb to peptides based on the pilin sequence. Maximum absorbance for the peptides not shown was 0.035. The lower figure shows the "core" sequence,* ^{137}*KDGKEI*^{142}*, common to the octapeptides with highly significant binding to the MAb.* **2b** *Binding of integrin receptor (IIb/IIIa) to octapeptides based on the fibrinogen sequence (this highlights the RDG motif known to bind with this receptor).*

Clearly a completely different strategy is required for the general case where either the antigen/ligand is unknown or the MAb/receptor recognizes a discontinuous or assembled epitope. In the context of antibody binding peptides, the average epitope (MAb) comprised six amino acids of which five showed limited replaceability and were considered to be essential or critical residues, and the longest linear epitope was eight amino acids long.[11] On this basis we proposed to synthesize a general screen of octapeptides in the format $\otimes\otimes\otimes D_1D_2\otimes\otimes\otimes$ where D_1 and D_2 are residue positions comprising single and defined amino acids, and $\otimes$ represents positions where residues are randomly incorporated by using a reaction mixture comprising each of the 20 common amino acids.[1,12-15] This procedure was named the "mimotope strategy" as peptides identified by its use could only be said to mimic the epitope.[1]

With an emphasis on identifying small peptide ligands, the present preferred mimotope strategy is based on a hexapeptide unit, i.e., $\otimes\otimes D_1D_2\otimes\otimes$. The size of the required set is 400 (20×20 preparations), each of which is now a mixture of peptides, rather than individual peptides; ideally, each consists of 160,000 different peptides (assuming equal incorporation of the residues). Using the pin technology, a convenient handle for both the synthesis and testing of large numbers of peptides,[9-10] the present derivatization level of 100 nmol per pin provides for an average of about 4×10^{11} copies of each individual peptide. This is comparable with

the 10^{12} to 10^{13} copies of epitopes found in a typical ELISA test with MAbs where 2-20 pmole of antigen is coated to the well.

Testing of this set of 400 peptide mixtures with an MAb/receptor in an appropriate test, will then only provide information about the identity of two of the amino acids, i.e., those residues incorporated at the two defined positions. A further round of synthesis based on this outcome in which the number of mixture positions is decreased by one, generates a new screen that allows the resolution or absolute identification of a further residue. The format usually used for the second screen (total of 40 peptide mixtures) is ⊗DXY⊗⊗ and ⊗⊗XYD⊗, where XY are the amino acids identified in the primary screen, and D equates to the position where single residues instead of mixtures are now incorporated. Whilst in principle the above step could simultaneously resolve the identity of two amino acids, this would only be the case where the residue at each position within the peptide is independent of the residues at other positions. In a number of cases we have observed the same or conserved amino acids selected as optimum at both the amino and carboxy terminal side of the currently defined residues, a result not supported by the next round of synthesis and testing. Therefore a more conservative approach is recommended in which only the residue corresponding to the single highest assay result from the screen of 40 peptide mixtures is used as the basis to synthesize the next screen. Clearly, repeating the cycle of synthesis and testing until either no further increase in the assay result is observed completes the identification of a binding peptide (mimotope), which may or may not share homology with the residues comprising the natural epitope. Table 1 shows the results obtained from the progressive delineation of a binding peptide for an MAb known to bind a linear epitope, based on a hexapeptide primary screen format of ⊗⊗D_1D_2⊗⊗.

Table 1.
Resolution of a Mimotope

Stage	No. of peptides made	Highest response	ELISA abs.¶ 45	10
Primary screen	400	⊗⊗LE⊗⊗	0.78	0.32
Second screen	40	⊗FLE⊗⊗	>2.50	1.20
		⊗⊗LEK⊗	1.45	0.61
Third screen	40	DFLE⊗⊗	>2.50	1.61
		⊗FLEK⊗	2.44	1.16
Fourth screen	20	DFLEK⊗	>2.50	1.16
Fifth screen	20	DFLEKI	>2.50	2.12
		⊗⊗⊗⊗⊗⊗	0.07	0.05

¶ ELISA absorbancies are given for incubation times with chromogenic substrate of 45 minutes and 10 minutes

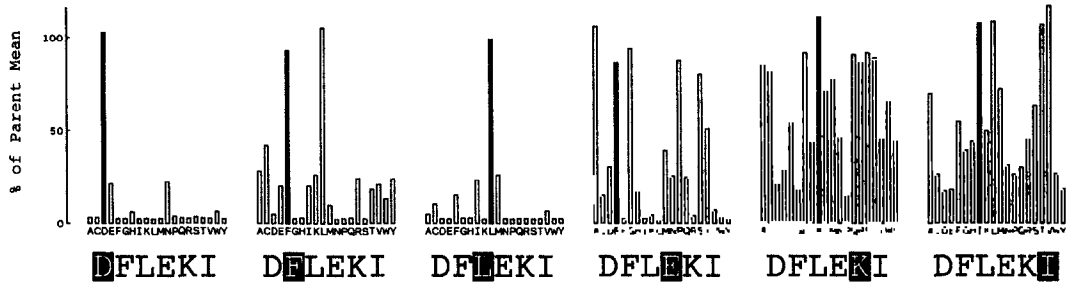

Figure 3. *Single amino acid substitution in an epitope, DFLEKI, of an anti-myohemerythrin MAb. Each block of 20 ELISA values represents the binding to peptides containing the single amino acid substitution identified by the single-letter code beneath each bar. The position of the substitution is indicated by the residue in inverse print in the sequence below each group. The parent sequence is shown as a black bar in each group. ELISA values are expressed as a percentage of the mean of binding with the six copies of the parent sequence.*

The several stages shown for the delineation of a binding peptide are outlined in Table 1. A general progression on the basis of the assay result is obtained, and demonstrates the definition of the sequence equivalent to the epitope, as originally identified by testing the smaller screen of linear peptides homologous with the antigen. The exception to a completely monotonic progression is associated with the substitution of the mixture at position 5 by lysine (K). In practice it has so far always been the case that at each stage of the delineation there are several options in terms of amino acid sequence which give comparable assay results.This necessitates a decision about whether one or more sequences (results) are selected for further resolution. If amino acid identity at each position in the sequence was fully independent, then some convergence would occur toward a set of solutions (peptides) consistent with all those "allowed" sequences defined from the set of single residue substitutions as shown in Fig. 3. However, it can be readily shown that identity at each position is not independent (H.M. Geysen, *et al.*, manuscript in preparation), and a divergent solution pathway leading to multiple local optima is therefore observed. A similar outcome has been found for the results obtained by screening phage libraries for binding peptides, where very different sequences, beyond those related by single residue substitutions, were readily identified.[8] Clearly, the uncertainty about having included the global optimum in the set of solutions decreases as the number of intermediate or partial solutions are followed to completion from each stage of the delineation procedure.

The total synthesis and testing requirement for the application of the mimotope procedure to a single antibody/receptor in a hexapeptide context, selecting only the optimum at each stage, is $400+3\times40+20=540$ in number (20 common amino acids only). Selecting for example, the top two from each solution pathway, at every stage, and resolving each to completion, requires a total of $400+2\times40+4\times40+8\times40+16\times20=1280$ syntheses and tests. An alternative strategy is to select say the top four, irrespective of the solution pathway, which would then require a total of $400+4\times3\times40+4\times20=960$ synthesis and tests. Clearly a compromise between the number of partial

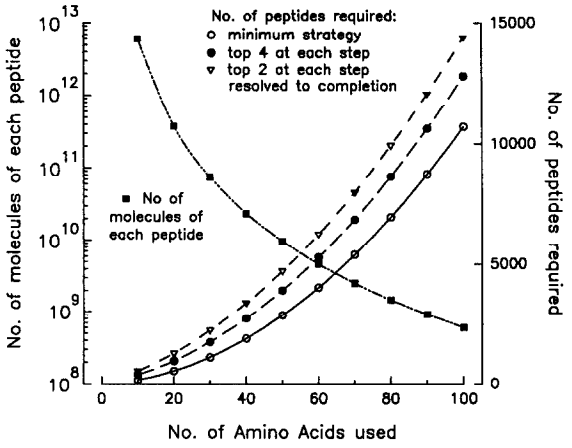

Figure 4. *The number of molecules of each individual peptide as a function of the number of monomers in the mixtures set (left hand axis). This represents the number of molecules of each peptide synthesized, assuming equi-molar coupling (total quantity of 100 nmol peptide), in the primary screen. Also shown is the number of peptide preparations that must be synthesized and tested to resolve a 6-mer mimotope for the three variations of the mimotope strategy (right hand axis) as discussed in the text.*

solutions resolved and the amount of syntheses and testing required is necessary, and the relationship between the two for the various strategies as described is shown in Fig. 4. Biologically generated peptide libraries are for now limited to the set of 20 genetically coded for amino acids, whereas chemically synthesized libraries are able to use a much larger repertoire of monomers, including the D-optical isomers corresponding to the common amino acids. However, increasing the set of starting amino acids very quickly escalates the magnitude of even the basic procedure in which only the optimum solution at each stage is selected for further screening. For example, including the D-optical isomers of the common amino acids into a minimum procedure requires a total of $39^2+3\times2\times39+39=1794$ syntheses and tests. Fig. 4 shows the relationship between the number of syntheses and tests required for each of the three options discussed

Table 2
Relative Reactivity on D-Optical Substitution of L-Amino Acids

No change	1 change	2 changes	3 changes	4 changes	5 changes
WQMGHS 1.000	wQMGHS 1.33	WQmGhS 1.31	wqmGHS 1.89	wqmGhS 0.90	wqmGhs 0.44
	WQmGHS 1.01	WQMGhs 1.06	WqmGhS 1.58	wqmGHs 0.43	
	WQMGhS 0.85	WQmGHs 1.17	wqMGhS 1.39	WqmGhs 0.36	
	WqMGHS 0.82	WqMGHs 0.86	wQmGhS 1.16	wQmGhs nt[¶]	
	WQMGHs 0.64	wQMGHs 0.83	wQMGhs 1.00	wqMGhs nt	
		wQMGhS 0.78	wQmGHs 0.89		
		wQmGHS 0.75	wqMGHs 0.88		
		WqmGHS 0.75	WqmGHs 0.81		
		WqMGhS 0.64	WQmGhs 0.33		
		wqMGHS 0.60	WqMGhs nt		

NOTE: ELISA responses are relative to the response with WQMGHS
¶ nt = not tested

above as a function of the number of monomers used. Also shown as a function of the number of monomers is the number of each of the different peptides ideally produced in each synthesis of the first screen. Clearly, a limit to the number of monomers included depends on the amount of MAb or receptor available (number of tests possible), the synthetic capability available (total syntheses required), and the sensitivity of the test (minimum amount of binding peptide present).

An alternative approach is to define one or more peptides using only the set of common amino acids, and then to further optimise these by screening substitution sets using either or both D-optical isomers or non-natural amino acids. This optimization strategy is equally applicable to peptides obtained from a screen of linear peptides homologous with a protein sequence as shown in Fig. 2. It has been common practice to screen analogues of biologically active peptides in which individual residues are substituted for by the D-optical isomer; this again assumes that each residue acts independently. A better strategy is to screen the complete set of D-optical isomers comprising substitutions at all single positions, all combinations of two positions, up to the fully substituted peptide. Table 2 summarizes the results of a binding assay, for a set of peptides comprising all combinations of D-optical isomer substitutions, for the case where the initial, completely delineated peptide had been progressively identified using only the set of 20 common amino acids.[1] What is evident is that the optimum peptide could not have been defined by the progressive accumulation of single substitutions showing increased binding. By extension, this observation would equally apply to the case where single substitutions only are made using an extended set of non-natural amino acids. One further useful strategy to optimize peptides after their *a priori* delineation, is to test for the effect of added flexibility introduced into the peptide by insertion of a non-alpha amino acid such as β-alanine. Alpha substituted amino acids impose constraints on the allowed conformations of the peptide which may preclude the optimum geometric arrangement of the constituent amino acids. Table 3 shows the effect on binding as a consequence of inserting either a glycine or β-alanine residue at each position between the residues in the peptide WQMGHS. A significant increase in binding to the defining MAb is observed for the peptide WQMβGHS, which was subsequently shown to be consistent with the peptide adopting a folded

Table 3
Relative Reactivity on adding "Spacer" Residues

Addition of Glycine		Addition of β-alanine (β)	
WQMGHS	1.000	WQMβGHS	2.38
WGQMGHS	0.93	WQMGHS	1.000
WQMGHGS	0.75	WQMGHβS	0.72
WQMGGHS	0.72	WQMGβHS	0.70
WQGMGHS	0.65	WβQMGHS	0.69
		WQβMGHS	0.46

conformation on binding.[1] Again, it could be argued that this constitutes a single site effect and assumes independence for each change.

The application of a general procedure for "fitting" by selection of a peptide to a binding molecule, for example, an MAb/receptor for which the normal ligand is a non-protein, would if successful translate from one class of chemical entity to another (to a peptide in this case). This in no way assumes that the interactions between the binding molecule and each of two dissimilar classes of ligands are the same, but rather that the requirements for binding could be satisfied in various ways. This "translation" was successfully demonstrated using a mimotope procedure, when small peptides were delineated that bound to monoclonal antibodies specific for blood-group antigens, well characterized polysaccharides.[12,15] Clearly, appropriate procedures for screening from large peptide libraries can in principle be equally applied to those molecules known to bind non-protein ligands such as DNA/RNA, lectins, steroid receptors and taste receptors to name a few.

In the absence of more rapid technology for solving the actual 3-dimensional structure of complexes between antigen/ligands and MAb/receptors, most new biologically active small molecules will continue to be identified by screening appropriate libraries. Modular synthesis strategies based on a diverse set of monomer units offer the greatest flexibility in terms of numbers of different compounds produced, and physico-chemical properties represented. At present, alternative chemistries to the amide linkage for the formation of a stable bond for compatible monomers sets are unavailable. However, the amide linkage used with an increased set of monomers (amino acids),[16] and including α-disubstituted and N-substituted monomers for increasing structure or alternatively the β-amino acids for reducing structure, provides for an enormous potential diversity. In addition, strategies have been developed (H.M. Geysen, *et al.*, manuscript in preparation) for the generation of chemical diversity based on various structural motifs, such as cyclized peptides (one to four loops), branched peptides and the inclusion of amino acid monomers onto rigid templates of various sizes. Clearly, present methods to generate chemical diversity exceed our ability to efficiently and practically screen these potentially large libraries. What is still needed is a better understanding of the components for an optimum library in terms of monomers, structures, and screening strategies, such that the global maximum is included in the outcome. One further goal should be to identify monomers associated with desirable properties, for example, bioavailability, and undesirable properties such as toxicity, and then to include or exclude these as appropriate in order to bias the outcome of the screen towards usable compounds.

Acknowledgement

We wish to acknowledge the work of the Synthesis and ELISA groups at Chiron Mimotopes, currently headed by Megan Jay and Sonja Plompen, respectively. Without the groups' skill, dedication and patience, much of the work reported here would not have been possible.

References and Notes

1. Geysen, H.M.; Rodda, S.J.; Mason, T.J. *Mol. Immunol.* **1986**,23, 709.

2. Cwirla, S.E.; Peters, E.A.; Barrett, R.W.; Dower, J. *Proc.Natl.Acad.Sci. USA* **1990**,87,6378.

3. Devlin, J.J.; Panganiban, L.C.; Devlin, P.E. *Science.* **1990**,249,404.

4. Fodor, S.P.A.; Read, J.L.; Pirrung, M.C.; Stryer, L.; Lu, A.T.; Solas, D. *Science.* **1991**, 251,767.

5. Houghten, R.A.; Pinilla, C.; Blondelle, S.E.; Appel, J.R.; Dooley, C.T.; Cuervo, J.H. *Nature.* **1991**, 354, 84.

6. Lam, K.S.; Salmon, S.E.; Hersh, E.M.; Hruby, V.J.; Kazmierski, W.M.; Knapp, R.J. *Nature.* **1991**, 354, 82.

7. Owens, R.A.; Gesellchen, P.D.; Houchins, B.J.; DiMarchi, R.D. *Biochem. Biophys. Res. Commun.* **1991**, 1,402.

8. Scott, J.K.; Smith, G.P. *Science.* **1990**, 249, 386.

9. Geysen, H.M.; Meleon, R.H.;Barteling, S.J. *Proc. Natl. Acad. Sci. USA.* **1984**, 81, 3998.

10. Geysen, H.M.; Rodda, S.J.; Mason, T.J.; Tribbick, G.; Schoofs, P.G. *J. Immunol. Meth.* **1987**,102, 259.

11. Geysen, H.M.; Mason, T.J.; Rodda, S.J. *J. Mol. Recog.* **1988**,1,32.

12. Geysen, H.M.; MacFarlan, R., Rodda, S.J.; Tribbick, G.; Mason, T.J.; Schoofs, P.G. *Towards Better Carbohydrate Vaccines;* Bell, R. and Torrigiani, G. Eds.; Wiley Chichester: **1987**,pp 103-118

13. Tribbick, G.; Edmundson, A.B.; Mason, T.J.; Geysen, H.M. *Mol. Immunol.* **1989**, 26, 625.

14. Edmundson, A.B.; Harris, D.L.; Tribbick, G.; Geysen, H.M. CIBA Foundation, London 1-5 July, **1990**.

15. Geysen, H.M.; Mason, T.J.; Tribbick, G.; Edmundson, A.B. Nordisk Insulin Symposium No. 2. *Molecular Mimicry in Health and Diseases*; Elsevier: Amsterdam, **1988**.

16. Roberts, D.C.; Vellaccio, F. *The Peptides*; Gross, E.; Meienhofer, J., Eds; Academic Press: New York, **1983**; Vol. 5B, pp 341-449.