The Journal of Microbiology, December 1997, p.347-353
Vol. 35, No. 4

Characterization of a Phage Library Displaying Random 22mer Peptides

Seung-Joo Lee[1], Jeong Hwan Lee, Brian K. Kay[3], Gideon Dreyfuss[4], Yong Keun Park[2], and Jeong-Kook Kim[2]*

[1]Department of Biology, and [2]Graduate School of Biotechnology, Korea University, Seoul 136-701, Korea
[3]Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3280, U.S.A.
[4]Howard Hughes Medical Institute, University of Pennsylvania, Philadelphia, PA 19104-6148, U.S.A.

(Received August 25, 1997 / Accepted November 21, 1997)

We have characterized a phage library displaying random 22mer peptides which were produced as N-terminal fusions to the pIII coat protein of M13 filamentous phages. Among the sixty phages randomly picked from the library, 25 phages had the 22mer peptide inserts. The DNA sequence analysis of the 25 inserts showed the following results: first, each nucleotide was represented almost equally at each codon position except that there were some biases toward G bases at the first position of the codons. Secondly, the expected 47 sense codons were represented. The deduced amino acid sequences of the 25 inserts were analyzed to examine its diversity. Glycine and glutamate were the two most overrepresented residues above the expected value, whereas cysteine and threonine residues were underrepresented. The range of diversity in dipeptide sequences showed that the amino acid residues were randomly distributed along the peptide insert. Acidic, basic, polar, and nonpolar amino acid residues were represented to the extent expected at most positions of the peptide inserts. The predicted isoelectric points and hydropathy indices of the 25 peptides showed that a variety of the peptides were represented in the library. These results indicate that this phage display library could be useful in fiuding ligands for a broad spectrum of receptors by affinity screening.

Key words: Phage display library, M13 bacteriophage, pIII coat protein, *E. coli*, peptide

The filamentous bacteriophages such as M13, fd, and fl consist of five different coat proteins and one molecule of single-stranded (ss), circular viral DNA. They are rod-shaped coliphages and can specifically infect male *Escherichia coli* by binding to F pili of host cells. The filamentous bacteriophage does not lyse the host cell but decreases its growth rate. Four of the five different coat proteins are the minor coat proteins, the products of genes III, VI, VII, IX, and each type is present at 3~5 copies per mature viral particle. The other type is the major coat protein, the product of gene VIII, and is present at approximately 2,700 copies per virus particle. The major coat protein is responsible for packaging the viral ss DNA. The pIII coat protein (product of gene III) at one end of the virus particle is responsible for proper phage assembly and infection by attachment to the pili protein of *E. coli* (9).

In a phage display technology, foreign DNA fragments cloned into the phage coat protein genes are expressed as a part of a fusion protein on the surface of the viral particles, thereby having two significant merits. First, since these phages could be applied to affinity matrices at high concentrations, large numbers of phages could be screened at one time. Second, if a particular phage is recovered from an affinity matrix in low yield, the phages could be amplified by infection (1, 2, 4, 10, 11, 12, 15). Smith and colleagues (15) constructed a hexapeptide library in which random hexapeptides were expressed as function proteins with the pIII proteins of fd phages. They screened the library with two different monoclonal antibodies (A2, M33) generated against the protein myohemerythrin (MHr) using the method of affinity enrichment, called panning. They found that phages displaying the specific antigenic inserts could be selected. Devlin and his colleagues (4) constructed a 15mer library expressed at the N-terminus of the pIII and screened the phage library with streptavidin, a non-peptide molecule, to isolate

* To whom correspondence should be addressed.
[1] Present address: Choongwae Pharma Co., Annyungri Taeaneup Hwasungkun, Kyungkido 146-141, Korea

binding phages. This result demonstrated that the recombinant phages could mimic the shape of non-peptide molecules. The octapeptide and 38mer peptide phage display libraries were also shown to be useful for selecting the phages that could bind the plant lectin concanavalin A and streptavidin, respectively (7, 12). The decapeptide phage display library was characterized for the biochemical diversity of its insert peptides, but not tested for its usefulness such as those of the phage display libraries described above (3).

The usefulness of the peptide phage display library depends on how rich a spectrum of biochemical structures the insert peptides represent. Therefore, we have examined the biochemical diversity in a sampling of clones in a phage library displaying random 22mer peptides prior to test for its usefulness.

Materials and Methods

Bacterial strains and a phage display library

The following *E. coli* strains were used as host cells for phages:

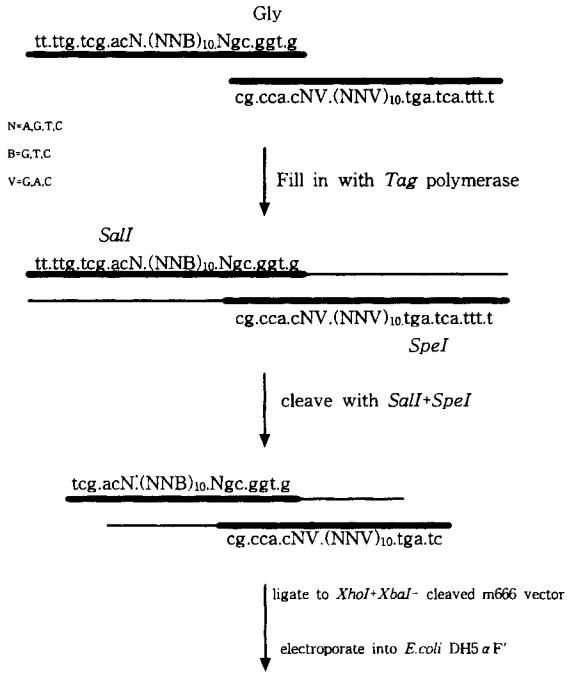
DH5αF' F'Φ80*dlacZ*ΔM15Δ(*lacZYA-argF*)U169 *deoR recA1 endA1 hsdR17*(r_k^-, m_k^+) *supE44*λ *thi-1 gyrA96 relA1*

MV1190 Δ(*lac-proAB*), *thi*, *supE*, (Δ*srl-recA*)306 ::Tn10(tet^r)[F':*traD36*, *proAB*, *lacI*^{q}ZΔM15]

A phage library displaying random 22mer peptides was constructed in Dr. Brian Kay's laboratory and named the TSAR-12 library. The construction strategy of the TSAR-12 library is shown in Fig. 1. The method for constructing the library was the same as that had been described previously (7). To summarize two degenerate oligonucleotides were annealed and extended with Taq DNA polymerase. Through sixty round of annealing and chain extension in the presence of dNTPs and Taq DNA polymerase, the two single-stranded oligonucleotides were converted into double-stranded DNA fragments. The DNA fragments were digested with restriction enzymes *Sal*I and *Spe*I. The digested DNA fragments were then ligated to the vector m666 cleaved with *Xba*I and *Xho*I restriction enzymes (5). The ligated DNA was ethanol-precipitated and electroporated into *E. coli* DH5α F' cells. Approximately 1×10^8 phage clones were generated. The recovered phages were concentrated to 4×10^{12} pfu/ml in 5% glycerol-containing LB medium and kept at -20°C.

Phage growth and its genomic DNA isolation

For phage growth and its genomic DNA isolation, Vierra and Messing's procedure (17) was followed



TSAR-12 library of pIII-random sequence fusion proteins

· · · S H S S (S/T) X_{10} Φ G δ X_{10} T R P A R T · · ·

↑

signal peptidase cleavage site

Fig. 1. The construction of a phage library displaying random 22mer peptides. Degenerate DNA inserts were prepared and cloned as described briefly in the Materials and Methods section. The arrow indicates the signal peptidase cleavage site at the N-terminus of M13 phage pIII coat protein. Amino acid sequences are shown in single letter symbols. The symbol Φ stands for any amino acid among Ser, Arg, Gly, Cys, or Trp. The symbol δ stands for any amino acid among Val, Ala, Asp, Glu, or Gly. Boxed residues, which include the random 22mer peptides, are absent from wildtype pIII. The symbol X10 indicates ten random amino acid residues.

with the following modification. In order to get phage plaques, *E. coli* DH5αF' cells were infected with the library phages and plated on LB top agar. Following overnight incubation at 37°C, individual plaques were picked to isolate phage genomic DNAs. To isolate the phage genomic DNA from the phage particles, *E. coli* MV1190 cells were used as hosts because they gave better DNA yield than the DH5αF' cells. A picked plaque was mixed with the MV1190 cells and incubated for 6hr in 5 ml of 2×YT media. Phage-containing supernatant was PEG-precipitated and the pellet was resuspended in 0.2 ml TE buffer.

DNA sequence analysis

DNA sequencing was performed according to

Sanger's method using Sequenase version 2.0 kit (USB) (14). In order to determine the nucleotide sequences of the 22mer inserts, an oligonucleotide primer (5'-TCCGTATGGGGTTTTGC-3') was used. It is located 35nt downstream of the *Xba*I cloning site of the vector m666.

Results and Discussion

As shown in Fig. 1, a phage library displaying random 22mer peptides was constructed. This phage library does not display random 22mer peptides in a strict sense, because two random 10mer peptides were connected through the Φ-G-δ tripeptides, in which Φ and δ stand for one of five different amino acid residues, respectively. Therefore, nucleotide and amino acid sequence diversities were determined for the two random 10mer peptides. However, we may say that this phage library displays random 22mer peptides in a general sense. Each variable amino acid in the random peptide sequence is encoded by a degenerate codon, either NNB or NNV, which includes any of the four nucleotides(N) in the first and second positions, and G, T, and C(B) or G, A, and C(V) in the third, respectively. Therefore, all 20 amino acids and one stop codon(UAG) are encoded by the 48 different codons represented by NNB or NNV.

Nucleotide sequence diversity of the library clones

Sixty phage clones were picked randomly from the phage library to examine for nucleotide sequence diversity. Initial attempts to sequence the library phage clones showed that some phage clones did not have the inserts. This led us to examine the restriction enzyme digestion profile of the sixty phage clones, which allowed us to distinguish the insert-containing phages from the other phages easily. It turned out that only 25 out of the sixty clones had the inserts (data not shown).

1. GAT TGG GAG AAT GGC GGG CGA GGC AGT GGT tgc ggt gtc AAG GGG CCG CAT GAT TAC CCT ATT TGC CCC
2. GGG GCG TTT CAG GGG TTT GAC CCC ACG AGT ggc ggt gcc CTG GCT GGG TTT CTT CCG TCC ACT GCC TCC
3. AGG GAG GGG ATT CGC AAT CTC GAT TGG ATG cgc ggt gtt CCC CAG TAC GGG TCA TAT GCT AAG CCC GAT
4. AAC GGG TGG TTT CAG AAG AGT GAG GTG TGG ggc ggt gga GCG ATC TTG GGC GAG CGC GAC TCC AAG CTC
5. CTG GGT TTT AAT CAG CCC AGT GCT GCG GGG agc ggt gtt GCG AGT ATC CAG CAC AAT CAT ATC TCC GTG
6. GGT GCG GGG CGG GGG GGT TTC GAG TTG TTT ggc ggt gat TAT AAC AGC AGT GCT GCC CTT CCG AAT GAG
7. GTC GGT GGC GTT CGG CAG ACC GGT GTG GCT agc ggt gac CAG GCC CTG ACC ATC CCT CGG ATC CCG GGT
8. ATG CAG AGG GGG GTC CAT GGT CTT ACT GTG cgc ggt gac AGT AAC TAC GAT AGT ATC TCT CAT CGT TAC
9. ATT GTG ACG GAG ATG GAT ACC TTG TTG GTC ggc ggt gat GAT CGT AAT TTC CAC TTT GGC CCC TCC ACG
10. AGT GGG GGC GTT TTC GTT GGT GAT GAA GGG ggc ggt gtg GTC CAC GAT AGC TCC AGT CTG CGC CGC GCC
11. GGT AAG GAT TTT TGC GGT CTT GTT CAT GGG ggc ggt gtt TCG TGT AAC TCC AAG GTC CCT AAG ATC TCC
12. GAC GCG ATG ATT TTT AGT AGG TCG GGC GTG ggc ggt gtt TAC AAC TCC CTC GAC GTG AGG ACT CGG AAT
13. GCG ACG GTG CTG GCT TCG GAG TGG TTC GAG ggc ggt ggt CCC AGC TCT AGC ACG GCG CCT CCG TCC TAT
14. GGT TCG GTC TGG TCG TCT CCG TAT CCT AAT cgc ggt gac TTT GGC ATC GAT AAC GCG CTT CAG TAC TCT
15. TGT ACG ACT CAT GGC CAT GAG GTT TTT ATG tgc ggt gac TAT CAT ATG CGG CTG TAC CCG CAT GCC GCG
16. CCC TGG GGC ATG CGG GGC GTT TGT TGG AAG tgc ggt gcc GAC CGC AGT GCG CGT ACC ACG CTC TCC AGC
17. TTG GGG GGG CAG GGG CAC GGC GAT AAG CAT ggc ggt ggt TGG CCC AGG TAC TTT TCG AAT GCC CAG CCT
18. GCC TAT GCC GAC ATC GGG TTT GGT TTT CGT agc ggt gga AAT GAG TCG CAC ATC GCT CCC TCC CGC GGG
19. GGC TCC GTG GGG ACG GAG CAG GGT GGT CTG agc ggt gga GCT CTC GGT CTT GAC ATG CCC GAT CGG AAC
20. CTT GTG GGC AGG TGT ACG GAG ATC CTT GGG tgc ggt gat CTG AGT TTC CCG CAT AGC CAG TCC ATG CAC
21. GTC CAT AGG ACG GTT AAT TCT GAT AGG ACT agc ggt gca GAG GGC AGC AGG GAT ACT CAC AAT TTT AAC
22. AAG GAG AAT CGC CTT GAG TTC CTG TAC GCG agc ggt gat ATT CTG GTC TCT TCA CGA TAT CGC AAT ACT
23. AAG ATG GAC GAT CCG TGG TGC GCC GGG TGG ggc ggt gtt CGC GAT AAG TCT GCC TCC TCC ATG CGC CCC
24. AGT ATT GGG AGC AGG GAT TCT AAC TTC GTG ggc ggt gta CTT GCT AAT CTT GCG TTG CAT AAC TTC CTC
25. GGT CTT GAG GAG GAG GTG TAT TGT GTG GCC ggc ggt gag TTA GCG TCG GCG CCT TCG ATG TAT AAG GGT

Fig. 2. The nucleotide sequences of the inserts of the 25 phage clones that were randomly selected from the phage display library. Only the nucleotide sequences of the $X_{10}\Delta G\delta X_{10}$ region are shown. The sequences of the δGδ region are indicated in small letters.

Table 1. Nucleotide distribution in the inserts of the 25 phage clones randomly selected from the library. The frequency (percentage) of the individual nucleotide was calculated at each nucleotide position of the 500 codons of the X10 regions shown in Fig. 2

Nucleotide	Codon position		
	1st	2nd	3rd
T	20	23	33
C	22	24	30
A	24	27	1
G	34	26	36

Table 2. Codon usage in the 25 random samples. The numbers indicate the frequency of each codon in the samples

First position	Second Position T	C	A	G	Third position
T	15	8	9	5	T
	9	16	9	3	C
	1	2	0	0	A
	5	9	0	11	G
C	23	7	13	4	T
	6	12	7	10	C
	0	0	0	2	A
	10	9	13	7	G
A	6	7	15	14	T
	11	4	10	8	C
	0	0	0	0	A
	12	10	13	10	G
G	7	9	18	18	T
	8	11	8	16	C
	0	0	1	0	A
	13	15	19	23	G

The inserts of the twenty-five clones were sequenced and their sequences are shown in Fig. 2. The frequency of the individual nucleotide was determined at each codon position of the 500 codons of the twenty five insert sequences and shown in Table 1. In general, nucleotides were represented with the expected frequency at the respective positions except that there was a bias toward G at the first position. The bias toward G may be due to the low fidelity of Taq DNA polymerase used for constructing the library; the Taq DNA polymerase prefers to misincorporate G opposite a template T rather than other nucleotides (13, 16). The bias toward G may also be introduced by the biology of the phage (2). Adenine bases were detected at the third codon position despite the fact that the degenerate oligonucleotide inserts were designed not to have A bases at that position. This is also presumably due to the low fidelity of Taq DNA polymerase for DNA synthesis. All forty seven sense codons were represented (Table 2). Since a host strain for a phage display library is a *sup*E mutant, about two UAG stop codons among the 500 codons examined were expected. However, no UAG stop codon was detected. A bigger sample size than the 500 codons would show UAG stop codons.

Amino acid sequence diversity of the library clones

The amino acid sequences deduced from the nucleotide sequences (Fig. 2) of the 25 phage clones are shown in Fig. 3. The tripeptides (ΦGδ) were represented in all clones. While there were biases toward glycine and trytophan residues for the Φ residues and glutamate residues for the δ residues, most of the expected residues were represented with the expected frequency. Fig. 4 shows the frequency of the individual amino acid residues of the 25 peptide inserts, in which the frequency of the tripeptides were omitted. In most cases, the observed frequency for each amino acid residue was well correlated with the expected frequency. However, cysteine residues were underrepresented. This could be partly because the cysteine residues may form disulfide bonds with other cysteine residues, causing the malfunction of the pIII protein, resulting in the low infectivity of the phage containing the cysteine residues. Indeed, the N-terminal region of the mature pIII protein contains a few cysteine residues which could form disulfide bonds with the cysteine residues of the inserts (9). Meanwhile, glycine and glutamate residues were overrepresented. This bias may be due to the biology of the phage (2).

In order to examine if individual amino acid residues were distributed in random along the inserts, the range of diversity in dipeptide sequence was examined. If distributed in random, the population of amino acid residues in the random sample should represent more than half of the 400 possible dipeptide pairs ($20 \times 20 = 400$) (7). As shown in Table 3, 256 different dipeptide pairs were identified, indicating that individual amino acid residues were distributed in random along the inserts of the phage library. It has been known that a peptide mimotope with apparently unrelated primary amino acid sequences could mimic a discontinuous antigenic determinant (7). This shows that the specificity and strength of antigen-antibody recognition is mediated not only by the linear amino acid sequence, but alsoby the amino acid size, charge, structure, and hydrophobicity. Therefore, we have examined the frequency and positional distribution of amino acid types in the inserts (Fig. 5) as well as the distribution of isoelectric points and hy-

1. Asp-Trp-Glu-Asn-Gly-Gly-Arg-Gly-Ser-Gly-C-G-V-Lys-Gly-Pro-His-Asp-Tyr-Pro-Ile-Cys-Pro
2. Gly-Ala-Phe-Gln-Gly-Phe-Asp-Pro-Thr-Ser-G-G-A-Leu-Ala-Gly-Phe-Leu-Pro-Ser-Thr-Ala-Ser
3. Arg-Glu-Gly-Ile-Arg-Asn-Leu-Asp-Trp-Met-R-G-V-Pro-Gln-Tyr-Gly-Ser-Tyr-Ala-Lys-Pro-Asp
4. Asn-Gly-Trp-Phe-Gln-Lys-Ser-Glu-Val-Trp-G-G-G-Ala-Ile-Phe-Gly-Glu-Arg-Asp-Ser-Lys-Leu
5. Leu-Gly-Phe-Asn Gln-Pro-Ser-Ala-Ala-Gly-S-G-V-Ala-Ser-Ile-Gln-His-Asn-His-Ile-Ser-Val
6. Gly-Ala-Gly-Arg-Gly-Gly-Phe-Glu-Leu-Phe-G-G-D-Tyr-Asn-Ser-Ser-Ala-Ala-Leu-Pro-Asn-Glu
7. Val-Gly-Gly-Val-Arg-Gln-Thr-Gly-Val-Ala-S-G-D-Gln-Ala-Leu-Thr-Ile-Pro-Arg-Ile-Pro-Gly
8. Met-Gln-Arg-Gly-Val-His-Gly-Leu-Thr-Val-R-G-D-Ser-Asn-Tyr-Asp-Ser-Ile-Ser-His-Arg-Tyr
9. Ile-Val-Thr-Glu-Met-Asp-Thr-Leu-Leu-Val-G-G-D-Asp-Arg-Asn-Phe-His-Phe-Gly-Pro-Phe-Thr
10. Ser-Gly-Gly-Val-Phe-Val-Gly-Asp-Glu-Gly-G-G-V-Val-His-Asp-Ser-Ser-Ser-Leu-Arg-Arg-Ala
11. Gly-Lys-Asp-Phe-Cys-Gly-Leu-Val-His-Gly-G-G-V-Ser-Cys-Asn-Ser-Lys-Val-Pro-Lys-Ile-Ser
12. Asp-Ala-Met-Ile-Phe-Ser-Arg-Ser-Gly-Val-G-G-V-Tyr-Asn-Ser-Leu-Asp-Val-Arg-Thr-Arg-Asn
13. Ala-Thr-Val-Leu-Ala-Ser-Glu-Trp-Phe-Glu-G-G-G-Pro-Ser-Ser-Ser-Thr-Ala-Pro-Pro-Ser-Tyr
14. Gly-Ser-Val-Trp-Ser-Ser-Pro-Tyr-Pro-Asn-R-G-D-Phe-Gly-Ile-Asp-Asn-Ala-Leu-Gln-Tyr-Ser
15. Cys-Thr-Thr-His-Gly-His-Glu-Val-Phe-Met-C-G-D-Tyr-His-Met-Arg-Leu-Tyr-Pro-His-Ala-Ala
16. Pro-Trp-Gly-Met-Arg-Gly-Val-Cys-Trp-Lys-C-G-A-Asp-Arg-Ser-Ala-Arg-Thr-Thr-Leu-Ser-Ser
17. Leu-Gly-Gly-Gln-Gly-His-Gly-Asp-Lys-His-G-G-G-Trp-Pro-Arg-Tyr-Phe-Ser-Asn-ala-Gln-Pro
18. Ala-Tyr-Ala-Asp-Ile-Gly-Phe-Gly-Phe-Arg-S-G-G-Asn-Glu-Ser-His-Ile-Ala-Pro-Ser-Arg-Gly
19. Gly-Ser-Val-Gly-Thr-Glu-Gln-Gly-Gly-Leu-S-G-G-Ala-Leu-Gly-Leu-Asp-Met-Pro-Asp-Arg-Asn
20. Leu-Val-Gly-Arg-Cys-Thr-Glu-Ile-Leu-Gly-C-G-D-Leu-Ser-Phe-pro-His-Ser-Gln-Ser-Met-His
21. Val-His-Arg-Thr-Val-Asn-Ser-Asp-Arg-Thr-S-G-A-Glu-Gly-Ser-Arg-Asp-Thr-His-Asn-Phe-Asn
22. Lys-Glu-Asn-Arg-Leu-Glu-Phe-Leu-Tyr-Ala-S-G-D-Ile-Leu-Val-Ser-Ser-Arg-Tyr-Arg-Asn-Thr
23. Lys-Met-Asp-Asp-Pro-Trp-Cys-Ala-Gly-Trp-G-G-V-Arg-Asp-Lys-Ser-Ala-Ser-Ser-Met-Arg-Pro
24. Ser-Ile-Gly-Ser-Arg-Asp-Ser-Asn-Phe-Val-G-G-V-Leu-Ala-Asn-Leu-Ala-Leu-His-Asn-Phe-Leu
25. Gly-Leu-Glu-Glu-Glu-Val-Tyr-Cys-Val-Ala-G-G-E-Leu-Ala-Ser-Ala-Pro-Ser-Met-Tyr-Lys-Gly

Fig. 3. The deduced amino acid sequences from the nucleotide sequences shown in Fig. 2. The amino acid sequences of the ΦGδ region are indicated in single letter symbols.

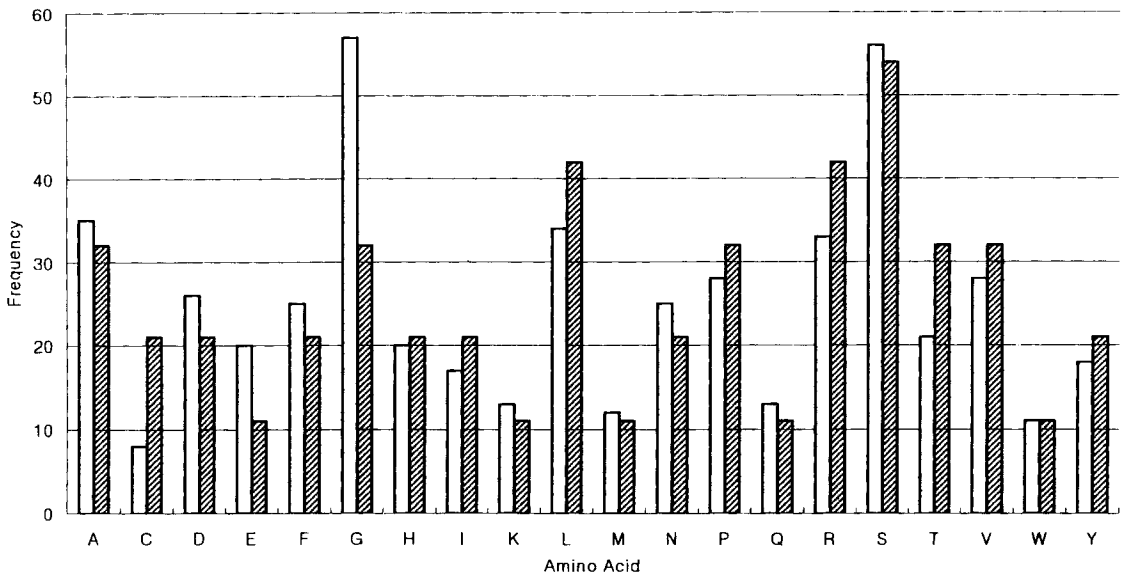


Fig. 4. Frequency of individual amino acid species in the 25 samples. The frequency of each amino acid residue in the sample was calculated by multiplying the fraction (n/48, n=no. of codons) of codons which encode each amino acid residue by the total number (500) of amino acid residue present in the samples. The expected or observed frequency of each amino acid residue in the samples is indicated by the open bar or hatched bar, respectively. The amino acid residues are indicated in single letter symbols.

Table 3. The presence of independent dipeptide within the samples. The amino acid sequences were divided into individual overlapping dipeptides. The single amino acid at each end was not paired. Each dipeptide was indicated by frequency number at the intersection of the labeled rows (first amino acid) and columns (second amino acid)

	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
Ala	3		1		1	4		1	1	5	1	1	3	1	1	5	1			1
Cys	1					1						1	1				2	1	1	
Asp	1		1	1	1			1	2		1	1	2		4	4	2	1	2	1
Glu				1	1	3		1			1	2		1	1	1		3	1	
Phe		1	1	2		4	1			3	1	2	1	2	1	2	1	2		
Gly	2		2	1	6	6	2	2	1	5	1		2	1	3	6	1	6	2	
His	1		2	1	1	4		2			1	3			2	1				
Ile	1	1	1		2	2				2			2	1	1	3		1		
Lys			1	1		2	1	1		1	1		1			2		1		
Leu	5		3	2	1	4	1			1			2	1	1	2	2	4		2
Met			2				1	1					1	1	3					1
Asn	2			2	4	2	1			2				1	1	4				1
Pro			2		1	1	3	1	1			2	1	1	2	6	1		2	1
Gln	1					3	1		1				2		1	1	1			2
Arg	1	1	4	1		5		1		2		5	1	1	1	2	4			3
Ser	5	1	1	2		3	2	3	2	2	3	3	1	1	5	7	2	3		2
Thr	2			3		1	2	1		2					1	1	2	3		
Val	2	1			2	4	4			1		1	1		2	1	1		2	1
Trp		1		1	2	1			1		1		1			1				
Tyr	3	1	1		1	1	1		1			2	3		1	1				

dropathic indices for the inserts (Fig. 6) (9). As shown in Fig. 5, acidic, basic, polar, and nonpolar amino acids were distributed mostly in proportion to their presence in the genetic code except that there seemed to be some bias against basic or acidic amino acids at the 8th, 19th, and 22th amino acid positions in the inserts. The data in Fig. 6 show that there is some bias for hydropathic indices, but no significant bias for isoelectric points for the 25 insert peptides.

The data in Fig. 6 show that the combined distribution of hydropathic indices and isoelectric points for the 25 insert peptides covers a relatively broad field. In sum, the 25 insert peptides of the phage display library examined in this study shows a broad spectrum of biochemical properties.

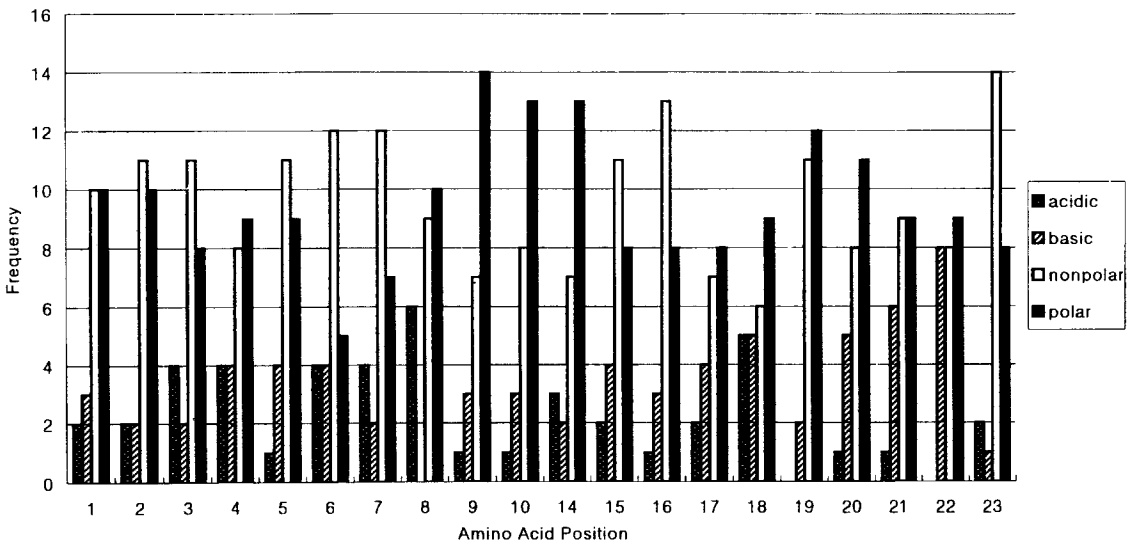


Fig. 5. Frequency and positional distribution of amino acid types in the 25 samples. The amino acid residues were divided into acidic (Asp, Glu), basic (Lys, Arg, His), nonpolar (Ala, Val, Leu, Ile, Pro, Phe, Trp, Met), and polar (Gly, Ser, Thr, Cys, Tyr, Asn, Gln) types. The total number of each amino acid type at each residue position in the samples is indicated by the specific bars. The amino acid position is numbered as follows : $X_1 \cdots X_{10}\Phi G\delta X_{14} \cdots X_{23}$.

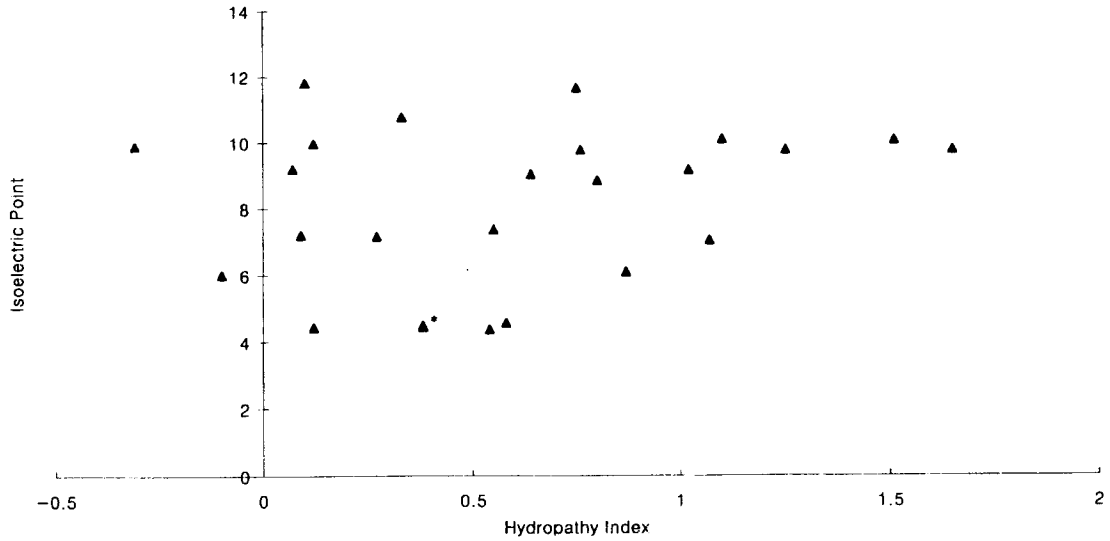


Fig. 6. Profiles of the combined isoelectric and hydropathic diversity in the 25 samples. The isoelectric point and hydropathic index(8) for each sample were determined using the DNA star program. The asterisk indicates two samples which have the similar combined scores.

Therefore, this library will be useful for mapping antibody epitopes and finding ligands for other binding proteins such as receptors and enzymes.

Acknowledgments

The authors are greatful to Jeong-Hyun Kim for his assistance in the preparation of the manuscript. We also thank Dr. Yong-Suk Nam for providing us with *E. coli* MV1190. This work was supported by Grant BSRI 95-4407 from the Ministry of Education, Korea.

References

1. **Bass, S., R. Greene, and J.A. Wells.** 1990. Hormone phage: an enrichment method for variant proteins with altered binding properties. *Proteins* **8**, 309-314.
2. **Cwirla, S.E., E.A. Peters, R.W. Barrett, and W.J. Dowers.** 1990. Peptides on phage; A vast library of peptides for identifying ligands. *Proc. Natl. Acad. Sci. USA* **87**, 6378-6382.
3. **DeGraaf, M.E., R.M. Miceli, J.E. Mott, and H.D. Fischer.** 1993. Biochemical diversity in a phage display library of random decapeptides. *Gene* **128**, 13-17.
4. **Devlin, J.J., L.C. Panganiban, and P.E. Devlin.** 1990. Random peptide libraries: A source of specific protein binding molecules. *Science* **249**, 404-406.
5. **Fowlkes, D.M., M.D. Adams, V.A. Fowler, and B.K. Kay.** 1992. Multipurpose vectors for peptide expression on the M13 viral surface. *BioTechniques* **13**, 422-427.
6. **Geysen, H.M., S.J. Rodda, and T.J. Mason.** 1986. A priori delineation of a peptide which mimics a discontinuous antigenic determinant. *Mol. Immunol.* **23**, 709-715.
7. **Kay, B.K., N.B. Adey, Y-S He, J.P. Manfredi, A.H. Mataragnon, and D.M. Fowlkes.** 1993. An M13 phage library displaying random 38-amino-acid peptides as a source of novel sequences with affinity to selected targets. *Gene* **128**, 59-65.
8. **Kyte, J. and R.F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105-132.
9. **Marvin, D.A. and B. Hohn.** 1969. Filamentous bacterial Vir uses. *Bacteriological Reviews* **33**, 172-209.
10. **McCafferty, J., A.D. Griffiths, G. Winter, and D.J. Chiswell.** 1990. Phage antibody: filamentous phage displaying antibody variable domains. *Nature* **348**, 552-554.
11. **McCafferty, J., R.H. Jackson, and D.J. Chiswell.** 1991. Phage-enzymes: expression and affinity chromatography of functional alkaline phosphatase on the surface of bacteriophage. *Protein Eng.* **4**, 955-961.
12. **Oldenburg, K.R., D. Loganathan, I.J. Goldstein, P. G. Schultz, and M.A. Gallop.** 1992. Peptide ligands for a sugar-binding protein isolated from a random peptide library. *Proc. Natl. Acad. Sci. USA* **89**, 5393-5397.
13. **Saiki, R.K., D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis, and H.A. Erlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-491.
14. **Sanger, F., S. Nicklen, and A.R. Coulson.** 1977. DNA sequencing with chain inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
15. **Scott, J.K. and G.P. Smith.** 1990. Searching for peptide ligands with an epitope library. *Science* **249**, 386-390.
16. **Tindall, K.R. and T.A. Kunkel.** 1988. Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Biochemistry* **27**, 6008-6013.
17. **Vierra, J. and J. Messing.** 1987. Production of a single-stranded plasmid DNA. *Methods Enzymol.* **153**, 3-11.