Cosmix-Plexing: A Novel Way to Generate and Use Combinatorial Phage-Display Libraries

Nathalie Horn,[1] ✉ Nelson S Vispo,[2] Michael Szardenings,[3] John Collins[1,3]

[1]Dept. Applied Genetics, GBF, Mascheroder Weg 1, Braunschweig, D-38124 Germany.
[2]Centro de Ingienería Genética y Biotecnología. Ave. 31 e/ 158 y 190, Playa. AP 6162, CP 10600, Ciudad de La Habana, Cuba. E-mail: nelson.santiago@cigb.edu.cu
[3]Cosmix Molecular Biologicals. GmbH, Mascheroder Weg 1b, Braunschweig, D-38124, Germany.

ABSTRACT

This work describes the Cosmix-Plexing® procedure. It provides a novel way to use and generate combinatorial libraries of peptides displayed on the surface of filamentous phages. Using the Cosmix-Plexing approach, the left and right sections of the variable domain can be induced by the use of type II restriction enzymes. Thus, a population with "optimized" sequences for a larger number of amino acid residues can be generated. It can also simplify the production of extension libraries by the introduction of specific cassettes into existing libraries. This technology enabled us to efficiently isolate ligands having a great affinity for several target molecules of the EVH1 of VesI domain from an initial library that was rich in proline.

Keywords: bacteriophages, combinatorial libraries, Cosmix-Plexing, phage-display, recombination

Biotecnología Aplicada 2001;18:221-231

RESUMEN

Cosmix-Plexing: modo novedoso de generar y utilizar bibliotecas combinatorias de péptidos presentados en la superficie de fagos. En este trabajo se describe la técnica denominada Cosmix-Plexing® que proporciona un modo novedoso de generar y utilizar bibliotecas combinatorias de péptidos presentados en la superficie de fagos. Mediante la técnica de Cosmix-Plexing se puede inducir la recombinación de las secciones izquierda y derecha del dominio variable del anticuerpo mediante el uso de enzimas de restricción de tipo II y se puede generar así una población con secuencias "optimizadas" para un mayor número de residuos aminoacídicos. También puede facilitar la producción de bibliotecas grandes mediante la introducción de casetes específicos en bibliotecas ya generadas. Con la aplicación de esta técnica se logró aislar de forma eficiente ligandos de gran afinidad para varias moléculas blanco del dominio EVH1 de Ves1 a partir de una biblioteca inicial rica en prolina.

Palabras claves: bacteriófagos, bibliotecas combinatorias, Cosmix-Plexing, exposición en fagos, recombinación

*I*ntroduction

Phage-display was initially described in 1985 by George P. Smith [1] as a way to present on the surface of the virus a genetic product, "*which is always coupled with the gene, which encodes that particular variant protein, the gene for this variant being packaged within the same virus particle".* Since then, an increasing number of laboratories are working with this technique, with a wide range of applications [2]. For the first time, in 1991, large libraries of proteins and peptides displayed on phages, were successfully employed to isolate clones in which the selected variant had defined binding properties. Kay *et al.* [3] reviewed recent advances in the search for peptide ligands, and Holliger and Hoogenboom [4] described the use of antibodies derived by this method, and in particular the successful application of these products for clinical use.

One factor limiting the success of the phage-display technology is the size of the library. To maximize interactions between the selected ligand and the target to which it binds, it would be desirable to display a mixture of relatively long peptides. However, the minimal number of clones required within a library (containing all the possible sequences) increases exponentially with the number of amino acid residues in the peptides. A phage library that displays a 9-amino acid (aa) peptide on its surface would require a population of at least 1.5 x 10^{12} variants (2.4 x 10^{17} for a 13-aa peptide). However, the largest combinatorial peptide libraries which have been constructed contain about 10^{11} variants, thus, only a fraction of all possible variants are represented in the library.

The development of a novel technology, the Cosmix-Plexing® approach, overcomes this limitation [5]. The potential diversity of the library will no longer be represented by the number of variants present in the initial library. Instead, those used in this technique contains a single recombination site within the hypervariable domain. Generally, recombination will be carried out with a preselected population of (for example) 10^4 variants, thus generating 10^8 possible recombinants, essentially containing all the possible combinations of sequences on either side of the recombination site [6]. The advantage of this technique due to the increase in the number or residues obtained through the diversity of a preselected population is shown in Figure 1.

Using the Cosmix-Plexing approach, the reassortment of the left and right sections of the variable domain can be induced by the use of type II restriction enzymes. The characteristic of these enzymes is that they bind at a defined sequence, but cleave at a site that is at a fixed distance away, i.e. in a region with no demands on sequence specificity. Thus, a population with "optimized" sequences for a larger number of amino acid residues can be generated.

✉ Corresponding author

1. Smith GP. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science 1985;228:1315–7.

2. Collins J. Phage display. In: Moos WH, *et al.*, editors. Ann Reports in Combinatorial Chemistry & Molecular Diversity. Vol. 1, ESCOM. Leiden; 1997. p.210–50.

3. Kay BK, Kurakin AV, Hyde-DeRuyscher R. From peptides to drugs via phage display. DDT 1998;3:370–8.

4. Holliger P, Hoogenboom H. Antibodies come back from the brink. Nat Biotechnol 1998;16:1015–6.

5. Collins J, Röttgen P. Evolutive phage-display –a method as basis for a company. GBF. Scientific Annual Report 1997;19–28.

6. Collins J, Röttgen P, inventors. Cosmix-Plexing patent EP 9800533 (pending) based on EP 97101539.1, (priority 6.02.1997).

The generation of recombination within a particular region is a powerful concept. DNA shuffling for example as applied by the Arnold and Stemmer group [7–9], clearly demonstrates the advantages of this idea. However, this method cannot be used in all cases. It can work only with gene families or gene variants, having a high homology among them. The Cosmix-Plexing technique is not limited by homology between recombination partners, and showing a higher efficiency for recombination within short regions.

Although the main advantage of Cosmix-Plexing-induced recombination is obtained through the increase in diversity during selection, it can also be used for the simple production of libraries with a large number of clones (e.g. $>10^9$ variants). In addition, it enables the production of extended libraries, by the introduction of specific cassettes into existing libraries (Figure 2). These cassettes are of particular interest if a "core motif" is known (for example, Pro Pro Xxx Pro motif of the domains 3 of homology with Src (SH3) or Ena-VASP homology domain 1 (EVH1), which show a preference for proline rich ligands).

The cassette, containing this Pro Pro Xxx Pro motif was used in our project to isolate and characterize strong ligands to the SH3 and EVH1 domains and it will be given as an example to illustrate the Cosmix-Plexing technology. We will first describe how to build a library featuring the Cosmix-Plexing properties, and secondly, focus on its recombination during selection.

Materials and Methods

Preparation of phagemid peptide libraries

Preparation of the linear vector DNA. The phagemid pROCOS4/7stuffer1 that we used in this work is an expression vector where the variant peptides can be fused to the N-terminal part of the pIII gene. Phagemid vectors contain both plasmid and bacteriophage replication origins. However, except for the gene to be fused with the peptide or protein codifying sequence to be displayed, these vectors do not carry any of the phage genes. The phagemids remain as plasmids, until a superinfection with a helper phage (M13KO7) activates their phage replication/packaging origins. The helper phage provides all bacteriophage functions necessary for the synthesis of the circular single-stranded DNA, and for the assembly of the phage-like particles, displaying the hybrid fusion proteins.

In the pROCOS4/7stuffer1 system, the fusion-pIII hybrid synthesis is under the control of the λ P_L promoter. The λ P_L promoter is repressed by the λ cI protein, which is produced in *Escherichia coli* strains containing an integrated copy of the phage λ genome (i.e. lysogens). It was found empirically, using this promoter, that the correct ratio of hybrid to normal pIII in the particles produced after superinfection was achieved for essentially monovalent presentations (i.e., one hybrid pIII molecule per viral particle). pROCOS4/7stuffer1 contains an ampicillin resistance gene, allowing the selection for phagemid-containing cells.

The stuffer sequence is a 950 bp DNA fragment of vector pBR322. We performed a complete digestion with two different restriction enzymes, in such a way that two different coheside ends were generated in the larger fragment corresponding to the vector. This fragment cannot undergo ring closure in the absence of an inserted cassette fragment. The site, into which the hypervariable oligonucleotide cassette is to be inserted, was prepared by digestion with the *Kpn*I and *Sac*I restriction enzymes. The hypervariable region was then inserted between a "signal" sequence and the

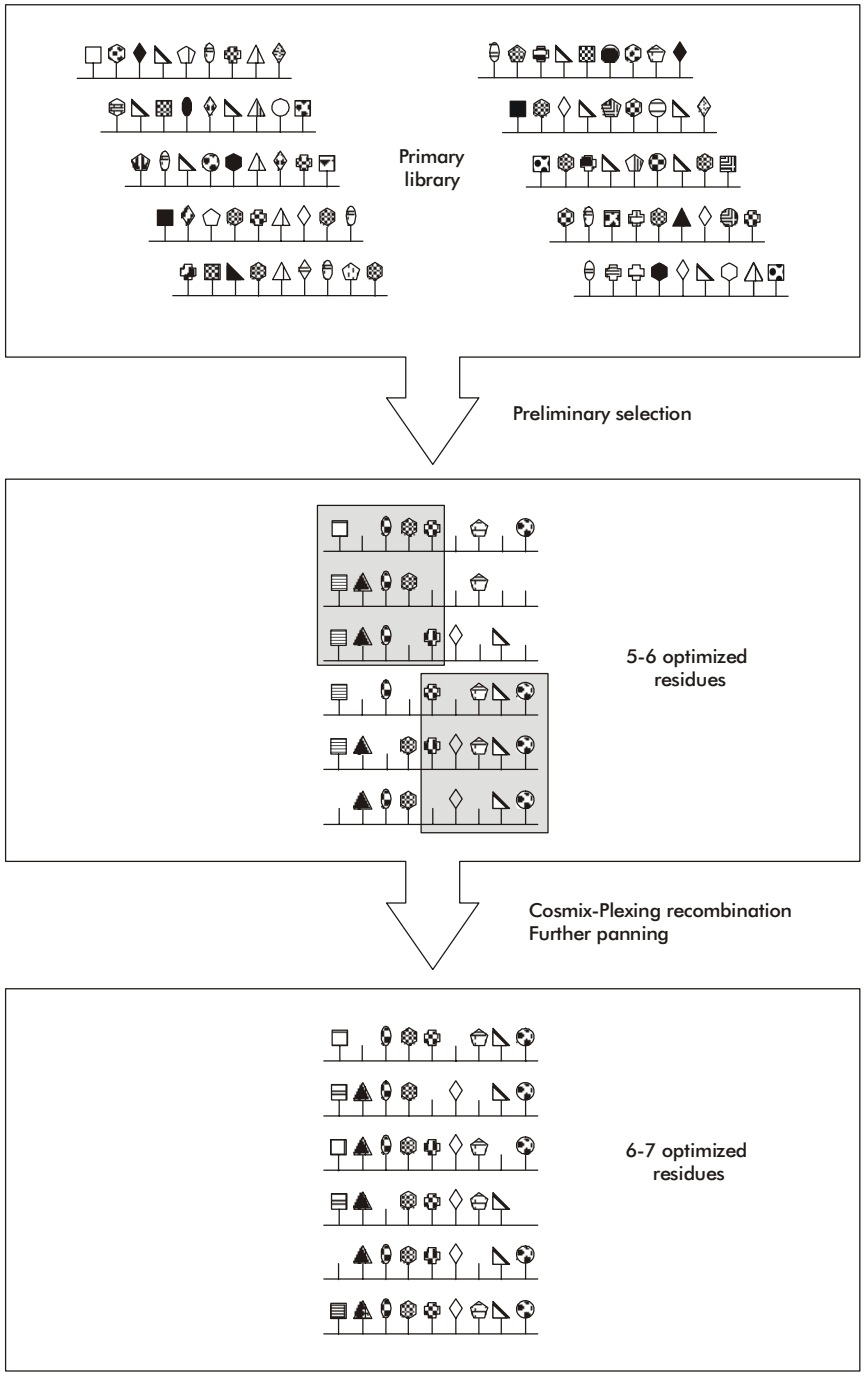


Figure 1. Chart of the principle of Cosmix-Plexing-induced recombination during screening. A medium-sized combinatorial library (upper section) is subjected to an affinity selection procedure, leading to the isolation of several thousand variants at least partially fulfilling the selected requirement, in which on average some 6-aa residues may be "optimal" (middle section). The left and right section of the hypervariable domain is recombined so that possible recombinants are produced. Further selection now leads to the isolation of "optimal" variants containing eight or nine optimized residues (lower section).

7. Stemmer WPC. DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. Proc Natl Acad Sci USA 1994;91: 10747–51.

M13*gpIII* gene in our phage-display vector so that the peptide is presented on the surface of the viral particle as an amino terminal extension. The resulting library displays an 8-aa peptide variant on the phage surface.

The second step was to introduce the project specific cassette within the 8-aa peptide insert already present in the vector. The following variant peptides are then displayed:

$(Xxx)_5$ Pro Pro Xxx Pro $(Xxx)_4$

These two steps in the construction of the final library are schematically illustrated in Figures 3 and 4, respectively.

Cosmix-Plexing recombination can be carried out on this extended library by cleavage at the type II restriction site *(Bpm*I), allowing the optimization of the whole region (see below). The vector is linearized according to standard methods.

Procedure

1. Cleave the vector at one of the restriction sites *(Kpn*I) 3 h at 37 ºC.
2. Verify the digestion on a 1% agarose gel. Purify the vector according to the GFX™ protocol (GFX purification kit; Amersham Pharmacia Biotech, Sweden).
3. Incubate the vector with the second enzyme *(Sac*I) at 37 ºC for 3 h. Verify the digestion on another 1% agarose gel. The two bands of the vector and the stuffer fragment appear clear and distinct.
4. Extract the large linear fragment of the gel with the GFX purification kit.
5. Determine the concentration of the DNA from its OD_{260}.

The vector is then ready for the ligation with the insert.

Preparation of the double stranded DNA insert from degenerated oligonucleotides. The codon scheme used to generate mixtures of amino acid sequences is NNB, where N means A, C, G or T, and B represents C, G or T. This scheme allows the reduction of the stop codon frequency—that would lead to unproductive clones—as only one of the three stop codons (TAG) can be encoded, while codons for all 20 aa are still represented.

Four sets of inserts are prepared, which differ only in the dinucleotide corresponding to the cohesive ends that will be generated on cleavage by the type II *Bpm*I restriction enzyme (at the ZZ position, in bold type in the DNA sequence shown in Figure 5). The *Bpm*I type II restriction enzyme was bound to a defined sequence within the invariable sequence adjacent to the inserts, but it cleaves at a fixed distance from the binding site, within the middle of the hypervariable domain, creating a 2-bp cohesive end. ZZ stands for AM (AA and AC), CT, GG, and TS (TC and TG). It is important that no homologous end bindings occur on religating the cleaved sequence. This was achieved by the set of base pairs used at the ZZ position. Furthermore, when the four inserts are pooled, this NZZ allows the representation of all amino acids, except cysteine. The *Bpm*I restriction site was necessary for the insertion of the cassette, and for the Cosmix-Plexing recombination, whereas the *Sac*I and *Kpn*I restriction sites were used to create the original insert.

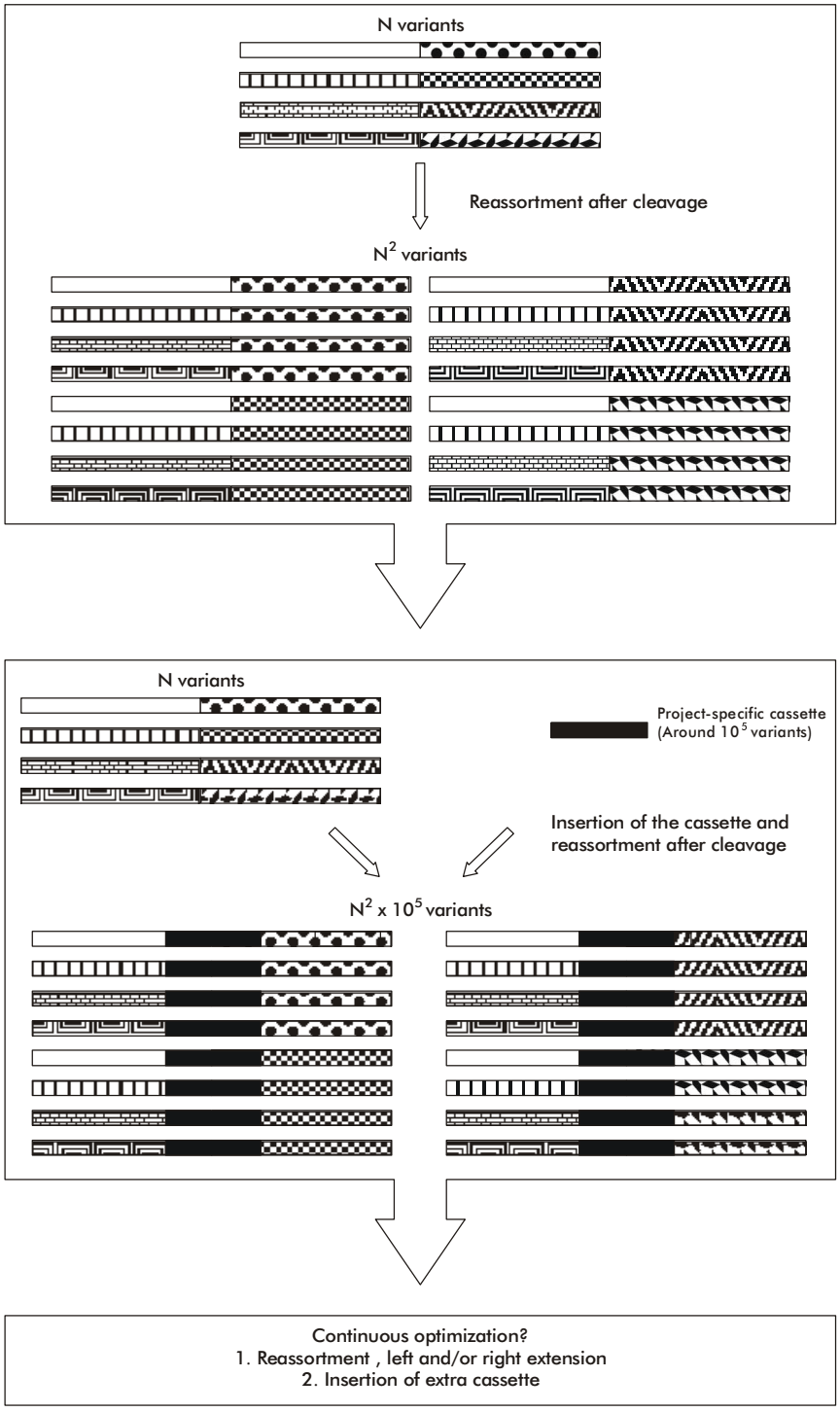


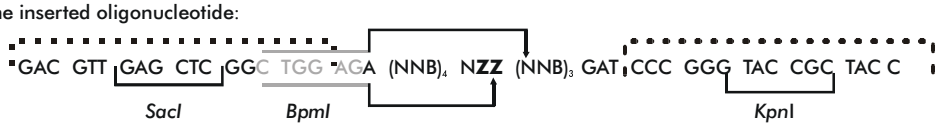
Figure 2. Scheme of the application of Cosmix-Plexing. The upper section shows the selection of an optimal primary ligand (principle outlined in Figure 1). The consensus motif can be inserted as an evolutionary strategy to take advantage of it.

Although the strategy chosen to prepare the oligonucleotides contains many steps, it did, however, guarantee the purity of the products (Figure 3). It should be noted that the quality of the digested inserts had a strong influence on the quality of the ligation with the vector, and thus, on the efficiency

8. Stemmer WPC. Rapid evolution of a protein *in vitro* by DNA shuffling. Nature 1994;370:389–91.

9. Patten PA, Howard RJ, Stemmer WPC. Applications of DNA shuffling to pharmaceuticals and vaccines. Curr Opin Biotechnol 1997; 8:724–33.

Sequence of the inserted oligonucleotide:



Sequence of the inserted peptide:

Asp Val Glu Leu Gly Trp Arg **Xxx**$_4$ **Yxx** **Xxx**$_3$ Asp Pro Gly Tyr Arg Tyr

Primer 1 DNA sequence:

5' biotin GAC GTT GAG CTC GGC TGG 3'

Primer 2 DNA sequence:

5' biotin GGT AGC GGT ACC CGG GAT 3'

Figure 3. DNA and peptide sequences of the inserts. On the DNA sequence, recognition sites of the enzymes are indicated: *Sac*I and *Kpn*I are responsible for cloning the inserts within the vectors. *Bpm*I allows the insertion of the cassettes within the hypervariable region. The cleavage site is described. DNA sequences of the biotinylated primers are given, as well as their complementary sequences with the inserts (▪▪▪▪▪▪, •••••• indicate the complementary sequences for the primers 1 or 2, respectively). ZZ represents AM (AA, AC), CT, GG, TS (TC, TG), and Yxx any amino acid other than cysteine.

of the introduction of the ligated DNA into the bacteria by electroporation.

The oligonucleotides that are to be inserted are converted into double-stranded DNA, and amplified via a polymerase chain reaction (PCR), using biotinylated primers (in order to easily remove the restriction fragments later on).

Procedure

1. Prepare the PCR reaction mixture in the following way:

Insert	1 µL (10 pmol)
Primer 1	2 µL (200 pmol)
Primer 2	2 µL (200 pmol)
PCR buffer (10x)	10 µL
$MgCl_2$ (25 mM)	10 µL
dNTP (2 mM each)	10 µL
Taq DNA polymerase	0.4 µL (2U)
H_2O	up to 100 µL

   Use a heated lid PCR thermocycler with the following program, for 10 cycles:

94 ºC	1 min
60 ºC	1 min
72 ºC	1 min

2. Verify the quality, size and purity of the double-stranded DNA on a 4.5% gel
3. Purify the double-stranded DNA by phenol extraction, chloroform precipitation (the inserts are too short to be purified on the column of the GFX purification kit). After each cleavage (or ligation) step, check the quality of the DNA on a 4.5% agarose gel, before being purified according to the phenol/chloroform extraction method.
4. Digest the inserts with the *Kpn*I restriction enzyme.
5. Purify the inserts according to the phenol/chloroform extraction method.
6. Digest the inserts with the *Sac*I restriction enzyme.
7. Capture and remove the biotinylated DNA with superparamagnetic polysterene beads with streptavidin covalently attached to the bead surface (Dynabeads® M-280 streptavidin from Dynal). A magnet particle concentrator (Dynal MPC®) is used to immobilize the beads.
8. Purify the inserts according to the phenol/chloroform extraction method.
9. Ligate the inserts among themselves in order to build a long chain of monomers (concatemer). Inactivate the ligase with heat.
10. Purify the inserts according to the phenol/chloroform extraction method.
11. Cleave the DNA fragment with *Sac*I, to obtain a dimeric insert with a *Kpn*I site in the center, and *Sac*I at the cohesive ends.
12. Purify the inserts according to the phenol/chloroform extraction method.

The insert is then ready to be ligated to the vector prepared according to the step 1 (Figure 3).

Ligation of the vector and insert DNA. To obtain a large library it is important to determine the optimal ratio of the insert and vector. This is carried out performing a series of test ligations, with various ratios of insert/vector (e.g. molar ratios of 2:1, 1:1, 1:2...). The ratio, yielding the largest number of recombinants, is used for preparative scale ligation (e.g. 10 µg).

The DNA quality is verified between the ligation and restriction steps, on a 1% agarose gel, and then purified using the GFX purification kit.

Procedure

1. Ligate the vector digested with *Sac*I and *Kpn*I and inserts digested with *Sac*I. The vector can not, thereby, religate because i) the ends are different and ii) the inserts can only be ligated with one end of the vector; the one digested with *Sac*I. Inactivate the ligase with heat
2. Digest the DNA with *Kpn*I.
3. Ligate the vector attached to one end of the insert to close the DNA molecule. This ligation should either be performed at a low DNA concentration to encourage religation and discourage reinsertion of the small cleavage products, or, optionally, the small fragments may also be removed by gel purification.

The hybrid thus formed (vector with insert) is, finally, introduced by electroporation into a WK6λmutS *E. coli* strain (ampicillin selection; 300 µg/mL). Clones are randomly selected and their DNA is sequenced to

verify the presence of the insert, and the integrity of the bank (very few clones should have changes in their reading frame, the frequency of the stop codons should be very low, and finally, the frequency of codons for each amino acid is determined, and compared to the expected frequency for a random NNB library). After the confirmation that the bank is of good quality, the packaging of the genetic material is induced by superinfection of these cells with the M13KO7 helper phage.

Introduction of ligated DNA into bacteria by electroporation. Electroporation produces high efficiency transformation by subjecting a cell/DNA mixture to a brief but intense electrical field of exponential decay. Because of this strong electrical field, the mixture must be free of salts to avoid the generation of an excessively high current during electroporation. A culture of *E. coli* is made electrocompetent by a series of washes. We used the WK6λmutS strain that allowed an efficiency of up to 10^{10} transformants/μg with control native pBR322 plasmid (supercoiled DNA control). The ligated vector gives yields of $>10^7$ (up to 10^8) transformants/μg of DNA.

Procedure

Preparation of competent cells [10]

1. Inoculate 2 L of Luria Bertani (LB) medium/tetracycline (20 μg/mL), (4 recipients x 500 mL) with 20 mL of a fresh overnight culture of WK6λmutS, and incubate it at 37 ºC with shaking until an optical density at 600 nm (OD_{600}) of 0.6 is reached.
2. Chill the culture on ice for 10 min. The rest of the procedure has to be performed at 4 ºC, using prechilled pipettes, rotors and tubes. Transfer aliquots of the culture into centrifuge tubes, and pellet the cells by centrifugation for 15 min at 4000 x*g* (Sorvall GS3 rotor). Decant the supernatant carefully, but remove as much of the liquid as possible.
3. Gently resuspend the cells in 2-L ice-cold Hepes 1 mM pH 7. Repeat the centrifugation step. Eliminate the supernatant.
4. Gently resuspend the cells in 1 L of cold sterile distilled water. Repeat the centrifugation step. Eliminate the supernatant.
5. Gently resuspend the cells in 40 mL of cold sterile glycerol (10%). Pellet the cells by centrifugation for 15 min at 4000 x*g* (Sorvall SS-34 rotor). Eliminate the supernatant.
6. Gently resuspend the cells in 4 mL of cold glycerol (10%). The final cell concentration should be of 2–4 x 10^{10}/mL.
7. Freeze 80-μL aliquots (in precooled sterile Eppendorf tubes) in liquid nitrogen, and store at -70 ºC.

Electroporation using the Bio-Rad Gene Pulser (USA).

8. For each electroporation, use a single 80-μL aliquot, and let it thaw on ice. Add 0,5 μg of DNA (in less than 10 μL of double distilled water (ddH_2O) and incubate on ice for 1 min.
9. Transfer the suspension to a cold electroporation cuvette (0,2 cm path length), place the cuvette in the electroporation sled, and give a pulse at a voltage of 2.5 kV, a capacity of 25 μF and a resistance of 200 Ω. This combination yields a pulse of 12.5 $kV.cm^{-1}$, and a pulse length of 4.5 to 4.7 msec [6].

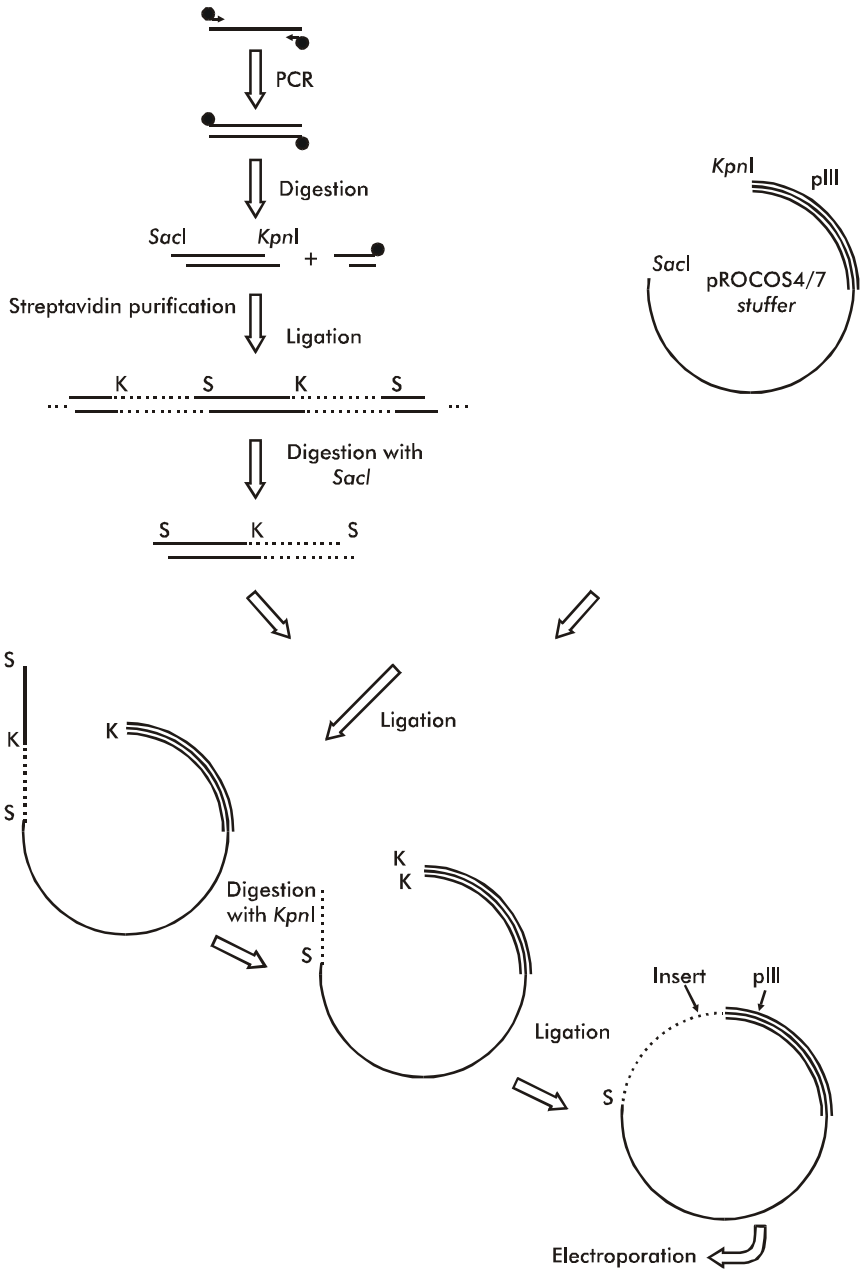


Figure 4. Schematic primary library construction. The insert oligonucleotides are amplified by PCR, using biotinylated primers, leading to a product carrying biotin at its ends. The PCR product is cleaved with *Sac*I and *Kpn*I, and the restriction fragments are removed by magnetic streptavidin-coated beads. The inserts are ligated together, producing concatemers, which are resolved as dimers by restriction with *Sac*I (upper left section). The vector-cloning site is opened by *Sac*I and *Kpn*I (upper right section). Vectors and dimers are ligated together. A restriction with *Kpn*I makes available the *Kpn*I site of the insert, which is now free to ligate with the second end of the vector. The vector with the fusion to the pIII protein can now be electroporated to give a high diversity library (lower section). ● Represents the biotin, linked to the primers, and the PCR products.

10. Immediately add 1 mL of LB medium at room temperature, mix and transfer the suspension to an Eppendorf tube.
11. Incubate for 1 h at 37 ºC with shaking, and pour the suspension on LB/agar plates containing ampicillin (300 μg/mL) and tetracycline (20 μg/mL). Incubate the plates at 37 ºC overnight.
12. To determine the complexity of the library, i.e. the

10. Dower WJ, Miller JF, Ragsdale CW. High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Res 1988;16:6127–45.

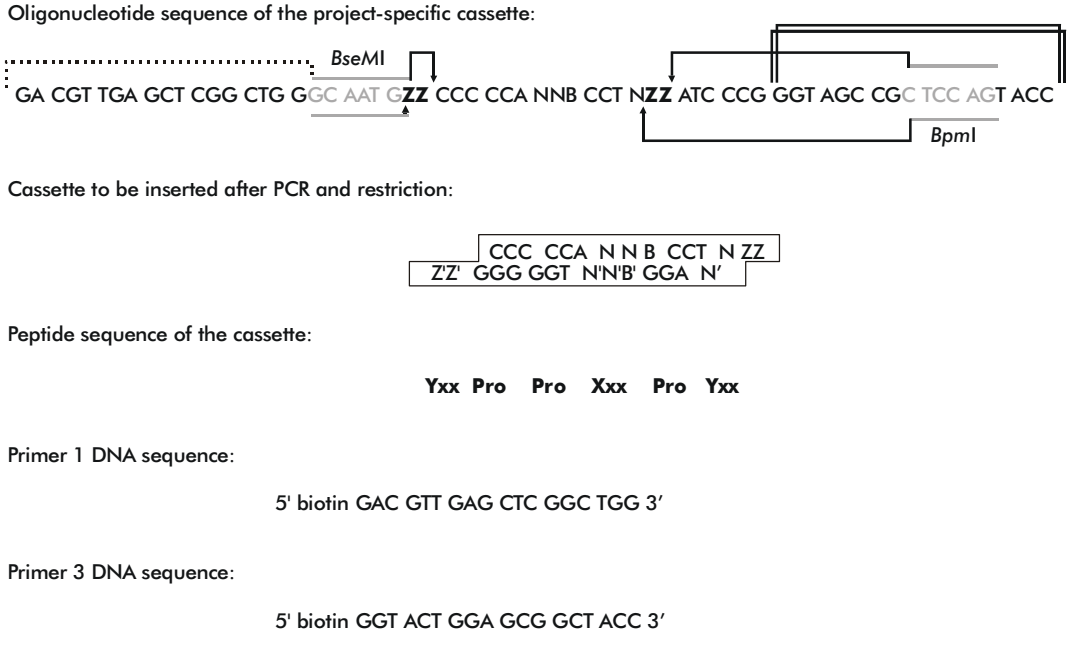


Figure 5. DNA and peptide sequences of the cassettes. On the DNA sequence, recognition and cleavage sites of the restriction enzymes are indicated: *BseM*I and *Bpm*I allow the insertion of the cassettes within the hypervariable region of the primary library. DNA sequences of the biotinylated primers are given, as well as their complementary sequences with the cassettes. ▪▪▪▪▪▪, ═══ indicate the complementary sequences for the primers 1 or 3, respectively. ZZ represents AM (AA, AC), CT, GG, TS (TC, TG), and Yxx, any amino acid other than cysteine. B', N' and Z' stand for the complementary base to B, N and Z respectively.

total number of independently generated clones, perform serial dilutions of the electroporated DNA and pour them on plates prepared according to the previous step. Cells electroporated with the whole vector and those with no DNA were used as positive and negative controls, respectively.

Preparation and insertion of the specific cassette. The next step is to insert this specific cassette within the hypervariable sequence created in the primary library, thus generating a secondary specific library. Invariable amino acid positions are chosen in the design of this cassette, so as to improve the probability of interactions between this secondary library and the particular target.

The codon scheme used for the hypervariable codons is the same one that was used for the initial inserts (i.e NNB, B represents C, G or T).

Sixteen sets of cassettes, which differ only by the two ZZ dinucleotides found in the cohesive ends generated by cleavage of two type II restriction enzymes *(BseM*I and *Bpm*I) were prepared (Figure 6). ZZ stands for AM (AA and AC), CT, GG, and TS (TC and TG). The ZZ dinucleotides present on both cohesive ends of the cassettes must differ (any ZZ dinucleotide found on the right end of the cassette, AM, CT, GG and TS has an equal chance of being present on the left side). These distinctly different ends promote the formation of concatemers during the cloning of the cassette into the primary library, as one cassette cannot be cloned within a single vector. This concatemer formation strongly increases the diversity of the bank.

The sets used at the ZZ positions prevented head to head joining of the homologous ends during the ligation of the cleaved sequence with the primary library. Furthermore, when the 16 cassettes were pooled, it allowed the representation of all amino acids, except cysteine, and it reduced the frequency of the stop codons.

The *BseM*I and *Bpm*I enzymes were used to clone the cassettes within the primary library, allowing a compatibility with the cohesive ends (set of 6) created by the cleavage of the hypervariable insert with *Bpm*I in the initial library. Figure 4 shows the optimized protocol to clone the cassette within the insert of the primary library.

The DNA integrity should be verified in agarose gels between all restriction and ligation steps, and then it should be purified. During the preparation of the cassette 4.5% agarose gels were used and the purification was performed with the phenol/chloroform extraction procedure. However, when preparing the vector and ligating it, 1% agarose gels were used as well as the GFX purification system.

Preparation of the double-stranded specific cassette from degenerated oligonucleotides. The oligonucleotides are converted into double-stranded DNA, and amplified via a PCR, using biotinylated primers (in order to easily remove the restriction fragments using streptavidin-coated beads).

Procedure

1. Prepare the PCR reaction mixture as follows:

Insert	1 µL (10 pmol)
Primer 1	2 µL (200 pmol)
Primer 2	2 µL (200 pmol)
PCR buffer (10x)	10 µL
$MgCl_2$ (25 mM)	10 µL
dNTP (2 mM each)	10 µL

Taq DNA polymerase 0.4 µL (2U)
H_2O up to 100 µL

Use a thermocycler with the following program, for 10 cycles:

94 ºC, 1 min
60 ºC, 1 min
72 ºC, 1 min

2. Test the degree of purity of the cassettes.
3. Purify the DNA (phenol/cloroform extraction protocol).
4. Digest the cassettes with the *Bpm*I enzyme.

These cassettes are ready to be ligated with the vector (Figure 4).

Preparation of the linear vector of the primary library. Use the vector containing the insert (the primary library) as a cloning vector to create a more extended library this extension library.

1. Digest the vector with *Bpm*I to cleave the hypervariable region in the middle.
2. Digest with *BseR*I to divide the vector into two parts, a small (S) and a large (L) fragment.
3. Separate both fragments by electrophoresis in 1% agarose gel and purify them with the GFX purification system. Keep the fragments apart.

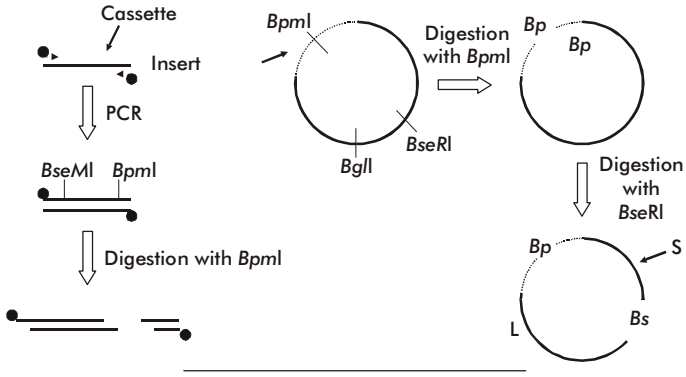
Insertion of the project-specific cassette within the primary library.

1. Ligate the *Bpm*I-digested cassettes with the small fragment (S) of the vector (S-cassette product), and then digest with *BseM*I. Eliminate undesired ligation products and the small biotinylated ends with streptavidin coated beads.
2. Ligate the large fragment (L) and the S-cassette product from the preceding reaction. Use an equimolar ratio and high DNA concentration (>200 µg/mL). This reaction reconstitutes the complete vector, and leads to the formation of concatemers. These concatemers contain monomer units that are oriented in the same direction, since the *BseR*I restriction site produces cohesive ends that are not compatible with those produced at the *Bpm*I or *BseM*I cleavage sites. The ZZ dinucleotide set used did not allow head to head joining of fragments. Since the two ends of the cassettes in most cases do not present identical ZZ dinucleotides, and since they cannot be cloned within the same vector, most of the ligation products will arise from fragments, which come from different clones from the initial library. Hence, the ligation ensured the recombination among the vectors, leading to the formation of a more diverse hypervariable region.
3. Digest with *Bgl*I and ligate at a low DNA concentration (<40 µg/mL) to separate the single phagemid molecules that form the concatemers.
4. Verify the formation of monomers by sequencing randomly sampled clones The analysis of the sequences showed that the extended library displayed a hypervariable region of 13 aa, among which 3 are fixed prolines (insert + cassette), as shown:

DNA sequence: $(NNB)_4$ NZZ CCC CCA NNB CCT NZZ $(NNB)_3$

Peptide sequence: Xxx_4 Xxx Pro Pro Xxx Pro Xxx Xxx_3

4.1. Reassort the 5' and 3' hypervariable blocks following these steps:
(a) digest the final library with *Bpm*I,

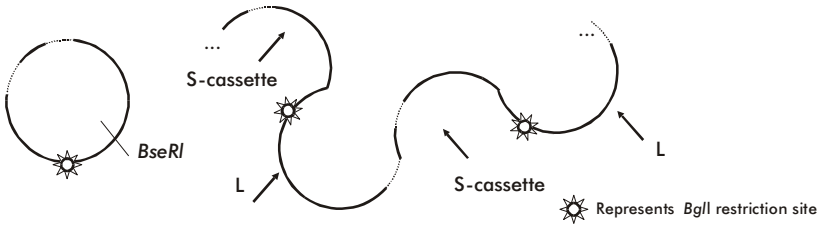


1- Gel extraction of the small (S) and large (L) fragments of the vector, which are kept separated

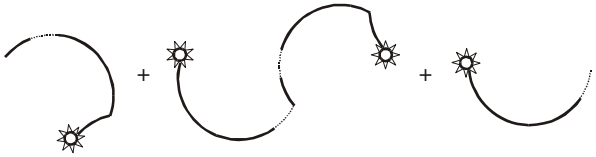
2- Ligation between the cassette and the small (S) fragment of the vector

3- Streptavidin purification

4- Ligation between the large (L) and the small (S) cassette fragments which mainly leads to the formation of concatemers



5- Restriction with *Bgl*I



6- Ligation at a low DNA concentration for the formation of monomers

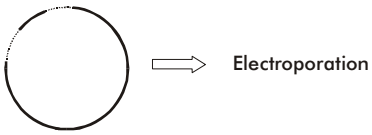


Figure 6. Schematic illustration of project-specific cassette insertion. At the upper section, the cassette is prepared by PCR with biotinylated primers, and it is restricted on one end, with *Bpm*I. The primary library is opened with *Bpm*I and *BseR*I, and the two fragments obtained (S and L) are purified by gel extraction, and remain separated. In the lower section, the half digested cassette is ligated with S (S-cassette), and digested with *BseM*I. Streptavidin-coated beads remove restriction fragments. Ligation occurs between L- and S-cassette in a way to form concatemers, which are finally resolved as monomers by cleavage with *Bgl*I. Ligation at a low concentration generates ring closure of the vector, which now contains the hypervariable region complemented by the project-specific cassette, which can be electroporated. ———, ▪▪▪▪▪▪, ----------, represent vector, insert, and cassette DNA, respectively. ● represents the biotin, linked to the primers. Bp and Bs represent the *Bpm*I and *BseR*I restriction sites, respectively. S and L stand for small and large vector fragments, respectively. S-cassette stands for the DNA ligated form between the cassette and the small fragment of the vector. ☼ represents the *Bgl*I restriction site.

(b) ligate to form concatemers,
(c) separate the monomeric units by cleavage with *Bgl*I,
(d) repeat ligation at a low DNA concentration to obtain a ring closure (Figure 7).

This reassortment is called Cosmix-Plexing.

When Cosmix-Plexing is performed, the type II enzyme divides the hypervariable domain into two segments: one is represented by five random residues, while the second one is composed of a hypervariable region of 8 aa, of which 3 are invariable prolines.

The peptide codified by the nucleotide sequence divided by the type II enzyme has the following structure:

Gly Trp Arg **Xxx_5** + **Pro Pro Xxx Pro Xxx_4** Asp Pro Gly

The bold residues correspond to the hypervariable domain, while the other amino acids belong to the fixed bordering sequences.

4.2. Introduce the construction in *E. coli* by electroporation, and induce the packaging of the secondary library according to that described above. The DNA prepared in this way gives extremely high transformation frequencies (around 10^8 transformants/µg DNA).

Preparation of M13KO7 helper phage stock solution. A phagemid vector carries a plasmid replication origin, in addition to a bacteriophage replication origin, but no bacteriophage genes other than the gene fusion with the protein or peptide to be displayed. The production of particles carrying phagemid is achieved by superinfecting a strain, transformed with the helper bacteriophage derived from M13. This helper phage provides the necessary functions for the synthesis of the circular single-stranded DNA and bacteriophage capsid.

Procedure

1. Use a sterile disposable pasteur pipette to pick a single, well separated M13KO7 plaque from a host of the *E. coli* WK6 strain grown overnight on a LB/agar plate. Inoculate 20 mL of the LB(2x) medium with kanamycin (50 µg/mL) with this agar slice, and incubate all day at 37 ºC on a shaker.
2. Inoculate 1 L of the LB(2x) medium with kanamycin (50 µg/mL) with 10 mL of the preculture, and incubate overnight at 37 ºC, on a shaker.
3. Centrifuge the culture at 11 000 x*g* and 4 ºC (Sorvall GS3 rotor). Transfer the supernatant into fresh centrifuge tubes, and repeat the centrifugation step.
4. Transfer the supernatant into fresh centrifuge tubes, and add a 0.15 volume of the polyethylene glycol 6000 (PEG)/NaCl solution (16.7%/3.3 M). Mix, and incubate on ice for at least 2 h.
5. Centrifuge 1.5 h at 13 000 x*g* and at 4 ºC (Sorvall GS3 rotor). Eliminate the supernatant by decanting, and centrifuge the sediment for another few minutes. Remove the last trace of the supernatant with a pipette, and resuspend the helper phage in 10 mL PBS.
6. Centrifuge the resuspension for 10 min at 17 000 x*g* and 4 ºC (Sorvall SS-34 rotor). Recover the supernatant, add NaN_3 to a final concentration of 0.02%. Store at 4 ºC.
7. *Titrate of the M13KO7 stock solution.* Inoculate 20 mL of the LB medium with tetracycline (20 µg/mL) with 200 µL of WK6λmutS culture

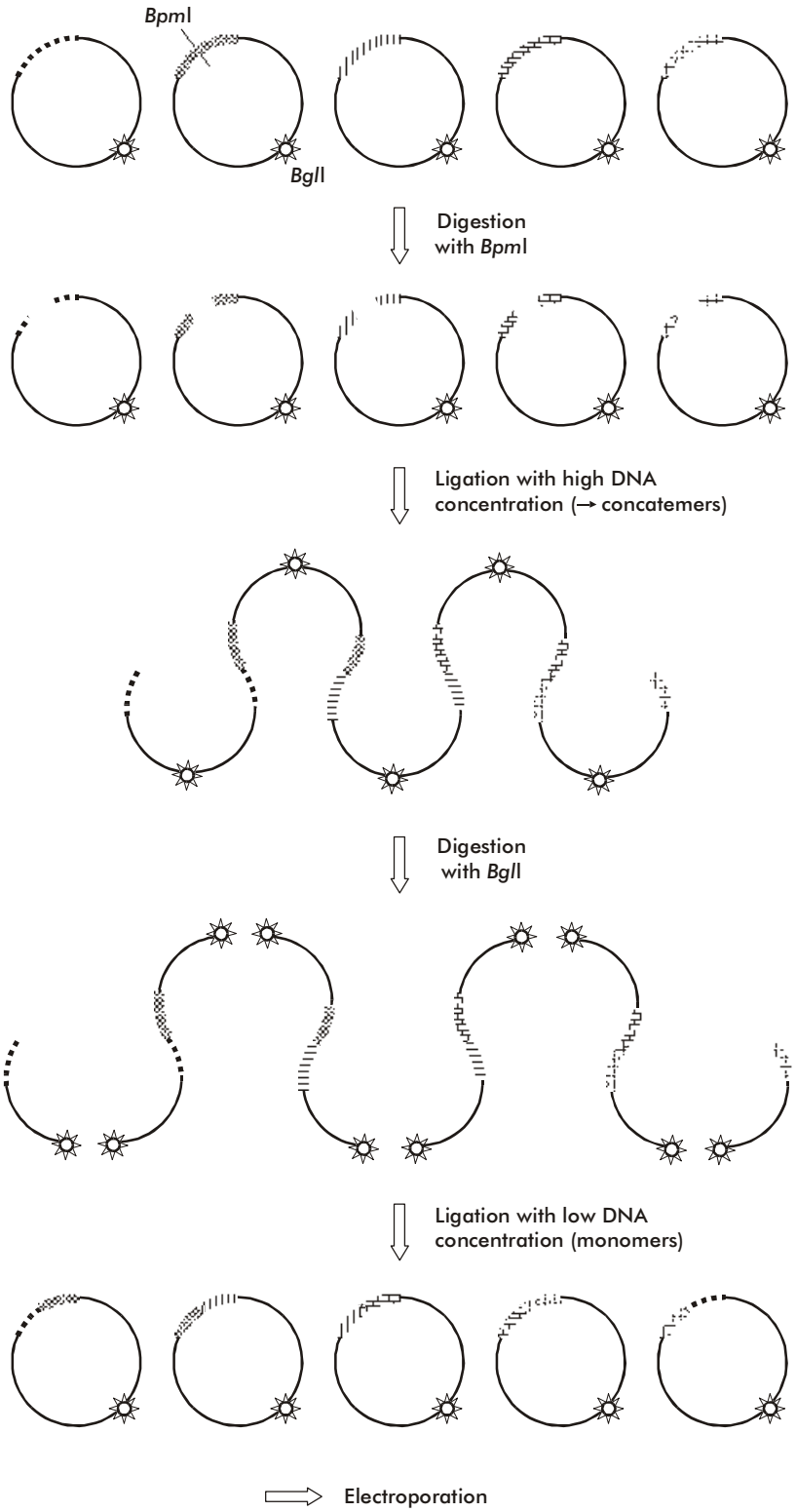


Figure 7. Schematic of Cosmix-Plexing recombination. The initial material is a phagemid preparation obtained from the resuspended colonies infected with the eluted phage from the first round of panning. The vectors are opened with *Bpm*I, and religated at a high DNA concentration to yield the formation of concatemers. The concatemers are resolved as single vectors by restriction with *Bgl*I, and are ligated with a low DNA concentration, as closed monomers. The recombined vectors can be electroporated, packaged, and used for further panning to isolate optimized ligands against a particular target. ☼ represents the *Bgl*I restriction site.

and incubate at 37 ºC with shaking, until an OD_{600nm} of 0.5 is reached.
8. During this time, prepare a serial dilution of the M13KO7 stock (10 µL phage in a total volume of 100 µL ddH_2O).
9. Add 100 µL of the logarithmic growing cells from the step 7, mix, and incubate for 30 min at 37 ºC.
10. Spot 20 µL aliquots of the serial dilution on LB/agar (ampicillin 300 µg/mL and tetracycline 20 µg/mL), and on LB/agar (kanamycin 50 µg/mL and tetracycline 20 µg/mL), and LB/agar (tetracycline 20 µg/mL) as a control. Incubate the plates overnight at 37 ºC.
11. Count the number of colonies in each spot and determine the titer expressed as colony forming units per milliliter (cfu/mL). The titer should be of 10^{12} to 10^{13} cfu/mL.

Amplification and packaging of the phagemid libraries.

1. Resuspend the colonies from the plates where the electroporated cells have been amplified overnight. For this, overlay the LB/agar plates with 15 mL of the LB medium with ampicillin (300 µg/mL) and tetracycline (20 µg/mL). Rock the plates for 20 min at room temperature, and wash the cells by flushing the medium up and down several times using a 10-mL pipette (the tip diameter is large enough so as to avoid damaging the cells).
2. Inoculate a 1-L LB medium complemented with ampicillin (300 µg/mL) and tetracycline (20 µg/mL), with 2.5 mL of the resuspension. Incubate at 37 ºC while shaking until the OD_{600} reaches 0.5.
3. Add 10^{12} cfu of the helper phage M13KO7 to the culture. Incubate at 37 ºC while shaking for 1 h, and then, overnight while shaking at 30 ºC.
4. Collect the cell culture by centrifugation for 10 min at 13 000 x*g*, at 4 ºC (Sorvall GS3 rotor). Transfer the supernatant into a fresh centrifuge tube. Repeat the centrifugation step.
5. Transfer the supernatant into a fresh centrifuge tube, and add a 0.15 volume of the PEG/NaCl solution (16.7%/3.3 M). Mix, and incubate on ice for at least 2 h.
6. Centrifuge for 1.5 h at 13 000 x*g,* at 4 ºC (Sorvall GS3 rotor), to harvest the phage particles. Eliminate the supernatant by decanting, and centrifuge for a few more minutes. Remove the last trace of supernatant with a pipette.
7. Resuspend the pellet in 10 mL PBS, and centrifuge for another 10 min at 17 000 x*g* (Sorvall SS-34 rotor).
8. Recover the supernatant, add NaN_3 to a final concentration of 0.02%, and store the phage particles at 4 ºC.

Application of the Cosmix-Plexing system

Cosmix-Plexing of the libraries. Cosmix-Plexing can be used in order to insert an additional specific cassette within the hypervariable domain of a primary library, as well as to amplify the number of variants of a particular library. However, in most cases, Cosmix-Plexing is used to optimize the binding capacity of ligands identified against a target of interest. As explained previously, the Cosmix-Plexing procedure allows the generation of recombination within the hypervariable domain of the bank, by reassorting the left and the right sections of the hypervariable region. This is achieved by using type II restriction enzymes.

At the start of the procedure a "standard" selection against a target of interest was performed, and after a few rounds of affinity selection, some variants were identified. These variants exhibited appropriate properties to interact with the target molecule. Typically, few clones were enriched among which some unique variants, highly related to the enriched dominant clone(s), are also present, so that a consensus sequence could be conceived. This consensus sequence represents the essential structure required for binding with this target. During the first round of panning, the future dominant clones were present in small amounts, as well as many other clones that exhibited weak affinity for the target. The more the dominant clone(s) are enriched, during further rounds, the greater their competition with the weaker binding clones, which were then eliminated during the washing steps.The Cosmix-Plexing system takes advantage of this.

The preselected population recovered after the first round of panning, is the working material submitted to the recombination. At this stage, most of the clones present are able to show an affinity capacity for the target, no matter how weak. The hypervariable regions of the clones of this preselected population may recombine, producing new variants. Hence, through the binding of initially weak domains may generate clones with new characteristics that may correspond to the "perfect" structure for binding with the target. Further selection rounds with this optimized library leads to the isolation of new clones, with an assumable optimun affinity with the target. Cosmix-Plexing was applied to a phage preparation obtained from an aliquot of resuspended cells conserved in glycerol and infected with eluted phage, after the first (and eventually second) round of panning. The strategy of this approach is described on Figure 7.

Between all ligation and restriction steps, the DNA integrity is verified on a 1% agarose gel, and then purified using the GFX system.

Procedure

1. Using *Bpm*I digest the DNA of the preselected population that was recovered as a plasmid from the preparation. This enzyme cleaves the hypervariable region in the middle.
2. Ligate the open vector at a high DNA concentration (>200 µg/mL) leading to the formation of concatemers. This ligation ensures a recombination among the vectors, increasing the diversity of the hypervariable region. Remember that the ZZ set used at the opened position of the variable domain, does not allow head to head joining of fragments.
3. Separate the monomeric units forming the concatamers through *Bgl*I digestion. Ligate at a low DNA concentration (<40 µg/mL).
4. Verify the formation of the monomers by sequencing randomly selected clones.

The optimized library is now ready for its introduction in *E. coli* through electroporation and for packaging. The DNA prepared in this way gives extremely high transformation frequencies (around 10^8 transformants/µg DNA).

11. Brackeman PR, Lanahan AA, O'Brien R, Roche K, Barnes CA, Huganir RL, *et al*. Homer: a protein that selectively binds metabotropic glutamate receptors. Nature 1997;386:284–8.

12. Tu JC, Xiao B, Yuan J, Lanahan AA, Leoffert K, Li M, *et al*. Homer binds a novel proline-rich motif and links group 1 metatropic glutamate receptors with IP3 receptors. Neuron 1998;21:717–26.

Results

Panning of the CPLPPXP library against different targets, from SH3 and EVH1 domain families, leads to the design of long consensus sequences, and to the isolation of strong binders against these particular targets. Only one example will be treated here, as a validation of our technology, on the Vesl EVH1 domain.

Vesl (also called Homer) is constitutively expressed in the brain and enriched at excitatory synapses, which selectively bind to the C-terminus of group 1 metabotropic receptors (mGluR1a and mGluR5) [11]. The Vesl region, which interacts with mGluR1a/5, is termed an EVH1 domain, based on the homology to domains in a family of proteins that include *Drosophilia* enabled, mammalian VASP, and the Wiscott-Aldridge syndrome protein (WASP). The Vesl EVH1 domain binds to an internal, proline-rich sequence that is at approximately 50 residues from the C-terminus of both mGluR1a and mGluR5 [12].

A polyproline motif, Glu/Asp Phe Pro Pro Pro Thr Glu/Asp, present in the ActA protein of the intracellular bacterial pathogen *Listeria monocytogenes,* serves as a ligand for the VASP, Mena, and Evl protein family, which share a highly conserved EVH1 domain [13]. Other natural ligands of EVH1 domains (vinculin, and zyxin) also bind to the EVH1 domains due to a similar sequence, with respect to the acidic charges present on both sides of the conserved proline core, and to the phenylalanine located immediately before the proline motif.

Recent work on Vesl performed by Tu *et al.* [14], proved that Shank proteins, postsynaptic molecules that function as a part of the PSD-95 complex associated to the NMDA receptor, bind to Vesl, through the sequence Leu Val Pro Pro Pro Glu Glu Phe Ala Asn. In this case, acidic residues are no longer surrounding the proline core, since we find them only on the C-terminal part of the proline motif. Moreover, the Phe is present, but no longer binding to the proline core. Furthermore, it is, in this case, located on the C-terminal end of the prolines, in contrast to that of the sequences of the natural peptides that bind with VASP, Mena and Evl. These results bring evidence to the fact that the EVH1 domains of VASP, Mena and Evl do not bind to the same ligands as the EVH1 of Vesl, although they share certain general similarities. It must be emphasized that the results cited above [13, 14] were obtained from the yeast two-hybrid system.

The work performed in our laboratory began with "normal" affinity selections achieved on Vesl. After the third round, in which 200-fold enrichment was obtained, we sequenced randomly picked clones, and analyzed them (Table 1).

The analysis of these clones showed the great importance of a tyrosine or tryptophan after the proline rich motif, on position 10, and seems to indicate a preference for Arg on position 2, and Aspartic acid, on position 12. Position 8, between the prolines, presents a good advantage for hydrophobic residues, as well as positions 3 and 4.

The population recovered after the first round of "normal" panning, was submitted to a Cosmix-Plexing, and screened once more against Vesl. Randomly selected clones were sequenced after the first round of panning, since our experience demonstrated that after Cosmix-Plexing, enriched clones are efficiently obtained from the first cycle of selection. The sequences are presented on Table 2. The first element to be observed after cosmix-plexing, regarding to the Vesl protein, is that it confirms the results found after the "normal" panning: the dominant clone is the same in both cases.

However, the Cosmix-Plexing approach allows to obtain a longer consensus sequence (only one of the thirteen positions is still undetermined). In addition,

13. Niebuhr K, Ebel F, Frank R, Reinhard R, Domann E, Carl UD, *et al*. A novel proline-rich motif present in ActA of *Listeria monocytogenes* and cytoskeletal proteins is the ligand for the EVH1 domain, a protein module present in the Ena/VASP family. EMBO J 1997;16:5433–44.

14. Tu JC, Xiao B, Naisbitt S, Yuan JP, Petralia RS, Brakeman P, *et al*. Coupling of mGluR/homer and PSD-95 complexes by the shank family of postsynaptic density proteins. Neuron 1999;23:583–92.

Table 1. Sequences of 25 randomly picked clones, after a third round of panning selection of the CPLPPXP library, against Vesl EVH1 domain (one clone was not readable). Underlined residues are fixed within the CPLPPXP library. φ and Ψ represent the aromatic and hydrophobic residues, respectively.

Sequences against Vesl EVH1 domain (third round of panning)												Frequency
Asp	Arg	His	Tyr	Arg	Pro Pro	Phe	Pro	Trp	Ala	Asp	Gly	13x
Asn	Arg	Leu	Tyr	Pro	Pro Pro	Trp	Pro	Tyr	Ser	Asp	Pro	4x
Ser	Arg	Ser	Val	Tyr	Pro Pro	Pro	Pro	Tyr	Pro	Phe	Ala	2x
Asp	Arg	Ile	Tyr	Leu	Pro Pro	Val	Pro	Trp	Ala	Asn	Ser	1x
Arg	Thr	Gly	Val	Trp	Pro Pro	Pro	Pro	His	Asp	Phe	Arg	1x
Tyr	Leu	Ile	Leu	Ser	Pro Pro	Ala	Pro	Trp	Arg	Asp	Arg	1x
Cys	Ser	Arg	Ser	Arg	Pro Pro	Val	Pro	Leu	Gly	Pro	Phe	1x
His	Met	Leu	Phe	Asp	Pro Pro	Phe	Pro	Tyr	Ser	Asn	Glu	1x
Consensus sequence												
		Xxx	(Arg)	Ψ(φ/Ψ)	Xxx	Pro	Pro (φ/Ψ)	Pro	W/Y	Xxx	D/F	Xxx

Table 2. Sequences of 10 randomly picked clones, after a first round of panning selection. The CPLPPXP population recovered after the first round of "normal" panning against the Vesl EVH1 domain, was submitted to cosmix-plexing, and screened in this panning. Underlined residues are fixed within the CPLPPXP library. φ and Ψ represent the aromatic and hydrophobic residues, respectively.

Sequences against Vesl EVH1 domain after cosmix plexing (first round of panning)													Frequency
Asp	Arg	His	Tyr	Arg	Pro	Pro	Phe	Pro	Trp	Ala	Asp	Gly	6x
Ser	His	Ile	Phe	Asp	Pro	Pro	Phe	Pro	Tyr	Gly	Pro	Met	1x
Ser	Arg	Ser	Val	Tyr	Pro	Pro	Pro	Pro	Tyr	Pro	Phe	Ala	1x
Ser	Leu	Leu	Tyr	Glu	Pro	Pro	Pro	Pro	Trp	Asn	Ser	Pro	1x
Ser	Arg	Val	Tyr	Pro	Pro	Pro	Phe	Pro	Trp	Arg	Ala	Val	1x
Consensus sequence													
Ser	Arg	Ψ	φ	(α)	Pro	Pro	F/P	Pro	W/Y	Xxx	Ψ	Ψ	

fewer ambiguities within the consensus sequence are found. A tyrosine or tryptophan residue was verified after the proline core motif, in position 10, along with the Arg2. The residue in position 8, between the prolines, appears to show a preference for a phenylalanine, or a proline, rather than just any hydrophobic residue. The fourth position was clearly determined as tyrosine, or eventually a phenylalanine residue (but in any case, an aromatic residue), which was not obvious after the "normal" panning.

Furthermore, the Cosmix-Plexing approach gives more information on both ends of the consensus sequence. In fact, in the first position a serine seems to be important, and in the last two positions (12, and 13), hydrophobic residues.

The clones characterized after the panning (Tables 1 and 2), were prepared as single phages, and individually checked on phage ELISA for affinity against the Vesl protein. The results of this ELISA confirm the preference of the target for the dominant clone, as well as the ability of the other selected clones to bind to the Vesl protein. BIAcore analysis has not yet been performed.

These results, obtained from a phage display library, may be comparable, to some extent, to those obtained from other groups, with the yeast two hybrid system [13, 14]. Our results mainly insist on a crucial role of an aromatic residue directly after the proline motif, fixed in our library (however Trp and Tyr appear to be more advantageous than Phe). In addition, acidic residues are also present on both sides of the prolines, even if a particular position was not clearly determined. To this respect, our consensus sequence looks like a mirror image sequence of the natural ligands of VASP, Mena and Evl.

However, it may also be considered that our consensus sequence shows similarities to the Shank sequence that binds with Vesl, since in both cases, we find the aromatic residue on the same side of the proline core, and separated from it (or a part of it as in the case of the CPLPPXP library) by two residues. In addition, a similar hydrophobic position toward the N-terminal of the proline motif was found homologing the Shank sequence, and the phage-display consensus sequence as shown in the following sequence: Leu/Ile Xxx Xxx Pro Pro Xxx Xxx Arom. It should not be emphasized that the interactions of proline-rich peptides with EVH1 domains takes place through the formation of a left-handed polyproline type II helix (PPII). This PPII helix contains three residues per turn, with a triangular cross-section. One face (two residues in each turn with the lateral chains facing inward) is contacting the floor of EVH1 domain binding groove, whereas the third residue of each turn forms a spine that does not interact with the domain, but is essential for the stability of the helix [15]. The regularity of the previous sequence, regarding the number of residues that might be important for binding, and the space in-between, might suggest the shape of this helix: $Leu/Ile_1\ Xxx_2\ Xxx_3\ Pro_1\ Pro_2\ Xxx_3\ Xxx_1\ Arom_2$, where residues $_1$ and $_2$ are located on the two spines which make a contact with the Vesl, whereas residues $_3$ stabilized the structure.

Conclusions

We succeeded in building a high quality bank considering several factors such as: its diversity ($>10^8$ recombinants), the frequency of the amino acids found at each position, that is very similar to the frequency expected for a random NNB library, the low frequency of non productive clones, very few stop codons or frameshifts of the open reading frame, and the absence of parental religated vectors, which would not contain any insert.

Our specialized library obtained through Cosmix-Plexing confirms that this approach is of great interest. We efficiently isolated strong ligands against several targets of the SH3 and EVH1 families, using the initial proline-rich library. Long consensus sequences have been readily obtained. Furthermore, the results of panning with the Cosmix-Plexing recombined proline rich library allows to extend the consensus sequences, and to reduce their potential ambiguities.

Acknowledgements

We thank Lawrence A. Quilliam and Brian K. Kay, for kindly providing the GST-SH3 domains, Prof Wehland and Dr. Uwe D. Carl, for kindly supplying GST-EVH1 proteins. Thanks also to Peter Röttgen for helpful discussions and encouragement.

15. Adzhubei AA, Sternberg MJE. Left-handed polyproline II helices commonly cccur on globular proteins. J Mol Biol 1993;229:472–93.

Received in March, 2001. Accepted for publication in August, 2001.