Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 7761–7766, July 1996
Evolution

A strategy of exon shuffling for making large peptide repertoires displayed on filamentous bacteriophage

(phage/self-splicing introns/combinatorial library)

IGOR FISCH*†, ROLAND E. KONTERMANN*, RICARDA FINNERN‡, OLIVER HARTLEY*, ANDRES S. SOLER-GONZALEZ§, ANDREW D. GRIFFITHS*¶, AND GREG WINTER*¶

*Medical Research Council Centre for Protein Engineering and ¶Medical Research Council Laboratory for Molecular Biology, Hills Road, Cambridge CB2 2QH, United Kingdom; §Medical Research Council Unit for Protein Function and Design, Department of Chemistry, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, United Kingdom; ‡Division of Transfusion Medicine, University of Cambridge, Long Road, Cambridge CB2 2PT, United Kingdom

Communicated by F. Sanger, Cambridge, United Kingdom, March 25, 1996 (received for review November 28, 1995)

ABSTRACT It has been suggested that recombination and shuffling between exons has been a key feature in the evolution of proteins. We propose that this strategy could also be used for the artificial evolution of proteins in bacteria. As a first step, we illustrate the use of a self-splicing group I intron with inserted *lox*-Cre recombination site to assemble a very large combinatorial repertoire (>10^{11} members) of peptides from two different exons. Each exon comprised a repertoire of 10 random amino acids residues; after splicing, the repertoires were joined together through a central five-residue spacer to give a combinatorial repertoire of 25-residue peptides. The repertoire was displayed on filamentous bacteriophage by fusion to the pIII phage coat protein and selected by binding to several proteins, including β-glucuronidase. One of the peptides selected against β-glucuronidase was chemically synthesized and shown to inhibit the enzymatic activity (inhibition constant: 17 nM); by further exon shuffling, an improved inhibitor was isolated (inhibition constant: 7 nM). Not only does this approach provide the means for making very large peptide repertoires, but we anticipate that by introducing constraints in the sequences of the peptides and of the linker, it may be possible to evolve small folded peptides and proteins.

Recently, attempts have been made to harness the forces of Darwinian evolution to create artificial peptides and proteins "ligands" with *de novo* binding activities to "receptors." In principle, this involves the synthesis of large and diverse repertoires of genes, the expression of the encoded polypeptides, and the selection of those with binding activities (for review, see ref. 1). In particular, the use of filamentous bacteriophage offers a powerful means of evolving polypeptide ligands, as the ligand can be displayed on the surface of the phage by fusion of the ligand gene to that of a phage coat protein. The fusion phage (and encapsidated ligand gene) can be readily selected by binding to a solid phase "receptor."

Most genes for eukaryotic proteins contain two or more exons separated by introns; the introns are spliced from the mRNA before translation. Indeed, it has been proposed that the presence of these introns might have facilitated protein evolution by recombination between exons ("exon shuffling") (2, 3). This is most evident in multidomain proteins comprising domains with different functions, where each domain is located on a different exon(s). For example, rabbit endothelial leukocyte adhesion molecule-1 (ELAM-1) consists of an N-terminal lectin domain, an epidermal growth factor domain, several complement regulatory elements, a single transmembrane sequence, and a short cytoplasmic tail (4). An intron is located at the boundary of each sequence coding for the different domains.

Introns are also located within domains of eukaryotic proteins (5–7), and it has also been suggested that the architecture of domains may have evolved through recombination between exons (for review, see ref. 8). This inspired us to develop a strategy of exon shuffling toward the artificial evolution of proteins on the filamentous bacteriophage fd. We created two exon repertoires, each comprising 10 peptide residues. Both repertoires were separated by a self-splicing group I intron from *Tetrahymena thermophila* 26S rRNA, which undergoes accurate and efficient cleavage-ligation in bacteria (9, 10).

For recombination between exons, we introduced a *loxP* site from P1 phage into the self-splicing intron (see Fig. 1). By locating each exon on a different replicon, we were able to generate a large combinatorial repertoire of exons using the process of "combinatorial infection" (11). Thus, the first (N-terminal) exon repertoire was located on a plasmid replicon and the second exon repertoire was fused to the gIII protein of filamentous bacteriophage on a phage replicon. Bacteria harboring the first repertoire were infected with phage harboring the second repertoire. Induction of the Cre recombinase led to recombination at the *loxP* sites to generate a single combinatorial repertoire of exons fused to the filamentous phage. The phage repertoire was selected for binding activities against one monoclonal antibody and two enzymes.

MATERIALS AND METHODS

Bacterial Strains. The *Escherichia coli* strains TG1 [F′ *traD*36 *lacI*^{q} Δ(*lacZ*)M15 *pro*A^{+}B^{+}/*sup*E Δ(*hsd*M-*mcr*B) 5($r_k^- m_k^-$McrB$^-$) *thi* Δ(*lac-proAB*)] (12), and MC1061 [F$^-$ *araD*139 Δ(*ara-leu*)7696 *galE*15 *galK*16 Δ(*lac*)X74 *rpsL* (Str^r) *hsdR*2 ($r_k^- m_k^+$) *mcrA mcrB*1] (13) were used.

Construction of the Intron *loxP* Wild Type (wt). The wt *loxP* site (14) was inserted at position 236 of the group I intron from *T. thermophila* 26S rRNA (15) contained in the vector M13 mICE10 (16). Site-directed mutagenesis was performed with an oligonucleotide containing the *loxP* wt site (oligo-2889, Table 1) according to the manufacturer's instructions (Sculptor *in vitro* mutagenesis system, Amersham) and the DNA transformed into *E. coli* TG1 (17).

Construction of the pUC19-Exon 1 Repertoire. The intron containing the *loxP* wt was amplified by PCR with *Taq* polymerase using oligo-3194 (Table 1) and oligo-3198 (Table 1). The reaction mixture (500 μl) was cycled 30 times (94°C for 1 min, 55°C for 1 min, 72°C for 45 s). The resulting fragment was then subcloned into *Sfi*I/*Eco*RI-digested pUC19–2lox (11) and electroporated (17) into 10 aliquots of 40 μl *E. coli* TG1/pACYCaraCre (S. C. Williams, unpublished work). The plasmid produced (pUC19-exon 1, see Fig. 2*A*) contains a group I self-splicing intron flanked at its 5′-end by random nucleotides encoding a 10 residue exon and at its 3′-end by a *Eco*RI restriction site. The frequency of inserts was checked by

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: wt, wild type; t.u. transducing units.

†To whom reprint requests should be sent at present address: University of Lausanne, Biochemistry Institute, 155 ch. des Boveresses, 1066 Eralinges, Switzerland.

Table 1. Oligonucleotide used for the construction and the characterization of the vectors pUC19-exon 1 and fdDOG-exon 2

Oligo-2889	5′-TGAACTGCATCCATATCAACAGAAGATAACTTCGTATAGCATACATTATACGAAGTTATATCTGTTGACTTAGGACTTGGCTGC-3′
Oligo-3249	5′-CATGACCACAGTGCACAGTAATAATATAACTTCGTATAATGTATACTATACGAAGTTATGCATGCAAATTCTATTTCAAGGAGACAGTC-3′
Oligo-3483	5′-TTACTCGCGGCCCAGCCGGCCATGGCCGCTCTCTAAATAGCAATATTTACCTTTGGAG-3′
Oligo-3193	5′-CTAATTAAGGGCGGCCGCMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNATACCTTACGAGTACTCCAAAACTAATCAATATA-3′
Oligo-3194	5′-TTACTCGCGGCCCAGCCGGCCATGGCCNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKGCTCTCTAAATAGCAATATTTACCTTTGGAG-3′
Oligo-3198	5′-CAGCGTCACCGGAATTCATACCTTACGAGTACTCCAAAACTAATCAATATA-3′
Oligo-4226	5′-GCGTGGTTAGGTCCATGTCCGTCAGC-3′
Oligo-4445	5′-ACTTGGTACTGAACGGC-3′
Oligo-3358	5′-GAAGTGATGCAACACTGGAGC-3′
LMB2	5′-GTAAAACGACGGCCAGT-3′
fdPCRFOR	5′-TAGCCCCCTTATTAGCGTTTGCCA-3′

N, an equimolar mixture of all four bases; K, an equimolar mixture of G and T; M, an equimolar mixture of A and C.

PCR screening; of 32 clones, all contained inserts of the correct size and the sequence of 12 of these clones was checked and shown to differ only in the exon. The size of the exon 1 repertoire was calculated as 1.2×10^7.

Construction of the fdDOG-Exon 2 Repertoire. The Fd fragment of the antibody NQ10.12.5 (18) was amplified from the vector pUC19 NQ10 κ (19) using oligo-3249 (Table 1) and oligo-LMB2 (Table 1). The DNA was then cloned into fdDOG cut with *Apa*L1 and *Not*I to produce the vector fd-BLX. The intron containing the *loxP* wt was amplified by PCR with *Taq* polymerase using oligo-3483 (Table 1) and oligo-3193 (Table 1). The reaction mixture (500 μl) was cycled 30 times (94°C for 1 min, 55°C for 1 min, 72°C for 45 s). The resulting fragment was then subcloned into *Sfi*I/*Not*I-digested fd-BLX and electroporated (17) into 10 aliquots of 40 μl *E. coli* MC1061. The phage produced (fdDOG-exon 2, see Fig. 2*A*) contains a group I self-splicing intron flanked at its 5′-end by the pelB leader peptide and at its 3′-end by random nucleotides encoding a 10 residue exon. As for pUC19-exon 1, the repertoire was checked by PCR screening; 20/24 clones contained inserts of the correct size, and the sequence of 12 of these clones was checked and shown to differ only in the exon. The size of the exon 2 repertoire was calculated as 6.2×10^5.

Combinatorial Infection and *in Vivo* Recombination. The recombination was performed essentially as in ref. 20, however, Cre recombinase was supplied by the plasmid pACYCaraCre, which carries the Cre gene under the control of an *araB* promoter (21) instead of using phage P1 (S. C. Williams, unpublished work). Approximately 10^9 *E. coli* MC1061 cells harboring the fdDOG-exon 2 library (Table 2, sample point 1) were used to inoculate 1 liter of 2xTY broth containing tetracycline at 15 μg/ml, and the culture was shaken for 30 hr at 30°C in a 2-liter baffled Erlenmeyer flask. Phage were purified from the supernatant by precipitation with polyethylene glycol as in ref. 22 and resuspended in 10 ml phosphate-buffered saline (PBS; 25 mM NaH_2PO_4/125 mM NaCl, pH 7.0). Yields were typically 10^{11} transducing units (t.u.) per ml of culture. Approximately 10^8 *E. coli* TG1 harboring pACYCaraCre cells and the pUC19-exon 1 library (Table 2, sample point 2) were used to inoculate 200 ml 2xTY broth containing 100 μg/ml carbenicillin, 25 μg/ml chloramphenicol, 1% (wt/vol) glucose, and 2 g/l of glycerol. After overnight growth at 37°C, 150 ml of this culture (Table 2, sample point 3) was used to inoculate 15 liters of the same medium in a 20 liter fermentor (BioFlo IV, New Brunswick Scientific) and the culture was grown at 37°C to an OD_{600} of 0.4. The library (2.2×10^{12} t.u.) in fdDOG-exon 2 was then added and the culture was grown for 15 min at 37°C before harvesting by tangential flow filtration using a Pellicon cassette (0.45 μm filter, Millipore) at 4°C to remove the excess phage (Table 2, sample point 4). The bacteria were resuspended in 15 l of 2xTY broth containing 100 μg/ml carbenicillin, 25 μg/ml chloramphenicol, 15 μg/ml tetracycline, 0.5 g/liter L(+)arabinose (Sigma), and 2 g/liter of glycerol, and was grown for a further 30 hr at 30°C. The phage were harvested from the filtrate after tangential flow filtration by precipitation with polyethylene glycol as before (22) and resuspended in a final volume of 32 ml PBS; the total yield of fd phage was 1.4×10^{14} t.u. (Table 2, sample point 5). Aliquots of bacteria were plated on media containing different antibiotics for determination of colony-forming units at various time point during the recombination to calculate the overall repertoire size (Table 2). The efficiency of the recombination was checked by PCR screening (23) using oligo-4226 (Table 1) and pelBBack (20).

Selection. The phage repertoire was selected by panning on coated immunotubes (Maxisorb, Nunc) as in refs. 20 and 22. Recombined phage (6.2×10^{12} t.u.) were used for the first round of selection and $\approx 10^{12}$ t.u. were used for the subsequent rounds. Antigens were coated overnight at 4°C at a concentration of 10 μg/ml in PBS.

Sequencing of Selected Peptides. Individual clones were amplified by PCR using *Taq* polymerase and oligo-fdPCRFOR (Table 1) and oligo-fdPCRBACK (20). PCR cycle sequencing reactions with fluorescent dideoxy chain terminators (Applied Biosystems) were carried out according to the manufacturer's instructions, with oligo-4445 (Table 1) and oligo-3358 (Table 1). Sequencing reactions were analyzed on an Applied Biosystems model 373A Automated DNA Sequencer and se-

Table 2. Determination of the repertoire size

Sample point	Total number of colony forming units ($\times 10^{-10}$)				
	No	$a^R + c^R$	t^R	$a^R + c^R + t^R$	t.u. total
1			0.05		
2		0.03			
3		34			
4					1.9×10^{11}
5					1.4×10^{14}
6	190*	80†	16‡	16§	
7	0.35*	0.3†	0.2‡	0.061§	
8	1.1*	0.9†	0.75‡	0.55§	

Sample point, see *Materials and Methods*: point 1, *E. coli* containing fdDOG-exon 2 from frozen stock; point 2, *E. coli* containing pUC19-exon 1 and pACYCaraCre from frozen stock; point 3, *E. coli* containing pUC19-exon 1 and pACYCaraCre after overnight culture; point 4, titer of fdDOG-exon 2 phage in the filtrate; point 5, titer of fdDOG-exon 1 phage after 30 hr; point 6, determination of the size of the repertoire; point 7, determination of the size of the reshuffled repertoire selected against AP; point 8, determination of the size of the reshuffled repertoire selected against β-Gluc. No, no antibiotics; a^R carbenicillin resistant; c^R, chloramphenicol resistant; t^R, tetracyclin resistant; AP, bovine alkaline phosphatase; β-Gluc, *E. coli* β-glucuronidase. *Total number of *E. coli*.

†*E. coli* containing pUC19-exon 1 and pACYCaraCre.

‡*E. coli* infected with fdDOG-exon 2.

§*E. coli* containing pUC19-exon 1, pACYCaraCre and fdDOG-exon 2.

quence analysis was performed using SEQED (Applied Biosystems) and MACVECTOR (IBI–Kodak).

ELISA Screening of Selected Peptides. Single tetracycline-resistant colonies were screened to identify those producing antigen-binding phage by ELISA essentially as in ref. 24. Binding was detected using horseradish peroxidase-conjugated anti-M13 antibody (Pharmacia) with 3′,3′,5′,5′-tetramethylbenzidine and hydrogen peroxide (TMB, Pierce) as a substrate. Reactions were stopped with 50 μl of 1M H_2SO_4 after 20 min and readings were taken by subtracting the A_{650} from the A_{450}.

Reshuffling of Exons. Single colonies harboring the appropriate sequence at the 5′-end of the intron were amplified by PCR with *Taq* polymerase using oligo-pelBBACK (20) and oligo-3198 (Table 1). The reaction mixture (50 μl) was cycled 20 times (94°C for 1 min, 55°C for 1 min, 72°C for 45 s). The DNA was then cloned into *Sfi*I/*Eco*RI-digested pUC19-2lox and electroporated into *E. coli* TG1/pACYCaraCre. The plasmids pUC19-AP4 and pUC19-Gluc4 that were produced contain a group I self-splicing intron flanked at their 5′-end by nucleotides encoding a defined 10 amino acid peptide and at their 3′-end by a *Eco*RI restriction site. Approximately 10^7 TG1/pACYCaraCre harboring pUC19-AP4 or pUC19-Gluc4 were used as the inoculum for *in vivo* recombination, which was performed essentially as above but in a flask, on a scale of 50 ml with 10^{10} t.u. of the fdDOG-exon 2 library added. The ultrafiltration was replaced by centrifugation for 10 min at 4000 rpm.

Peptide Synthesis and Inhibition Assays. Peptides were synthesized on a Synergy Personal Peptide Synthesizer (Applied Biosystems), which performs solid-phase synthesis using 9-fluorenylmethoxycarbonyl (Fmoc) protection with free termini. Cleavage from the resin was performed with trifluoracetic acid/phenol/thioanisol/ethanediol/water mixture and the peptide purified by HPLC and characterized by mass spectrometry using a laser desorption mass spectrometer (Kratos Analytical Instruments Kompact MALDI 2).

The purified peptides were lyophilized and stored at −20°C. The enzyme inhibition assay for β-glucuronidase was performed essentially as in ref. 25. The inhibition of the binding of the phage displayed peptide by the free peptide was analyzed by phage-ELISA as described above, adding different concentrations (between 100 to 0.01 μM) of free peptide to the phage supernatants. The activity of the alkaline phosphatase was measured using *p*-nitrophenyl phosphate (PNPP, Sigma) as a substrate.

RESULTS

Exon Shuffling. The repertoire of 25 residue "spliced" peptides displayed on phage was created by *in vivo* recombination between phage and plasmid replicons, followed by self-splicing of the RNA between the two exons (Figs. 1 and 2*b*). About 40% of the phage was recombined (as shown by PCR screening); as there are multiple copies of plasmid and phage within the bacterium when the Cre recombinase is active and as phage carry 3–5 copies of pIII (27), we would expect most phage particles to display both exon 2–pIII and exon 1–exon 2–pIII fusions. (However, phage displaying a single species of pIII fusion should segregate on reinfection of fresh bacteria after the first round of selection.) The fact that the recombinant phage was infective (5×10^{10} t.u. per ml of culture) indicates that the splicing has been successful, as stop codons in all three reading frames of the intron would prevent translation of the phage pIII from unspliced RNA.

The number of bacteria containing both replicons and the plasmid encoding Cre recombinase was estimated immediately after infection from the resistance to the three antibiotics carbenicillin, chloramphenicol, and tetracycline (Table 2, sample point 6) as about 1.6×10^{11}.

Selections. The phage repertoire was selected against the antibody PAb240 and the enzymes bovine intestinal alkaline phosphatase and *E. coli* β-glucuronidase. The titer of eluted phage (t.u. total) increased substantially in later rounds of selection: after rounds 1, 2, 3, and 4, respectively, the titers were 1×10^5, 5×10^5, 1×10^7, and 2×10^8 for PAb240; 6×10^5, 8×10^4, 9×10^5, and 5×10^7 for bovine alkaline phosphatase; 3×10^5, 5×10^3, 1×10^5, and 1×10^7 for β-glucuronidase. For each selection, single phage clones were assayed for binding activities by ELISA after four rounds.

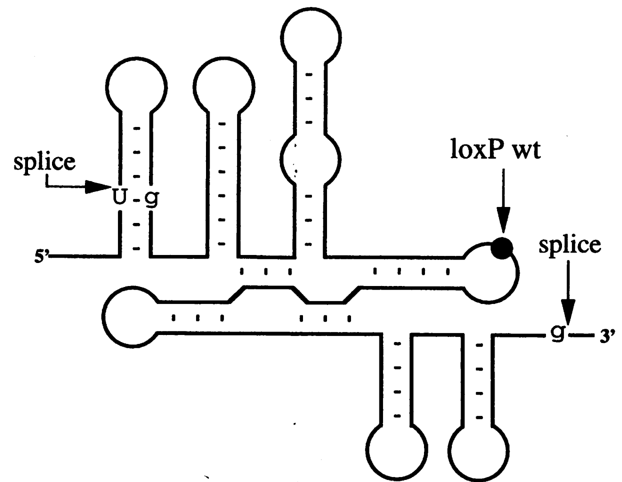

FIG. 1. Typical secondary structure of group I introns. The splice sites are indicated with arrows.

Against PAb240 (IgG_1, κ), which recognizes a linear epitope (RHSVV) of p53 (28), 21 phage clones with binding activities were sequenced and shown to encode different peptide sequences. All the peptide sequences included the consensus motif (R/K)HS(V/I/L) (Table 3). The majority of the selected peptides were full-length, but several encoded exon 1-fusions in which exon 1 and pIII are linked through the intron. The exon-1 fusions appear to have arisen due to loss of exon 2 from the fdDOG-exon 2 repertoire (probably during the assembly of this repertoire).

Against alkaline phosphatase, nine different peptide sequences were identified from 29 clones with binding activities; one sequence (AP8, Table 3) dominated (19 clones) and seven sequences (AP1-6 and AP9, Table 3) proved identical in the first exon but differed in the second. Against β-glucuronidase, four different peptide sequences were identified from 13 clones with binding activities. All the sequences contained at least two aromatic residues (phenylalanine, tyrosine, or tryptophan) in the first exon. In addition, the motif Asp-Pro (DP) was found at the N terminus of three of the peptides (Gluc2-4, Table 3).

The binding of the phage selected against the antibody or enzymes was highly specific. For example, the phage selected against antibody PAb240 did not bind to two other antibodies Fogl (IgG_1, κ) and Fog B (IgG_1, λ) (data not shown). The phage selected against the two enzymes did not bind to the enzymes β-lactamase, glucose oxidase, β-galactosidase, or alkaline phosphatase (Fig. 3).

Role of Each Exon. In all the peptides selected against PAb240, the consensus motif was found in either exon 1 or exon 2. With the peptides selected against the enzymes, common sequence elements appeared to be confined to exon 1 (see above). To determine the role of each exon in AP4 and Gluc4, phages were constructed to display only the peptides encoded by the first or the second exon. Phages displaying only the first exon bound well to the enzyme; although the ELISA signal was weaker for AP4 exon 1 than for AP4 (Fig. 4). Phages displaying only exon 2 did not bind to the enzymes. Therefore, exon 1 appears to contribute most of the binding activity for these two peptides.

This was confirmed by the binding of synthetic peptides (Fig. 5). The peptide pepAP4 did not show any inhibition of the alkaline phosphatase activity (using PNPP as a substrate), but did inhibit the binding of the corresponding phage with an IC_{50} of 1.1 μM (pepAP4, Fig. 5*a*). A synthetic peptide correspond-

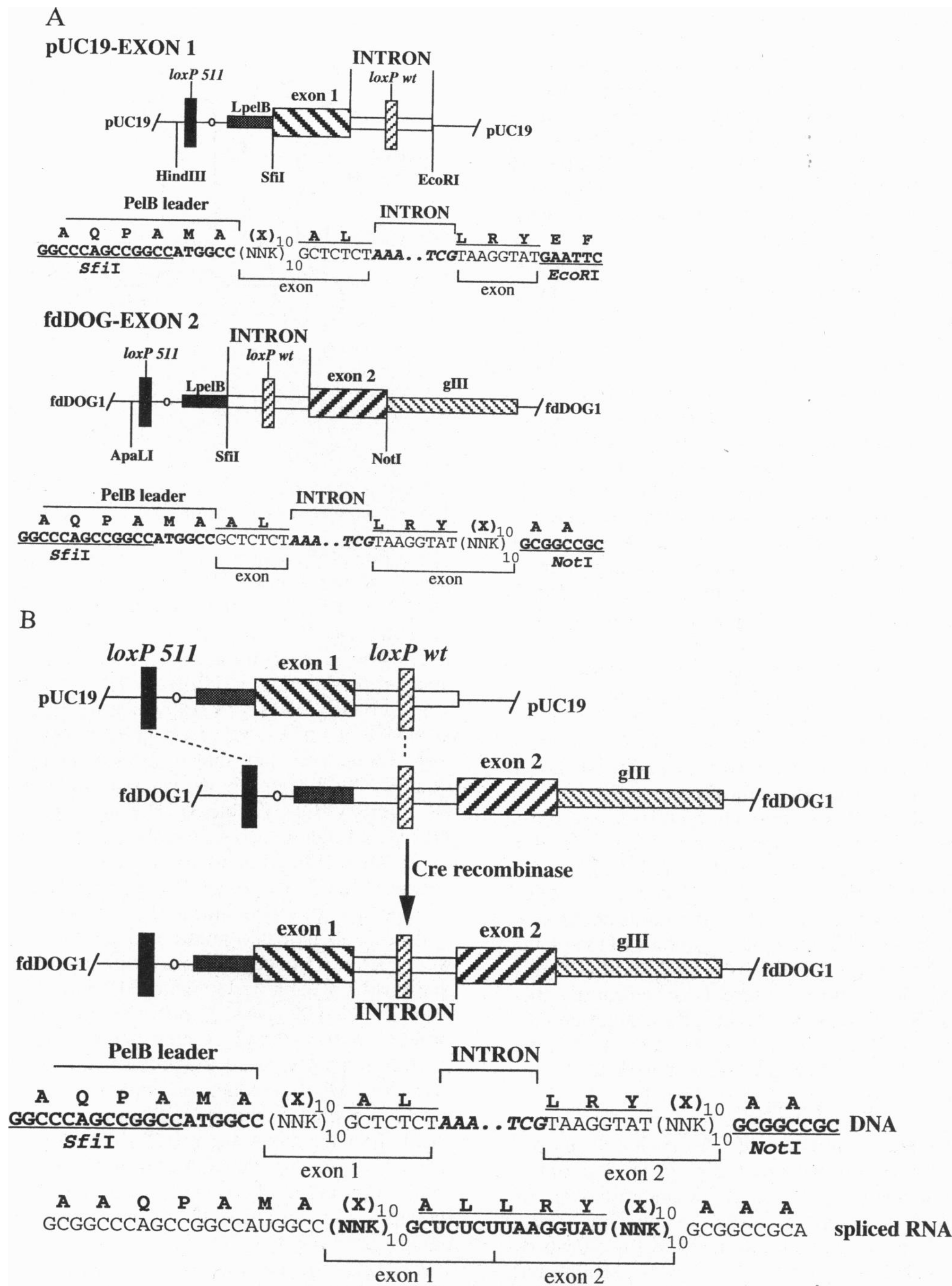

FIG. 2. (*A*) Vectors, pUC19-exon 1 "donor" vector, and fdDOG-exon 2 "acceptor" vector. *loxP* wt, wild-type *loxP* site (14); *loxP* 511, a mutant *loxP* site with a single point mutation (26); LPelB, PelB leader sequence; gIII, fd phage gene III. (*B*) Recombination scheme.

ing to the sequence of the first exon inhibited the binding of the phage with an IC_{50} of 2.6 μM (pepAP4/N, Fig. 5*a*). The peptide pepGluc4 inhibited the activity of β-glucuronidase with a K_i of 17 nM (pepGluc4, Fig. 5*b*) and the sequence corresponding to the first exon had a similar activity (K_i = 23 nM, pepGluc4/N, Fig. 5*b*).

Exon Reshuffling. We made further variations of the peptides by keeping the first (and dominant) exon and reshuffling the second exon. Thus, the first exon was cloned into the pUC donor vector and recombined *in vivo* (as above) with the repertoire of second exons on phage. This yielded 6.1×10^8 infected colonies for AP4 and 5.5×10^9 colonies for Gluc4 (Table 2, sample points 7 and 8), with recombination efficiencies of 29% and 75%, respectively. In these cases, the size of the reshuffled repertoires is therefore limited by that of the acceptor repertoire (6.5×10^5, see *Materials and Methods*).

The reshuffled repertoires were selected against the enzymes as above. After four rounds, phage clones were identified with improved binding activities by ELISA and were sequenced (APC4sh and GlucB3sh, Table 3). After the reshuffling of the second exon, it was still not possible to detect the binding of phage bearing only the second exon, although the ELISA signal of the phage displaying the entire sequence was improved by at least twofold (Fig. 4). The synthetic peptide corresponding to the reshuffled Gluc4 (GlucB3sh, Table 3) was synthesized and shown to inhibit the activity of β-glucuronidase with a K_i of 7.4 nM (Fig. 5*b*).

DISCUSSION

We have developed a method for exon shuffling involving the introduction of a *loxP* recombination site within a group I

Table 3. Sequences of clones

Clone no.	Sequence	f	rnd
Pab240			
A3	EF**KHSV**VGCE*ALLRY*NTPPDPELIC	1	4
A4	CP**RHSI**VEAA*ALLRY*	1	4
A8	S**RHSV**LEP*ALLRY*SEIQAQSFHG	1	4
A11	CP**RHSI**VETA*ALLRY*	1	4
B5	G**RHSV**LGPSM*ALLRY*	1	4
B8	AEL**RHSV**MLG*ALLRY*GAIEPRGKSH	1	4
C4	S**RHSV**LAPALLRYEEIIAYSGSS	1	4
C9	AASTSPGIGA*ALLRY*YQIK**KHSL**IP	1	4
C10	S**RHSV**LAP*ALLRY*	1	4
D1	VF**RHSL**VWSQ*ALLRY*	1	4
D7	**RHSM**VSVDVR*ALLRY*	1	4
D8	T**RHSI**LRPVN*ALLRY*CIETNNDC	1	4
E1	SSLSDASMVD*ALLRY*YQIK**KHSL**IP	1	4
E6	VHADLHDNTK*ALLRY*YQIK**KHSL**IP	1	4
E8	EF**KHSV**VGCE*ALLRY*HVTHGLT	1	4
E10	CIM**RHSV**VPD*ALLRY*APEDQRQICH	1	4
F2	VF**RHSL**VWSQ*ALLRY*APEDQRQICH	1	4
F5	G**RHSV**LGPSM*ALLRY*	1	4
G5	S**RHSV**LEP*ALLRY*SIIEAHSGTC	1	4
F10	EF**KHSV**VGCE*ALLRY*TRSALTSDPC	1	4
H12	EF**KHSV**VGCE*ALLRY*GGIAPHRCTC	1	4
AP			
AP1	SISFGQLWRP*ALLR*LLHKDNAFVR	1	3
AP2	SISFGQLWRP*ALLRY*ESHAGSPRGR	1	3
AP3	SISFGQLWRP*ALLRY*TEASNIIRRT	2	3
AP4	SISFGQLWRP*ALLRY*TDAVNTSLRI	1	3
AP5	SISFGQLWRP*ALLRY*DVKILHGSQR	2	4
AP6	SISFGQLWRP*ALLRY*DHSTTARINL	1	3
AP7	EVRWFDWIHK*ALLRY*PTVLINVRNP	1	3
AP8	DLMGLRNSVL*ALLRY*KLPKPTPGPN	19	3-4
AP9	SISFGQLWRP*ALLRY*EISPASVRLR	1	3
AP4/N	SISFGQLWRP*ALLRY*		
AP4/C	*ALLRY*TDAVNTSLRI		
APC4sh	SISFGQLWRP*ALLRY*GSTPVTLAIS	4	4
APC4sh/C	*ALLRY*GSTPVTLAIS		
b-Gluc.			
Gluc1	WYEYGWDETV*ALLRY*IEGTSISNAA	3	3
Gluc2	DPVTDEWVWE*ALLRY*QQATARILLS	3	3-4
Gluc3	DPLSAFGWNA*ALLRY*KHDIQTVYAQ	4	4
Gluc4	DPVFYVDVLP*ALLRY*QKIFANNI	3	4
Gluc4/N	DPVFYVDVLP*ALLRY*		
Gluc4/C	*ALLRY*QKIFANNI		
GlucB3sh	DPVFYVDVLP*ALLRY*TGSTIPTTIR	2	4
GlucB3sh/C	*ALLRY*TGSTIPTTIR		

f, Number of times the sequence was observed; rnd, round of selection. Sequences of clones selected against the anti-p53 antibody Pab240 with the consensus sequence of Pab240 highlighted, bovine alkaline phosphatase, and *E. coli* β-glucuronidase.

self-splicing intron from *T. thermophila* 26S rRNA. As already described for antibodies, the use of *lox*-Cre recombination allows the construction of very large repertoires by a process of "combinatorial infection" (11): the larger and more diverse a repertoire, the more likely it is to provide ligands with high binding affinity (20). Thus, in earlier work (20), antibody heavy and light chain gene repertoires on plasmid and phage replicons, respectively, were locked together (using *lox*-Cre recombination) on the same replicon within infected bacteria to create large ($>10^{10}$) and highly diverse combinatorial repertoires. This process overcomes the limits to library size (about 10^8-10^9 clones) imposed by the transfection efficiency of phage DNA into bacteria (17). Following this process, we were able to generate a very large repertoire (1.6×10^{11}) of polypeptides fused to the protein pIII of the filamentous phage fd.

Although the process of combinatorial infection is highly suitable for making combinatorial repertoires of two polypeptide chains (such as with antibodies), it is more difficult to apply to single polypeptide chains, as the *loxP* site (34 nucleotides encoding at least 12 amino acids) has to be located within the polypeptide reading frame. However, by placing the *loxP* site within a self-splicing intron it can be removed from the RNA. Nevertheless, the join is not seamless; we retained nucleotides from *Tetrahymena* exons adjacent to the splice sites as they are known to make base pair interactions with sequences within the intron (29). Seven nucleotides in the first exon and eight nucleotides in the second exon were retained and together encode five amino acid residues (ALLRY) located in the middle of the displayed peptide. The spliced peptide repertoires therefore comprise two hypervariable sequences with an intervening five-residue spacer. However, as mutations in the splice sites can be suppressed by intron mutations that restore base pairing (30), it seems likely that spacers with a wide range of different sequences and structures can be devised.

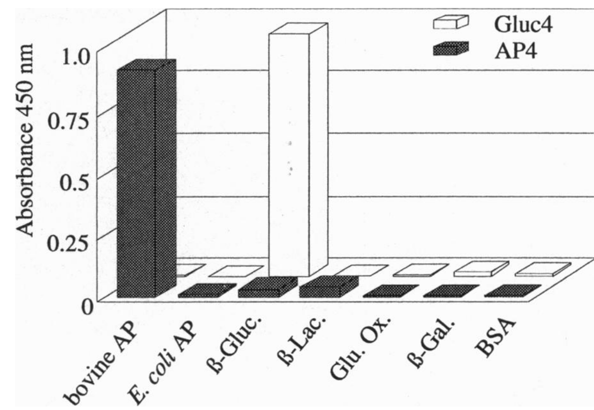

FIG. 3. Specificities of monoclonal phage selected against bovine alkaline phosphatase (AP4) and *E. coli* β-glucuronidase (Gluc4) measured by ELISA. AP, alkaline phosphatase; β-Gluc., *E. coli* β-glucuronidase; β-Lac, *E. coli* β-lactamase; Glu. Ox., *Aspergillus niger* glucose oxidase; β-Gal., *E. coli* β-galactosidase; BSA, bovine serum albumin.

From the spliced repertoire, we isolated peptides with specific binding activities against an antibody (PAb240) and two enzymes (alkaline phosphatase and β-glucuronidase). The peptides selected against PAb240 included the consensus motif

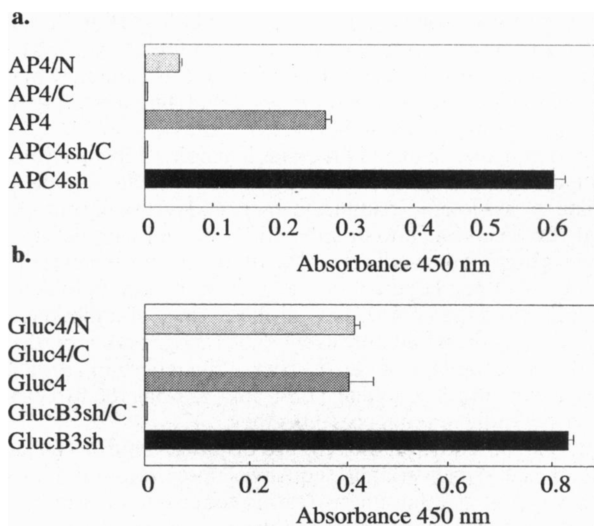

FIG. 4. ELISA of full-length peptide and peptide fragments displayed on phage. (*a*) Anti-alkaline phosphatase peptides (AP4/N, AP4 exon 1; AP4/C, AP4 exon 2; APC4sh/C, C4 exon 1). (*b*) Anti-β-glucuronidase peptides (Gluc4/N, Gluc4 exon 1; Gluc4/C, Gluc4 exon 2; GlucB3sh/C, B3 exon 1). The quantity of 2×10^9 t.u. of phage per well was used. See Table 3 for the peptide sequences.

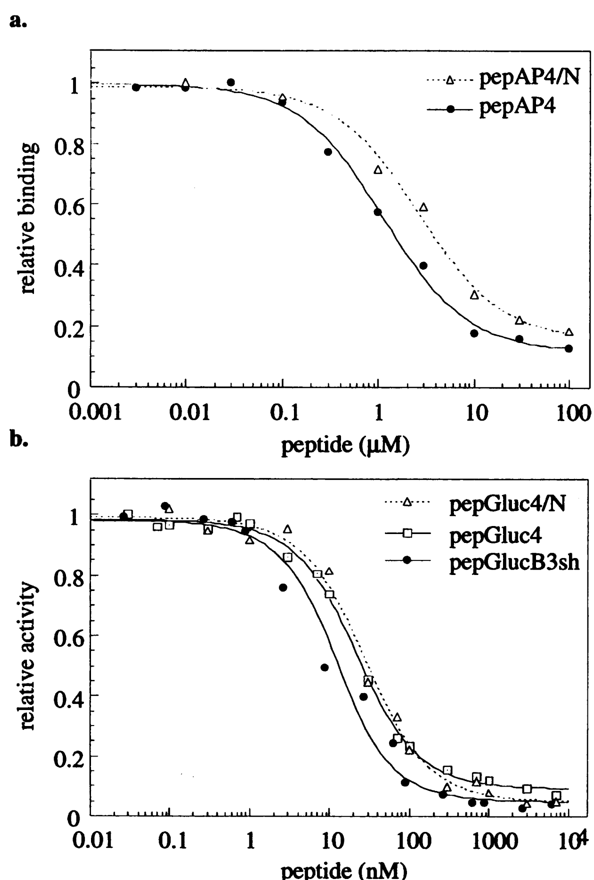

FIG. 5. Inhibition measurements. (*a*) Inhibition of the clone AP4-phage binding to alkaline phosphatase by varying concentrations of synthetic peptides pepAP4 and pepAP4/N (pepAP4 exon 1). (*b*) Inhibition of the β-glucuronidase activity by the synthetic peptides pepGluc4, pepGluc4/N (pepGluc4 exon 1), and pepGlucB3sh. See Table 3 for the sequences.

in either exon 1 or exon 2, indicating that either exon can contribute to binding. However, with the peptides AP4 and Gluc4 selected against the enzymes, binding was mediated mainly through exon 1 (Fig. 4). Although AP4 exon 2 does appear to make some contribution to binding (Fig. 4*a*), for pepGluc4 this was insignificant (K_i for pepGluc4 = 17 nM; K_i for pepGluc4 exon 1 = 23 nM). However, this exon may make some binding contacts as after exon reshuffling a peptide was isolated with improved activity (K_i = 7 nM).

The dominance of one of the exons in binding might have been anticipated. Peptide repertoires appear to be highly suitable for binding to cavities, for example, many peptides (usually hexapeptides) have been found to fit into antigen binding sites but not to other portions of the antibody (31). Although our sequence of 25 residues is rather larger, it is likely to be mainly unfolded in solution, and large compared with the cavities of antibodies or enzymes. It is therefore only likely to be recognized in portions. To take full advantage of the diversity of the repertoire, it will be necessary to introduce folding constraints to bring the two exons (and hypervariable sequences) together.

Thus, it should be possible to use disulfide bonds to fashion each of the hypervariable sequences into loops and more directly to link together the two exons, for example, by bridging from the N terminus of exon 1 to the C terminus of exon 2 to form a cyclic peptide. Alternatively, the sequence of the spacer region might even be developed as a nucleus for folding of the hypervariable sequences, as in the hydrophobic core of barstar (32). The spacer also may offer advantages for the chemical synthesis of peptides or domains. For example, the spacer might be synthesized as a central building block onto which other segments could be ligated (33, 34).

We anticipate that the shuffling of peptide exons and the introduction of folding constraints will allow the construction of folded peptide ligands *de novo*. It will be interesting to see whether such nuggets of structure can be assembled into complete protein domains. Nevertheless, the use of introns containing recombination sites may provide alternative strategies for building domains, for example by shuffling blocks of sequence within preexisting domains. It should certainly be useful for making combinatorial repertoires of domains in multidomain proteins. We propose that the use of exon shuffling may offer a means of recapitulating and testing possible strategies of protein evolution.

We thank D. Lane and C. Stephen for the gift of the PAb240 monoclonal antibody, W. Ouwehand for the gift of the monoclonal antibodies Fogl and FogB, and I. C. Eperon for the gift of the vector M13 mICE10 containing the intron.

1. Stemmer, W. P. C. (1995) *Bio/Technology* **13,** 549–553.
2. Gilbert, W. (1978) *Nature (London)* **271,** 501.
3. Patthy, L. (1994) *Curr. Opin. Struct. Biol.* **4,** 383–392.
4. Larigan, J. D., Tsang, T. C., Rumberger, J. M. & Burns, D. K. (1992) *DNA Cell Biol.* **11,** 149–162.
5. Ikeo, K., Takahashi, K. & Gojobori, T. (1995) *J. Mol. Evol.* **40,** 331–336.
6. Rogers, J. (1985) *Nature (London)* **315,** 458–459.
7. Donate, L. E., Gherardi, E., Srinivasan, N., Sowdhamini, R., Aparicio, S. & Blundell, T. L. (1994) *Protein Sci.* **3,** 2378–2394.
8. Doolittle, R. F. (1995) *Annu. Rev. Biochem.* **64,** 287–314.
9. Waring, B. W., Ray, J. A., Edwards, S. W., Scazzocchio, C. & Davies, R. W. (1985) *Cell* **40,** 371–380.
10. Price, J. V. & Cech, T. R. (1985) *Science* **228,** 719–722.
11. Waterhouse, P., Griffiths, A. D., Johnson, K. S. & Winter, G. (1993) *Nucleic Acids Res.* **21,** 2265–2266.
12. Gibson, T. J. (1984) *Studies on the Epstein-Barr Virus Genome* (University of Cambridge, Cambridge, U.K.).
13. Wertman, K. F., Wyman, A. R. & Botstein, D. (1986) *Gene* **49,** 253–262.
14. Hoess, R. H., Ziese, M. & Sternberg, N. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 3398–3402.
15. Cech, T. R. (1990) *Annu. Rev. Biochem.* **59,** 543–568.
16. Deeney, C. M., Eperon, I. C. & Potter, B. V. (1987) *Nucleic Acids Symp. Ser.* **18,** 277–280.
17. Dower, W. J., Miller, J. F. & Ragsdale, C. W. (1988) *Nucleic Acids Res.* **16,** 6127–6145.
18. Griffiths, G. M., Berek, C., Kaartinen, M. & Milstein, C. (1984) *Nature (London)* **312,** 271–275.
19. Hoogenboom, H. R., Griffiths, A. D., Johnson, K. S., Chiswell, D. J., Hudson, P. & Winter, G. (1991) *Nucleic Acids Res.* **19,** 4133–4137.
20. Griffiths, A. D., Williams, S. C., Hartley, O., Tomlinson, I. M., Waterhouse, P., Crosby, W. L., Kontermann, P., Jones, P. T., Low, N., Allison, T. J., Prospero, T., Hoogenboom, H. R., Nissim, A., Cox, J. P. L., Harrison, J. L., Zaccolo, M., Gherardi, E. & Winter, G. (1994) *EMBO J.* **13,** 3245–3260.
21. Cagnon, C., Valverde, V. & Masson, J. M. (1991) *Protein Engin.* **4,** 843–847.
22. McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. (1990) *Nature (London)* **348,** 552–554.
23. Güssow, D. & Clackson, T. (1989) *Nucleic Acids Res.* **17,** 4000.
24. Clackson, T., Hoogenboom, H. R., Griffiths, A. D. & Winter, G. (1991) *Nature (London)* **352,** 624–628.
25. Jefferson, R. A., Kavanagh, T. A. & Bevan, M. W. (1987) *EMBO J.* **6,** 3901–3907.
26. Hoess, R. H., Wierzbicki, A. & Abremski, K. (1986) *Nucleic Acids Res.* **14,** 2287–2300.
27. Glaser-Wuttke, G., Keppner, J. & Rasched, I. (1989) *Biochim. Biophys. Acta* **985,** 239–247.
28. Stephen, C. & Lane, D. (1992) *J. Mol. Biol.* **225,** 577–583.
29. Cech, T. R., Damberger, S. H. & Gutell, R. R. (1984) *Nature (London) Struct. Biol.* **1,** 273–280.
30. Been, M. & Cech, T. R. (1986) *Cell* **47,** 207–216.
31. Lane, D. P. & Stephen, C. W. (1993) *Curr. Opin. Immunol.* **5,** 268–271.
32. Buckle, A. M., Schreiber, G. & Fersht, A. R. (1994) *Biochemistry* **33,** 8878–8889.
33. Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. H. (1994) *Science* **266,** 776–779.
34. Schnolzer, M. & Kent, S. B. H. (1992) *Science* **256,** 221–225.