Anal Bioanal Chem (2003) 376 : 1006–1013
DOI 10.1007/s00216-003-1876-3

SPECIAL ISSUE PAPER

Krzysztof Bialek · Andrzej Swistowski · Ronald Frank

Epitope-targeted proteome analysis: towards a large-scale automated protein–protein-interaction mapping utilizing synthetic peptide arrays

Received: 8 November 2002 / Revised: 17 February 2003 / Accepted: 20 February 2003 / Published online: 3 April 2003

Abstract We describe the development of a process for the genome-wide mapping of interactions between protein domains and peptide ligands entirely based on high-throughput biochip technologies. A phage library displaying protein domains from a randomly fragmented and cloned cDNA library will be "panned" on an array of synthetic peptide ligands. After multiplexed affinity enrichment, peptide-specific phage populations will be automatically eluted, propagated, labelled and identified by hybridisation to a DNA microarray. Peptide arrays are synthesized in situ by SPOT synthesis on a planar substrate. By utilizing a commercially available library of human brain cDNA plus a set of distinct model domains cloned into T7-phage, we could show that a single panning round on an array of known peptide ligands for these model domains synthesized on a cellulose membrane can yield an enrichment of better than a factor of 1,000. This is sufficient to detect peptide-specific enrichment of Cy3(post-panning) against Cy5(pre-panning)-labelled phage DNA inserts on a cDNA microarray. Thus, the proof-of-principle of our approach could be successfully demonstrated and first interaction data are being collected.

Keywords Peptide arrays · DNA microarrays · Protein domains · Epitope · Phage display · Automation · Proteomics · Functional genomics

Introduction

Function is tantamount to interaction. Thus, global functional genome and proteome research aims to provide a complete description of the network of protein interactions within a cell or organ(ism) that is diagnostic for a specific cellular state such as foetal or adult, brain or liver, healthy or oncogenic/pathogenic etc. Proteins interact via surface accessible interaction sites which involve amino acid side chain and backbone contacts along (but not necessarily contiguously) a linear segment of the protein chain (linear epitopes) or involve amino acid residues from two or more segments of the protein chain brought together by the folded conformation (conformational epitopes). Note that the term epitope is used here in its broadest sense and far beyond the mere immunological meaning. Linear epitopes can be effectively represented by small peptide fragments that are readily available through simultaneous and parallel chemical synthesis. This is also partly true for conformational interaction sites that can specifically recognize mimotope peptides [1]. Many proteins, most prominently those of regulatory function, are built from smaller domains which are stably folded structural modules still displaying their specific functional property. The catalogue of such domains that recognize linear epitopes is rapidly growing (kringel, SH2, SH3, PH, EVH1, PDZ, WW, etc. [2]) indicating a more general principle utilized by nature. These domains are found to be involved in diverse molecular organization and regulation phenomena.

Complementary to other biochemical approaches such as large-scale analysis of protein complexes [3, 4] and molecular biology approaches such as the yeast-two-hybrid method [5, 6], a genome-wide peptide screening approach will directly address functional protein interaction sites, leading to a detailed insight into the discovered molecular recognition events, placing them in the context of the whole genome and even allows one to rapidly decipher the chemical nature of these interactions [7]. This information can then be transferred into powerful small-peptide tools that interfere with these interactions in vivo and help to link targets with phenotypes [8].

One important aim of functional genomics/proteomics is to gain access to new targets for drug discovery. Thus, it is logical to set up a genome-wide search for all "drugable" proteins and then validate these as relevant pharmaceutical targets by modern proteome analysis. It can be concluded that these "drugable" targets primarily belong to that repertoire of proteins that can bind small molecule

K. Bialek · A. Swistowski · R. Frank (✉)
Research Group for Molecular Recognition,
GBF (German Research Centre for Biotechnology),
Mascheroder Weg 1, 38124 Braunschweig, Germany
e-mail: frank@gbf.de

ligands. Synthetic peptides are practical tools readily at hand to address this property. Although peptides themselves have lost attractiveness as pharmaceutical drugs, they are perfect molecular probes for the search for new pharmaceutical targets.

SPOT synthesis [9] is a well-established technique for a rapid and flexible generation of large repertoires of peptide fragments in a high-density array format. SPOT peptide arrays are particularly suited for in situ biological activity screening in binding or enzymatic transformation assays. The above considerations consequently led us to exploit this synthetic peptide array format for the development of an automated system to allow a genome-wide systematic mapping of protein–domain to peptide–ligand interactions, which is entirely based on high-throughput parallel microarray technology.

Experimental

Materials

All media components, gelatine and skimmed milk were purchased from BD Difco (Heidelberg, Germany). Heparin (H-5284), SDS (L-4522), BSA (A-3059), EDTA (E 4884), Su buffer (Sigma membrane-blocking buffer SU-07–250) were purchased from Sigma (Deisenhofen, Germany). MB grade fish sperm DNA was obtained from Roche (Mannheim, Germany). Formamide was purchased from Roth (Karlsruhe, Germany). Oligonucleotides were synthesised by MWG-Biotech AG (Ebersberg, Germany) and Invitrogen (Karlsruhe, Germany).

E. coli carrying plasmid P18 with the coding sequence of the EVH1 domain of mouse MENA (E) was a kind gift of Melanie Barzik (Dept. of Structural Biology, GBF). *E. coli* carrying plasmids with the coding sequences of the WW domains of rat FE 65 (F) and of the human YAP (Y) proteins were a kind gift of Dr Marius Sudol from Mount Sinai School of Medicine, New York, USA. The mouse cDNA was a kind gift of Jadwiga Jablonska (Dept of Cell Biology, GBF). T7Select 10-3b phage display system, human normal brain T7Select 10-3b cDNA library, *E. coli* BL 5403 and BL 5615 strains were purchased from Novagen (Madison, USA). The λDisplay1 vector was a kind gift of Prof Gianni Cesareni from the University of Rome, Italy. The host strain *E. coli* BB4 was purchased from Pharmacia Biotech (Piscataway, USA).

Construction of T7 phages presenting the model protein domains

Plasmid DNA was isolated from bacterial lysates using Qiagen Miniprep Kit (Qiagen, Hilden, Germany). With primers extended by a 5′ EcoRI or a 3′ HindIII restriction site (underlined), corresponding to the 5′ and 3′ ends of the E domain (5′-dCGTGAATTCCATGGCTGAACAGAGT-3′ and 5′-GGTAAGCTTCGACGTAGATCCTGTCAAT-3′), the F domain (5′-TCTGAATTCCGATCTACCGGCTGGA-3′ and 5′-ATGAAGCTTCCCCTGTGATGGGGAG-3′), and the Y domain (5′-GGCGAATTCTTCTTTTGAGATACCTGATGA-3′ and 5′-AATAAGCTTCGACTGGTGGGGGCTG) respectively, all domains were amplified by PCR. The 100 µL PCR contained: 2.5 units of Taq DNA polymerase, 10 mM Tris–HCl (pH 9.0), 50 mM KCl, 1.5 mM $MgCl_2$, 0.1% Triton X-100, 0.5 µM of each primer pair, 0.2 mM of each dNTP (Promega, Madison, USA) and 1 µg plasmid DNA. Conditions used for amplification were as follows: hot start with 5 min pre-incubation at 95 °C, 30 cycles at 95 °C for 1 min, 58.5 °C (E and Y) and 61 °C (F) for 30 s, 72 °C for 1 min, followed by 6 min incubation at 72 °C. 2 µg (E) and 1 µg (F and Y) PCR products were digested with EcoR I and Hind III in EcoR I buffer (New England Biolabs, Frankfurt am Main, Germany) overnight at 37 °C with 20 units of enzyme each. The reaction products were purified with the Qiagen PCR Purification Kit (Qiagen, Hilden, Germany).

The DNA fragments obtained were then ligated with T7Select 10–3b vector arms. The 5-µL ligation reactions contained: 0.4 Weiss units of T4 DNA ligase, 20 mM Tris–HCl (pH 7.6), 10 mM $MgCl_2$, 25 µg mL^{-1} acetylated BSA, 10 mM DTT, 1 mM ATP 0.02 pmol of T7Select 10–3b vector arms (Novagen Madison, USA) and about 0.02 pmol of DNA of each domain inserts. Ligation was carried out for 3 h at 16 °C. For in vitro packaging into phages particles, 5 µL of ligation reaction was added to 25 µL of T7Select Packaging Extracts and incubated for 2 h at room temperature. The reaction was stopped by addition of 270 µL of sterile LB medium [10]. Infection, propagation, titering and preparation of phages lysates were made according to the manufacturers instructions (http://www.novagen.com/docs/ndis/C166–000.pdf).

The coding DNA sequences of the mouse Il-4, fas antigen and ezrin proteins were amplified from mouse cDNA as a template by PCR with the corresponding primers: 5′-TTAGAATTCGATGGGTCTCAACCC-3′ and 5′-CTAAAGCTTGGTGGCTCAGTACTACG-3′, 5′-CAAGAATTCAGACATGCTGTGGATC-3′ and 5′-CCTAAGCTTCACTCCAGACATTGTC-3′, 5′-TAAGAATTCAGCCAAGATGCCCAAG-3′ and 5′-ATCAAGCTTCTACATGGCCTCGAA-3′, respectively. PCR condition, digestion, ligation and packaging were as describe above (the annealing temperature was 58.5 °C for Il-4 and fas antigen and 61 °C for ezrin).

Construction of λ phages presenting the model protein domains

λDisplay1 phage was diluted to 10^4 pfu mL^{-1} concentration in SM buffer (50 mM Tris–HCl pH 7.5, 100 mM NaCl, 8 mM $MgSO_4$, 0.01% gelatine). 400 µL of diluted phages was mixed with 400 µL of *E. coli* BB4 plating bacteria (OD_{600}=2.0 in 10 mM $MgSO_4$) and incubated 20 min at 37 °C. Then 12 mL of LB top agarose was added, the mixture was poured onto 145/20-mm-large Petri dishes and incubated at 37 °C overnight. Phages were eluted from the dish by filling with 20 mL of SM buffer and incubation for 5 h at 4 °C under gentle agitation on a rocker table. The liquid was decanted, supplemented with chloroform (2% final concentration) and centrifuged at 4 °C with a speed of 4,500 g. The phage lysate was separated from the pellet and placed into a new tube [10]. The λDisplay1 phage vector DNA was isolated from this lysate by using the Nucleobond AX100 kit (Machery-Nagel, Düren, Germany). 20 µg of vector DNA was digested with SpeI and NotI in NEB 2 buffer (New England Biolabs, Frankfurt am Main, Germany) in a total volume of 50 µL at 37 °C overnight with 25 units of each enzyme and then isopropanol-concentrated to 20 µL (0.8 µg $µL^{-1}$). The coding sequences of the three model domains were transferred from the T7Select 10–3b constructs above to the λDisplay1. For this purpose the following primers were chosen: the sense primer 5′-TCAACTAGTATGCTCGGGGATCCGAA-3′ introduced the SpeI restriction site (underlined) at the 5′ end of amplified DNA sequences and the anti-sense primer 5′-CGTTACCTAGTTACTCGAGTGCGGC-3′. The NotI restriction site was taken from the T7Select 10–3b vector. The 100-µL PCR reactions were carried out as described above (construction of T7 phages) with an annealing temperature of 58.5 °C. The PCR products were SpeI/NotI-digested for 4 h at 37 °C with 5 units of each enzyme and purified using a Qiagen PCR Purification Kit (Qiagen, Hilden, Germany). These DNA fragments were ligated with previously digested λDisplay1 vector arms. The 15-µL ligation reactions contained: 6 Weiss units of T4 DNA ligase, 20 mM Tris–HCl (pH 7.6), 10 mM $MgCl_2$, 25 µg mL^{-1} acetylated BSA, 10 mM DTT, 1 mM ATP (Novagen Madison, USA), 4.2 µg of λDisplay1 Vectors Arms and about 11.6 ng of DNA from each domain inserts. Ligation was carried out for overnight at 16 °C. 8 µL of the ligation mixture was directly in vitro packaged by using lambda-packaging kit (Pharmacia Biotech, Piscataway, USA). Infection, propagation and titering were carried out according to the manufacturer's instructions. Several single phage plaques were transferred to separate 1.5-mL tubes containing. 200 µL of SM buffer and incubated for 2 h at room temperature. 50 µL of each lysate was taken for DNA template preparation and 1 µL of 0.5 M EDTA (pH 8) was added, heated at 65 °C

Table 1 Characteristics of the three model protein domains selected for proof-of-principle studies

Code	Type of domain	Parent protein	Peptide ligand	K_d (μM)	DNA size (bp)[a]	Domain size (aa)	Reference
E	EVH1	mMena	SFEFPPPPTDEELRL (pepE)	5	467	113	[11]
Y	WW	hYAP	GTPPPPYTVG (pepY)	54	278	64	[12]
F	WW	rFE65	PPPPPPPLPAPPPQP (pepF)	n.d.	228	46	[13]

[a]Size of the PCR product with T7 Up and Down primers

for 10 min and centrifuged at 14,000 g for 3 min. The identity of DNA inserts was determined by PCR using domain-specific primers followed by agarose gel electrophoresis (see construction of T7 phages). The preparation of larger stocks of λphage displaying the model domains followed the procedure as described above for λDisplay1.

Analysis of phage clones by DNA insert sizes

A portion of the top agarose of an individual phage plaque was dispersed into 100 μL of 10 mM EDTA (pH 8.0), heated at 65 °C for 10 min, centrifuged at 14,000 g for 3 min and 0.5 μL of this phage lysate was taken to perform PCR amplification of the DNA insert with T7-Up primer 5′-GGAGCTGTCGTATTCCAGTC-3′ and T7-Down primer 5′-AACCCCTCAAGACCCGTTTA-3′. The 20-μL PCR reaction mixture contained 1 unit of Taq DNA polymerase, 10 mM Tris–HCl (pH 9.0), 50 mM KCl, 1.5 mM $MgCl_2$, 0.1% Triton X-100, 0.1 μM of each primer and 0.2 mM of each dNTP (Promega, Madison, USA). Conditions used for amplification were as follows: hot start with 3 min pre-incubation at 80 °C, 35 cycles at 94 °C for 50 s, 50 °C for 1 min, 72 °C for 1 min, followed by final extension at 72 °C for 6 min. The λphage inserts were amplified using λ Up and Down primers: 5′-CGGCAATCAGCATCGTTACT-3′ and 5′-CGAATTCCTTAGCGGCCG-3′ respectively, under the same conditions as above. Sizes of inserts were determined by electrophoresis of the PCR products in a 2% agarose gel containing 0.5 mg mL ethidium bromide [10].

DNA sequencing

Inserts of the phages were amplified by PCR using T7 Up and Down primers. The reaction products were purified with the Qiagen PCR Purification Kit (Qiagen, Hilden, Germany). The sequencing reaction was performed using ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kits (PE Biosystems, Warrington, England) according to the manufacturer's instructions (http://docs.appliedbiosystems.com/pebiodocs/04339923.pdf). Samples were then sent for analysis to GATC Biotech AG (Konstanz, Germany).

Peptide arrays

The three target peptide sequences of the respective model domains (Table 1) together with a control peptide (NRPPPAVGPQPAP) were chemically synthesised as spots on a small square of 1×1-cm of a modified cellulose membrane (AC01–13; AIMS Scientific Products GmbH, Braunschweig, Germany) following described procedures [14]. Peptides are *C*-terminally anchored onto a stable amide PEG spacer and *N*-terminally acetylated. The 4-spot peptide array was synthesized in many copies on a continuous membrane sheet and separated post synthesis by cutting apart with a pair of scissors. One SPOT of about 6 mm^2 carries roughly 5 nmol of peptide.

Biopanning of phages on the spot membrane

The spots membrane was placed in 2-mL Eppendorf tubes and blocked with 2 mL of blocking solution (Table 2) for 1 h at room temperature under gentle agitation on a rocker table. The model phages were diluted to various degrees as indicated with either phages lacking an insert, or with a mixture of Il-4, fas and ezrin protein-presenting phages, or with the human normal brain T7Select 10–3b cDNA library. 0.9 mL of phage mixture was blocked with 0.1 mL of 5× concentrated blocking solution for 10 min at room temperature. Phages were transferred to the tubes containing the spot membranes and incubated for 2.5 h at room temperature or overnight at 4 °C under gentle agitation. The membranes were washed ten times for 5 min each with blocking solution and ten times for 5 min each with TBST (25 mM Tris–HCl pH 7.4, 150 mM NaCl, 0.1% Tween 20). The individual spots were cut out and transferred to 0.5-mL tubes. T7 phages were eluted by incubation in 150 μL elution buffer (Table 2) for 10 min. 100 μL of eluate was used to make a series of dilutions (from undiluted up to $1:10^5$ depending upon the expected amount of eluted phages) which were plated onto Petri dishes with *E. coli* BL 5403 or BL 5615. λ phages were eluted by direct infection utilizing 150 μL BB4 plating bacteria ($OD_{600\,nm}$=2.0, grown in LB medium containing 0.2% maltose (w/v) and 10 mM $MgSO_4$) and incubation for 20 min at 37 C. 100 μL of eluate was used to make a series of dilutions as above.

Table 2 Experimental conditions evaluated for optimal phage binding and elution using peptide ligands bound on cellulose membrane substrates

Blocking reagent	Binding conditions	Elution reagent
3% BSA in 0.1% TBST	2.5 h at RT	5 M NaCl
Su Buffer in 0.1% TBST	ON at 4 °C	100 mM DTT
2% skimmed milk in 0.1% TBST		1% SDS
5% skimmed milk in 0.1% TBST		0.5% SDS
5% skimmed milk+50 μg mL^{-1} DNA+ 50 μg mL^{-1} heparin in 0.1% TBST		0.1% SDS
5% skimmed milk+50 μg mL^{-1} DNA+ 500 μg mL^{-1} heparin in 0.1% TBST		Direct infection of bacteria (only λ phages)
5% skimmed milk+500 μg mL^{-1} DNA+ 50 μg mL^{-1} heparin in 0.1% TBST		
10% skimmed milk+1.25% BSA in 0.5% TPBS		

Printing DNA arrays on CMT–GAPS glass slides

To prepare DNA of the model domains for printing, PCR with the domain specific primers (see above) was performed. As a hybridisation control we used DNA coding for the variable part of a mouse IgM heavy chain (about 360 bp) amplified by specific primers. About 5 µg of each DNA was separately re-suspended in 20 µL of 3×SSC (0.45 M NaCl, 45 mM sodium citrate, pH 7.0) and printed on the CMT–GAPS-coated slides from Corning (New York, USA) with a GMS 427 ring and pin device (Genetic MicroSystem, Woburn, USA). Spotted DNA was immobilized by UV cross-linking at 254 nm for 5 min using Desaga MinUVIS lamp (Sarstedt, Germany) followed by baking at 80 °C for 3 h. Slides were transferred into a 95 °C water bath for 2 min, next for 1 min into 95% ethanol and finally placed into 50-mL Falcon tubes and dried by centrifugation (500 g for 1 min). Slides were stored dry in the slide container at room temperature.

PCR amplification and labelling of DNA after panning

98 µL of the phage lysate was mixed with 2 µL of 0.5 M EDTA (pH 8.0), heated at 65 °C for 10 min and centrifuged at 14,000 g for 3 min. A 2-µL aliquot of this solution was taken to perform PCR using T7 Up and Down primers. The 100-µL PCR contained 2.5 units of Taq DNA polymerase, 10 mM Tris–HCl (pH 9.0), 50 mM KCl, 1.5 mM $MgCl_2$, Triton X-100, 0.3 µM of each primer, 200 µM of dATP, dTTP, dGTP, 180 µM dCTP (Promega, Madison, USA), 20 µM of Cy3- or Cy5-dCTP (Perkin–Elmer, Boston, USA). Conditions used for amplification were as follows: hot start with 5 min pre-incubation at 80 °C, 35 cycles at 94 °C for 50 s, 50 °C for 1 min, 72 °C for 3 min, followed by final extension at 72 °C for 6 min. PCR products were purified with Qiagen PCR purification Kit (Qiagen, Hilden, Germany). DNA concentration was determined with the Pharmacia GeneQuant II RNA/DNA calculator using 5-mm Hellma quartz Suprasil glass cuvettes. The obtained labelled DNA was immediately used for hybridisation.

Microarray hybridisation

The slides were pre-incubated in 50% formamide, 5×SSC, 0.1% SDS for 30 min to 1 hour at 42 °C, then washed by immersing in water, followed by isopropanol, and dried by centrifugation (500 g for 1 min.). About 4 µg of labelled DNA was dried in a vacuum centrifuge (UNIVAPO 150H; UniEquip, Martinsried, Germany) and re-dissolved in 20 µL of a solution containing 50% formamide, 5×SSC and 0.1% SDS. The sample DNA was heated at 95 °C for 5 min, cooled on ice for 3 min and briefly spun in a microcentrifuge. The sample was then applied onto the slide, carefully covered with a cover slip, sealed with Fixogum Rubber Cement and incubated overnight at 42 °C. The cover slips were then removed by immersing the slides in 2×SSC at room temperature and the slides were washed twice in 0.1×SSC, 0.1% SDS for 5 min each, twice in 0.1×SSC for 5 min each, dried by centrifugation (500 g for 1 min) and evaluated with a GMS 418 scanner (Genetic MicroSystem, Woburn, USA).

DNA-printed slides could be reused up to 5 times. After documentation of the hybridisation image of one experiment, the slides were incubated in stripping buffer (100 mM sodium carbonate/bicarbonate pH 9.8–10, 0.01% SDS) at 90 °C for 2 min followed by washing extensively with water, 70% ethanol, and finally absolute ethanol and then dried by centrifugation (500g for 1 min).

Results and discussion

Experimental concept

The task as defined in the introduction involves the screening of a library of peptide fragments versus a library of protein domains; both libraries should include all relevant components from a biological model system of interest such as a microorganism, eucaryotic cell or tissue. Both libraries, thus, are more or less genome spanning and considerably large. Our experimental concept to cope with this complexity is outlined in Fig. 1. Peptide fragments will be displayed as immobilized compounds in the format of macro- or microarrays. These are manufactured on planar continuous membrane supports by in situ parallel chemical synthesis utilizing the SPOT technique [9]. Such peptide macroarrays are applied as synthesized or are cleaved into separate reservoirs and printed in a higher density onto another planar carrier such as a glass microscope slide. A bacteriophage library displaying protein fragments expressed from a randomly fragmented cDNA library cloned into the phage genome to yield fusions with an envelope protein will be the source for the protein domains. This format of a protein library was shown to be compatible with a protein selection process based on binding to immobilized peptides arrayed on a membrane support [15]. Furthermore, such phage particles are particularly easy to propagate and analyse by means of DNA technology, which is prerequisite to the anticipated systematic analysis of thousands to millions of peptide ligands. Obviously, our approach requires a very competent analytical tool for a rapid, automated and sensitive simultaneous detection of whole populations of phage-displayed enriched protein domains preferably after only one round of binding and elution (panning). These analytical conditions are perfectly met by modern DNA microarray analysis as applied for example to whole-genome expression profiling [16]. The protein domain-presenting phage library will be incubated with the synthetic peptide arrays. This affords a multiplexed affinity enrichment simultaneously on all peptide array elements. Peptide-specific phage populations will be automatically eluted from the array elements, propagated, their DNA labelled with either Cyanine dyes Cy3 or Cy5, mixed with an equal amount of respectively labelled pre-panning phage library DNA and finally analysed by hybridisation to DNA microarrays.

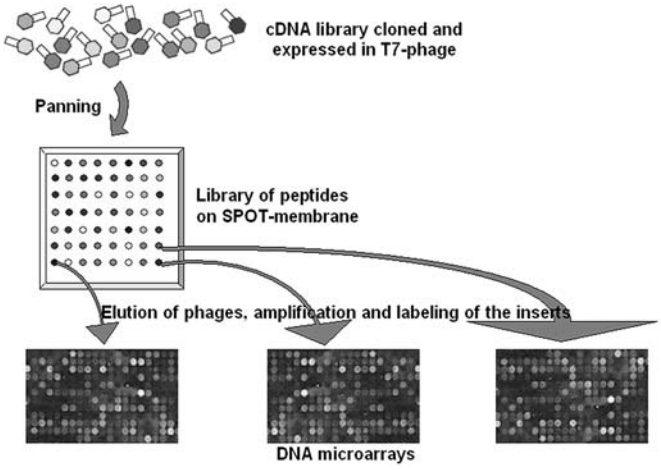

Fig. 1 Schematic illustration of the experimental concept

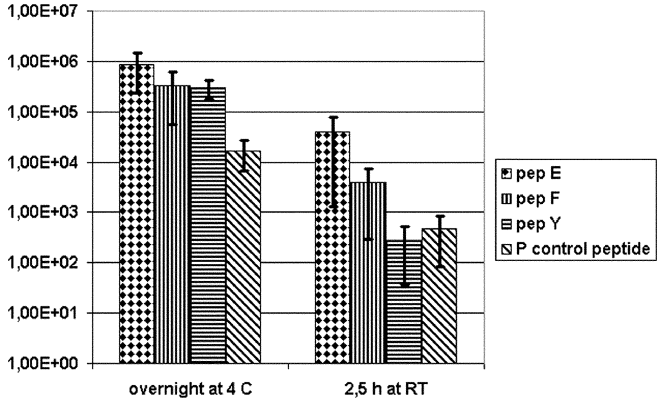

Fig. 2 Peptide selective recovery of phages after panning on membrane-bound peptide ligands with the model phages diluted 1:100 with T7 phages without any insert

The technical prerequisites of this concept were evaluated using a set of three model domains and their respective known peptide ligands. Table 1 lists the properties of the ones chosen which belong to the group that binds proline-rich ligands. Such domains are found in a large number of proteins without any obvious common function and are involved in a number of cellular processes, including subcellular localization, G-protein signalling, tyrosine kinase regulation and cytoskeleton assembly [17, 18, 19, 20]. The most important features for this study are their different affinity for the peptide ligands covering lower and higher affinity interactions as well as their characteristic difference in the size of the coding DNA to allow for an easy identification by PCR and gel electrophoresis sizing. The respective three peptide ligands together with an unrelated control peptide of comparable amino acid composition were synthesized as a 2×2 array on a 1-cm^2 cellulose membrane.

We also compared two different phage display systems, T7 [21] and lambda [22], with respect to peptide-specific enrichment on our type of peptide arrays. The major difference between these phage vehicles is, as reported in ref. [23], the density of protein presentation on the phage capsids, which is considerably higher for lambda. This results in a higher avidity of lambda complexes with a multitude of immobilized ligands. It was also reported that the selectivity for high-affinity ligands is best with T7 and for low-affinity ligands best with lambda [23]. A commercially available library of 10^7 primary clones made from human brain cDNA, with inserts size from 300 bp to greater then 3,000 bp and cloned into T7 phage *C*-terminal to the major capsid protein 10, was chosen as a first biological project.

Establishment of biopanning conditions

With the aim to optimise efficiency and selectivity of the binding and elution of phage-displayed protein domains using the peptide ligands synthesized on a cellulose membrane support, several experiments were carried out comparing various blocking and dissociation reagents as well as incubation conditions. To keep the number of unspecific bound phages on the SPOT membrane at a minimum, blocking buffers presented in Table 2 were tested. In this case the model phages were combined in an equal ratio and diluted 1:10 or 1:100 with parent phages having no inserts. We then compared the average quantities of phages eluted from the domain-specific peptides to that eluted from the control peptide. The best signal-to-background ratio was obtained for all peptide ligands with a combination of 5% skimmed milk + 50 µg mL^{-1} DNA + 50 µg mL^{-1} heparin in 0.1% TBST (data not shown).

The effect of temperature and time on selective binding was evaluated qualitatively as above by comparing the quantity of eluted model phages (Fig. 2) as well as quantitatively by further determining the percentage of correct phages through DNA insert analysis (Fig. 3). The average quantity of eluted phages after incubation overnight at 4 °C was about 300-fold higher than after incubation for 2.5 h at room temperature. Phages eluted from each peptide spot were separately plated on *E. coli* agar plates. From the plaques obtained for each peptide spot, at least 20 were randomly chosen for PCR analysis. True enrichment factors were calculated from the ratio of verified to analysed phages divided by the initial ratio of model to total phages (in this experiment 1:100). We observed that the enrichment factor for domain E under both binding conditions was almost equal (100) but increased for F from 60 to 85 and for Y from 25 to 95 at 4 °C overnight. This indicates that a longer incubation at lower temperature is more efficient when targeting weaker interactions. Moreover we did

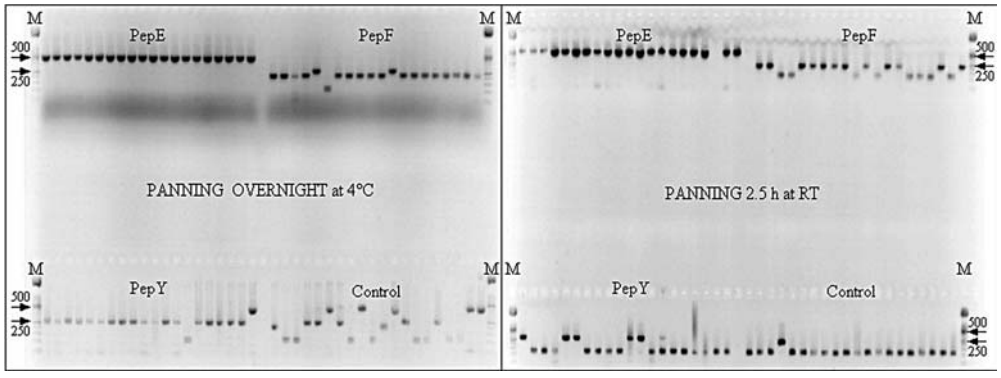

Fig. 3 Analysis of the DNA inserts of phage populations shown in Fig. 2. Domain-specific insert sizes are given in Table 1. T7 phage without insert yields a PCR product of 144 bp. *M* indicates the lanes with DNA molecular weight markers

Table 3 Panning studies with the membrane-bound peptide ligands using the model domains cloned into T7 phage and diluted with three distinct T7 phages displaying indifferent proteins

Domain	Peptide ligand	Dilution ratio of model phages				
		1:100	1:1,000	1:10,000	1:100,000	1:1,000,000
		Percentage of correct inserts after 1 panning				2 pannings
E	pepE	90	80	52.5	15	95
Y	pepY	92.5	85	44	0	22.5
F	pepF	90	55	7.5	2.5	12.5
		Enrichment factor after 1 panning				2 pannings
E	pepE	90	800	5,250	15,000	950,000
Y	pepY	92.5	850	4,400	n.d.	225,000
F	pepF	90	550	750	2,500	125,000

not observe enrichment of F and Y domains after incubation for 2.5 h at room temperature when the models phages were higher diluted (data not shown). However, after incubation overnight at 4 °C we got enrichment of all domains even when diluted down to $1{:}10^4$ (also see next section). The application of SDS as an elution reagent allowed us to elute the highest number of phages in comparison to the other reagents tested (data not shown). The SDS concentration itself did not show any significant influence on the amount of eluted phages. However SDS cannot be used for λ phage elution because of its sensitivity to this reagent. Therefore, only direct infection of bacteria could be applied in this case.

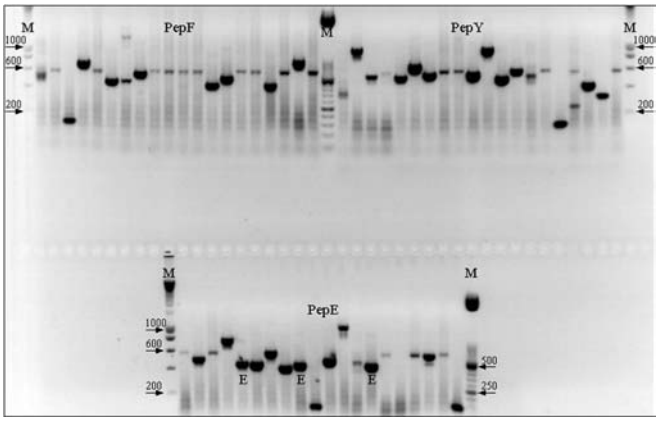

Fig. 4 Analysis of the DNA inserts of phage populations recovered from the membrane-bound model peptide ligands after two rounds of panning a T7 phage library that displayed human brain cDNA fragments supplemented with $1{:}10^6$ phages of each model domain. *M* indicates the lanes with DNA molecular weight markers

Model panning experiments

In order to extend to more realistic enrichment factors, the mixture of model phages was diluted to various degrees with a mixture of phages displaying indifferent proteins (mouse Il-4, fas antigen and ezrin). Still all phages could be discriminated via their insert sizes. Phages eluted from each spot were separately plated on *E. coli* agar plates. From plaques obtained for each spot, at least 20 randomly chosen were PCR-analysed and the enrichment factor was calculated (Table 3). Down to a $1{:}10^4$ dilution the enrichment of specific phages after a single panning round was good enough to detect correct inserts for each peptide spot. At a dilution of $1{:}10^5$ we analysed 40 plaques for each spot and, in spite of this, we could not find any correct insert for Y. Consequently, at a $1{:}10^6$ dilution, two consecutive rounds of panning were carried out before the PCR analysis. In this case, all model phages were identified on their respective target peptide and quite remarkable enrichment factors were observed.

Low expressed genes, which could be most problematic with our approach, are represented in cDNA libraries at a ratio of about 1 in 10^5 down to 1 in 10^6 [24]. In our simplified model experiment we could achieve a satisfactory enrichment of phage-displayed domains present at such a low concentration. To check enrichment under real conditions, a T7 phage display library made from human brain cDNA was applied to dilute the three model phages by a factor of $1{:}10^6$. After two consecutive rounds of panning on the SPOT membrane, 20 plaques from each ligand peptide were analysed. Figure 4 shows that the pools of phages obtained from each peptide look quite heterogeneous. Among those phages eluted from pepE, three carried the E domain. Neither the Y and F domains were identified within this rather small set of phage clones. However, within the phages eluted from for example the pepY ligand we identified by DNA sequencing the coding regions of other proteins containing WW domains such as the human atrophin-1 interacting protein 4 (AIP4) or the ubiquitin protein ligase ITCH mRNA. Many of the other protein ORFs collected need to be analysed in more detail. However, the PCR analysis of phage plaques is much too laborious and not parallel enough to suit a genome-wide approach. Obviously, considerably more phages need to be statistically analysed in order to recognize specific enrichment from such complex mixtures as exemplified here with the missing Y and F domains. A more sensitive and parallel method as the DNA microarray analysis should solve this problem.

Comparison of T7 phage with λ phage

We analysed both phage display systems with respect to the enrichment efficiencies on the peptide SPOT membrane,

Table 4 Comparison of λ and T7 phages(diluted 1:10^4 with their corresponding indifferent phages) for enrichment of the displayed model domains on their respective membrane-bound peptide ligands

Domain	Peptide	λ display 1	T7
		Percentage of correct inserts	
E	pepE	15	52.5
Y	pepY	10	44
F	pepF	65	7.5
		Enrichment factor	
E	pepE	1,500	5,250
Y	pepY	1,000	4,400
F	pepF	6,500	750

applying our model domains diluted 1:10^4 with a mixture of T7 phages displaying il-4, fas antigen, ezrin and correspondingly with λ phages lacking an insert. With both types of phages the same panning conditions were applied except the elution step, which had to be carried out by direct infection of bacteria in the case of λ phages. After a single round of panning, PCR and gel electrophoresis was used to identify the number of inserts matching to the respective peptide ligands. In contrast to T7, we observed a better enrichment of the F domain displayed by λ phages and lower enrichment of the other two domains (Table 4). This confirms the observation reported in ref. [23], as the F domain certainly belongs to the weaker binding domains. Thus, the λ phage display system is definitely preferable when weak protein interactions are to be identified. However, the T7 system is much more robust and many cloned cDNA libraries are already commercially available.

DNA microarray analysis

We tested whether expression profiling methods using DNA microarray techniques could be exploited to detect more sensitive and in a highly parallel mode the peptide-specific enrichment of phages after panning. We prepared simple microarray slides by contact printing those cDNAs that correspond to our model domains plus a heavy chain of a mouse antibody as control. Phages with domains E and F were mixed with human normal brain T7 select cDNA library in a ratio of 1:10^4. After a single round of panning, peptide spots were cut apart and bound phages were eluted separately. Eluted phages were amplified in liquid culture of *E. coli* BL 5615 until lysis occurred. Insert DNA from these phage populations obtained after panning was labelled with the cyanine dye Cy3. The insert DNA of phages from the mixture before panning was labelled with the cyanine dye Cy5. Additionally, the mouse control DNA was labelled with both Cy5 and Cy3 in two separate reactions. 1 µg each of the labelled DNA of one of the peptide-specific post-panning phage inserts, of the pre-panning DNA and the two control DNA preparations were carefully combined and the mixture was applied to a microarray slide. Figure 5 shows the result obtained with phages isolated from the pepE ligand: the DNA probe on the slide corresponding to the E domain hybridised preferentially with Cy3-labelled DNA (green colour), whereas the DNA probe corresponding to the F domain hybridised mainly with Cy5-labelled DNA (red colour). This means that DNA coding for the E domain is more abundant in the post-panning pool compared to the pre-panning pool. Thus, it clearly indicates enrichment of the E domain presenting phages on the expected target peptide. The DNA probe coding for the F domain shows an opposite behaviour, which confirms the lack of interaction with this pepE ligand. The control DNA hybridised to nearly equal amounts of Cy5- and Cy3-labelled DNA results in a yellow colour

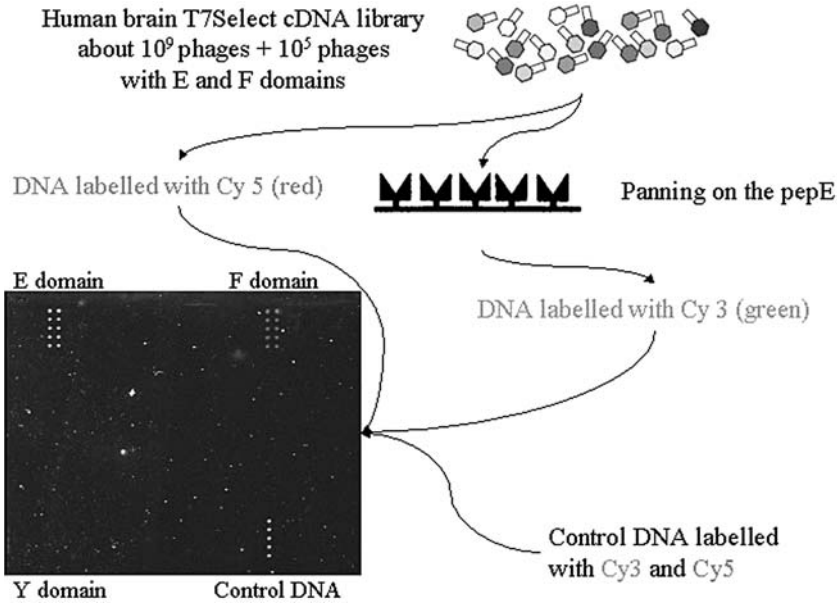

Fig. 5 Differential hybridisation of pre- and post-panning phage library DNA to domain-specific DNA probes printed onto a glass slide

and proves a proper quantification and composition of all DNA preparations. The DNA probe corresponding to the Y domain did not show any signal which suggests that this domain is not present in the library or only in such a low amount that it could not be detected. The panning of the F domain on its target peptide was also investigated by using DNA microarray analysis; however, as we expected the enrichment signal was much lower as compared to E.

Conclusion

Our concept of a peptide array-based multiplexed affinity enrichment of protein domains displayed on bacteriophage capsids was successfully applied to three model domains and their respective known peptide ligands. Now that first and new results are being collected from application to a library of protein fragments expressed from human brain cDNA, this challenges us to advance high-throughput performance of the process in order to achieve the anticipated genome-wide mapping of peptide-mediated protein–protein interactions. This, however, requires several further achievements related to:

- the optimised design of the peptide arrays: a realistic length for a biologically functional peptide fragment is about 12 amino acid residues but can be as small as three; generic peptide libraries as well as a genome-wide overlapping peptide scan covering all (e.g. human) ORFs would be adequate peptide collections for this approach; but these could be too large; utilizing combinatorial search strategies, less voluminous libraries of partially randomized peptide pools (10^3–10^6 components) with built-in information about recognition motives could be applied to reduce the number of peptide preparations, yet covering the whole relevant sequence space; these and other strategies are discussed in ref. [8].
- the massive and cheap production of cDNA microarrays: a respective approach based on the ultra-thin slicing with a microtome of a bundle of micro-fibres each loaded with for example a different DNA probe has been already conceived [25; also see http://zib.gbf.de/merk/].
- the process automation: most of the experimental steps of our process have been implemented into high-content screening formats using simultaneous parallel array methods and are carried out by only very few operations; yet, the spot-wise transfer of enriched phage populations from the peptide array sites to the DNA array analysis requires a series of standard operations which need to be carried out and evaluated by robotic instrumentation and software in order to achieve an adequate throughput; furthermore, an automated performance will definitely improve the quality and reliability of the whole process.

Compared to other approaches more focussed on certain types of peptide ligands [7, 26], we particularly expect the discovery of new families of peptide-recognizing protein domains from genome-wide screening.

Acknowledgements We are thankful to Lesley Mühle for endless PCR and gel analyses and to Susanne Daenicke for expert help in peptide array synthesis. We want to thank G. Cesarini for donation of λDisplay1 vector and A. Zucconi for helpful advice with the λ display system. This work was supported by BMBF grant 031U102G in the frame of the National Genome Research Net (NGFN). The authors are solely responsible for the content of this publication.

References

1. Geysen HM, Rodda SJ, Mason TJ (1986) Mol Immunol 23: 709–715
2. Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, Biswas M, Bucher P, Cerutti L, Corpet F, Croning MDR et al (2001) Nucleic Acids Res 29:37–40; http://www.ebi.ac.uk/interpro/. Cited 7 Nov 2002
3. Gavin AC, Bösche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz et al (2002) Nature 415:141–147
4. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P et al (2002) Nature 415:180–183
5. Fields S, Song OA (1989) Nature 340:245–246
6. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight J R, et al (2000) Nature 403:623–627
7. Tong AHY, Drees B, Nardelli G, Bader GD, Brannetti B, et al (2002) Science 295:321–324
8. Frank R (2002) Comb Chem High Throughput Screening 5: 429–440
9. Frank R (1992) Tetrahedron 48:9217–9232
10. Sambrook J, Fritsch EF, Maniatis T (1989) In: Molecular Cloning. CSH Laboratory Press, Cold Spring Harbor, NY
11. Niebuhr K, Ebel F, Frank R, Reinhard M, Domann E, Carl U D, Walter U, Gertler FB, Wehland J, Chakraborty T (1997) EMBO J 16:5433–5444
12. Chen HI, Sudol M (1995) Proc Nat Acad Sci 92:7819–7823
13. Ermekova KS, Zambrano N, Linn H, Minopoli G, Sudol M (1997) J Biol Chem 272:32869–32877
14. Frank R, Overwin H (1996) In: Morris GE (ed) Methods in molecular biology, vol 66: epitope mapping protocols. Humana Press, Totowa, pp 149–169
15. Hawlisch H, Müller M, Frank R, Bautsch W, Klos A Köhl J (2001) Anal Biochem 293:142–145; Hultschig C, Baensch M, Köhl J, Frank R (2001) In: Epton R (ed) Solid phase synthesis & combinatorial libraries 2000. Mayflower Worldwide, Birmingham, UK, pp 7–10
16. Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen B, Storz G, Botstein D, Brown PO (2000) Mol Biol Cell 11: 4241–4257
17. Feng S, Chen JK, Yu H, Simon JA, Schreiber SL (1994) Science 266:1241–1247
18. Callebaut I, Cossart P, Dehoux P (1998) FEBS Lett 441: 181–185
19. Duilio A, Zambrano N, Mogavero AR, Ammendola R, Cimino F, Russo T (1991) Nucleic Acids Res 19:5269–5274
20. Sudol M, Bork P, Einbond A, Kastury K, Druck T, Negrini M, Huebner K, Lehman D (1995) J Biol Chem 270:14733–14741
21. Rosenberg A, Griffin K, Studier FW, McCormick M, Berg J, Novy R, Mierendorf R (1996) InNovations 6:1–6
22. Castagnoli L, Zucconi A, Quondam M, Rossi M, Vaccaro P, Panni S, Paoluzi S, Santonico E, Dente L, Cesareni G (2001) Com Chem High Throughput Screen 4:121–133
23. Zucconi A, Dente L, Santonico E, Castagnoli L, Cesareni G (2001) J Mol Biol 307:1329–1339
24. Bishop JO, Morton JG, Rosbash M, Richardson M (1974) Nature 250:199–204
25. Frank R (2001) German Patent Application P 101 17 135.8
26. Sparks AB, Hoffman NG, McConnell SJ, Fowlkes DM, Kay B K (1996) Nat Biotechnol 14:741–744