A protein-domain microarray identifies novel protein–protein interactions

Alexsandra ESPEJO*, Jocelyn CÔTÉ†, Andrzej BEDNAREK‡, Stephane RICHARD† and Mark T. BEDFORD*[1]

*The University of Texas M.D. Anderson Cancer Center, Science Park – Research Division, P.O. Box 389, Smithville, TX 78957, U.S.A., †Lady Davis Institute for Medical Research, Departments of Oncology, Medicine, Microbiology and Immunology, McGill University, Montréal, Québec, Canada H3T 1E2, and ‡The Medical University of Lodz, Institute of Physiology and Biochemistry, Department of Biochemistry, Lindleya 6, 90-131 Lodz, Poland

Protein domains mediate protein–protein interactions through binding to short peptide motifs in their corresponding ligands. These peptide recognition modules are critical for the assembly of multiprotein complexes. We have arrayed glutathione S-transferase (GST) fusion proteins, with a focus on protein interaction domains, on to nitrocellulose-coated glass slides to generate a protein-domain chip. Arrayed protein-interacting modules included WW (a domain with two conserved tryptophans), SH3 (Src homology 3), SH2, 14.3.3, FHA (forkhead-associated), PDZ (a domain originally identified in PSD-95, DLG and ZO-1 proteins), PH (pleckstrin homology) and FF (a domain with two conserved phenylalanines) domains. Here we demonstrate, using peptides, that the arrayed domains retain their binding integrity. Furthermore, we show that the protein-domain chip can 'fish' proteins out of a total cell lysate; these domain-bound proteins can then be detected on the chip with a specific antibody, thus producing an interaction map for a cellular protein of interest. Using this approach we have confirmed the domain-binding profile of the signalling molecule Sam68 (Src-associated during mitosis 68), and have identified a new binding profile for the core small nuclear ribonucleoprotein SmB′. This protein-domain chip not only identifies potential binding partners for proteins, but also promises to recognize qualitative differences in protein ligands (caused by post-translational modification), thus getting at the heart of signal transduction pathways.

Key words: arginine methylation, proline-rich motifs, Sam68, signalling, SmB′.

INTRODUCTION

As we pass from the genomic to the proteomic era, it is becoming necessary and possible to complement existing techniques of gene expression profiling with chip-size protein microarrays. The development of protein microarrays is in its infancy and, as with all emerging technologies, there are different approaches being taken to generate such arrays. Initial studies arrayed cDNA expression libraries on PVDF membranes [1]. This approach has been successful in identifying protein binding partners and enzyme substrates [2,3]. Recent studies have paved the way for the rapid development of high-density protein arrays on glass slides by establishing the concentration of protein needed, identifying surfaces for protein immobilization, and demonstrating that fluorophore-tagged proteins can be used effectively to trace and compare protein–protein interactions [4–6]. MacBeath and Schreiber [5] arrayed proteins on aldehyde slides that were probed with fluorophore-tagged proteins to identify protein–protein interactions. A single specific interaction was detected in a field of 10000 spots. Using a synthetic ligand for FKBP12 (FK506-binding protein 12), this microarray format was also used to demonstrate the feasibility of identifying protein–small-molecule interactions. The group led by Pat Brown has also made inroads into the protein microarray field [7], using poly(L-lysine) slides. Their study focused on the arraying of antibody/antigen pairs. Two different fluorophore-tagged protein solutions were used as probes to assay for the relative specificity and abundance of proteins at physiologically relevant concentrations. Recently, nickel-coated slides were used to array glutathione S-transferase (GST)/His_6-tagged yeast proteins [4]. In that study, 5800 yeast open reading frames were arrayed, and new phospholipid- and calmodulin-interacting proteins were identified. A different approach was taken by Ziauddin and Sabatini [6], who printed cDNA expression vectors on glass slides and then cultured mammalian cells on the arrayed slide. This resulted in small areas of transfection and protein expression where the cDNA was arrayed. This technique has been termed 'transfected cell microarrays', and has successfully identified gene products that alter cellular physiology.

The stimulation of cells from outside triggers cascades of signal transduction that result in cellular responses such as growth, differentiation and movement. These signals are transduced by networks of interacting proteins [8]. As a result of the enormous body of data gathered in recent years regarding protein–protein interactions, it has become clear that a large proportion of protein interactions occur between a domain in one protein and a small motif (usually 8–15 amino acids) in its ligand [9,10]. These diverse associations are mediated through interactions of a limited number of modular signalling units or protein domains. Protein-interacting domains are classified in the protein family (Pfam) database (http://pfam.wustl.edu/). Protein interactions involving domains are often regulated by post-translational modification (phosphorylation, methylation and acetylation) of the smaller protein motif. The phosphorylation of proteins on serine and threonine residues can regulate 14.3.3, forkhead-associated (FHA) and WD40 (conserved

Abbreviations used: FF domain, a domain with two conserved phenylalanines (F); FHA, forkhead-associated; GST, glutathione S-transferase; KH, ribonucleoprotein K homology; PBST, PBS containing 0.1% Tween 20; PDZ domain, a domain originally identified in PSD-95, DLG and ZO-1 proteins; PGM, proline/glycine/methionine; PH, pleckstrin homology; PTB, phosphotyrosine-binding; Sam68, Src-associated during mitosis 68; SH3, Src homology 3; SH2, Src homology 2; WBP, WW-domain-binding protein; WW domain, a domain with two conserved tryptophans (W).

[1] To whom correspondence should be addressed (e-mail mbedford@sprd1.mdacc.tmc.edu).

sequence of 40 amino acids that ends in Trp–Asp) domain binding [11], and the phosphorylation of proteins on tyrosine residues can regulate Src homology 2 (SH2) and phosphotyrosine-binding (PTB) domain binding [12]. In addition, lysine acetylation [13], and lysine and arginine methylation [14–17], have also been implicated in the regulation of protein–protein interactions.

Here we use characterized modular protein domains to generate a chip that can be used to screen for protein interactions. Included in the protein domain array are WW (a domain with two conserved tryptophans) [18,19], SH3 [20], SH2 [9], 14.3.3 [21], FHA [11,22], PDZ (a domain originally identified in PSD-95, DLG and ZO-1 proteins) [23], pleckstrin homology (PH) [24] and FF (a domain with two conserved phenylalanines) [25] domains. In order to define the parameters of this approach we have focused our attention on the WW- and SH3-domain region of this array. These domains (WW and SH3) bind proline-rich ligands. The signalling molecule Sam68 (Src-associated during mitosis 68) and the core small nuclear ribonucleoprotein SmB′ are two such proline-rich molecules. Sam68 interacts with SH3 and WW-domain-containing proteins [16], and SmB′ associates with the WW domains of the spliceosome-associated protein formin-binding protein ('FBP') 21 [26]. Using peptides derived from these proteins and specific antibodies, we have been able to demonstrate distinct and reproducible binding to subsets of protein domains, thus generating domain-binding profiles for cellular proteins.

EXPERIMENTAL

Purification of GST fusion proteins

GST fusion proteins were overexpressed in *Escherichia coli* DH5α cells (Life Technologies, Rockville, MD, U.S.A.) by induction with a final concentration of 0.4 mM isopropyl β-D-thiogalactopyranoside. Cells were broken by sonication. The resulting lysates were centrifuged at 12000 ***g*** for 10 min, and the GST fusion proteins were then batch-purified from extracts by binding to glutathione–Sepharose 4B beads (Amersham Pharmacia Biotech, Piscataway, NJ, U.S.A.) and washed in PBS according to the manufacturer's instructions. The purified proteins were eluted from the beads with 30 mM glutathione, 50 mM Tris/HCl, pH 7.5, and 120 mM NaCl. The purified proteins were stored in the elution buffer at -70 °C in 384-well plates.

Production of protein microarrays

The proteins were arrayed in duplicate using a FLEXYS® robot (Genomic Solutions, Ann Arbor, MI, U.S.A.). We used a high-density arrayer (HAD) 48 Pin Head (FLX 12021) for arraying. The GST fusion proteins were arrayed from a 384-well plate which contained 10 μl of each protein at a concentration of 1 μg/μl. The protein stocks were in elution buffer (no glycerol was added), and each protein was arrayed five times on to the same spot to increase the local concentration of protein. Proteins were spotted on to a glass slide precoated with nitrocellulose polymer (FAST™ Slide; Schleicher & Schuell, Keene, NH, U.S.A.) to generate a rectangular array (3.1 cm $\times$ 2 cm). The arrayed proteins were air-dried. The protein chip is composed of 20 grids each, in a 5 row $\times$ 5 column format, with a distance of 700 μm between spots. Each grid thus contains 12 fusion proteins arrayed in duplicate, with GST alone spotted in the middle of the grid. A 384-well master plate containing 10 μl of each protein (1 μg/μl) was sufficient for arraying 35–40 slides. Thus each spot contains approx. 250 ng of fusion protein.

Probing the protein-domain array with a labelled peptide or proteome

Peptide probes

Peptides were synthesized by the W. M. Keck Biotechnology Resource Center (New Haven, CT, U.S.A.). Biotinylated peptides (10 μg) were pre-bound to 5 μl of Cy3–streptavidin or Cy5–streptavidin (Fluorolink™; Amersham Pharmacia Biotech) in 500 μl of PBST (PBS containing 0.1 % Tween 20). The fluorescently labelled peptide was then incubated with 20 μl of biotin–agarose beads (Sigma, St. Louis, MO, U.S.A.) to remove the free streptavidin label. Arrayed slides were blocked in PBST containing 3 % (w/v) powdered milk, followed by the addition of 400 μl of fluorophore-tagged peptide. Blocking and hybridization were performed in an Atlas Glass Hybridization Chamber (Clontech, Palo Alto, CA, U.S.A.). After 1 h of incubation at room temperature, the unbound peptide was washed away with PBST, three times for 10 min each.

Proteome probes

Human MCF7 cells were grown to 80 % confluency. Cells were then scraped into a mild lysis buffer [PBS, pH 7.2, containing 100 mM NaCl and Complete™ protease inhibitor cocktail (Roche, Indianapolis, IN, U.S.A.)]. Cells were subsequently broken by two 30 s sonicator pulses. The array was first blocked in PBST/3 % (w/v) powdered milk and then incubated for 1 h with 2 mg/ml MCF7 cell lysate. The array was washed for 3 $\times$ 10 min in PBST. Bound protein was then detected by probing for 1 h with a primary antibody to the protein of interest, working at a dilution used for Western blot analysis (1:1000). The array was washed for 3 $\times$ 10 min in PBST. The primary antibody was recognized with an appropriate FITC-conjugated secondary antibody. All incubations were performed at room temperature. The anti-peptide antibody against Sam68 has been described previously [27]. The anti-SmB′ antibody, Ana128, was obtained from ICN Biomedicals (Costa Mesa, CA, U.S.A.). When performing endogenous protein profiling, it is important that the primary antibody is raised against a peptide and not against a GST fusion protein, as this would result in cross-reactivity with the arrayed proteins.

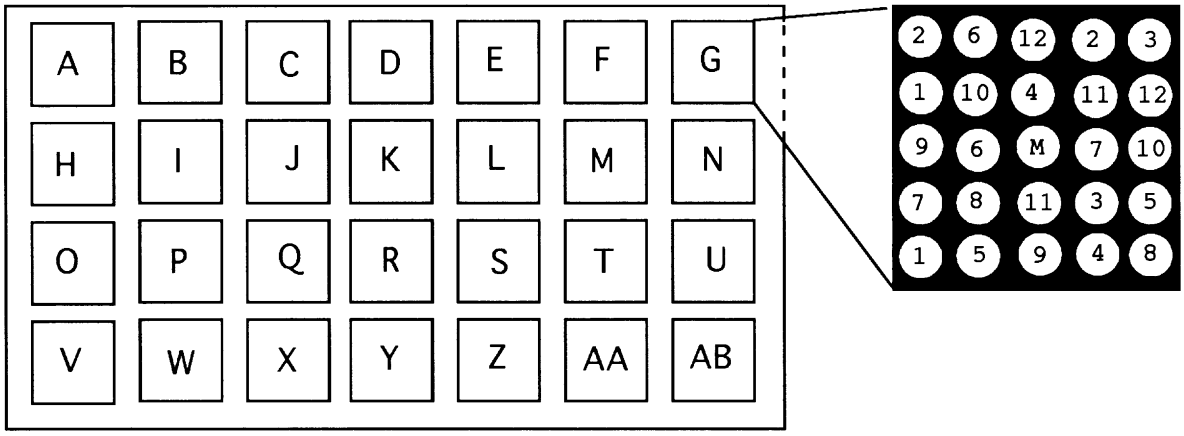
Probe detection

Following the washes, the slides were centrifuged dry and the fluorescent signal was detected using a GeneTAC™ LSIV scanner (Genomic Solutions). A 550 nm long pass filter was used for the detection of Cy3-labelled probes and FITC-conjugated secondary antibodies. A 675 nm band pass filter was used for the detection of Cy5-labelled probes. A positive signal is seen as two dots at varying angles.

RESULTS

Generation of protein-domain microarrays

We have generated and gathered, from fellow researchers, GST expression vectors harbouring protein domains as well as proteins of general interest. A list of 212 constructs is shown in Figure 1, from which we generated GST fusion proteins for our master plate. A total of 145 protein domains are represented, broken down as follows: 33 WW, 23 SH3, 17 SH2, 23 PH, 23 PDZ, seven 14.3.3, five PTB, four FHA, eight FF and two ribonucleoprotein K homology (KH). An additional 67 GST fusion proteins without canonical protein domains are listed in the 'Other' section. We have arrayed this set of GST fusion proteins in



WW		WW		WW		SH3		SH3		SH2	
A1	FBP11A (f)	B1	WWOX2(f)	C1	SMN4(q)	D1	Src(f)	E1	p120 ras Gap(o)	F1	Shc (f)
A2	FBP11 (f)	B2	WWOX1(f)	C2		D2	p85(f)	E2	Fyn-3 SH3/SH2(o)	F2	Src (f)
A3	NEDD4(f)	B3	ESS I(q)	C3	HYAP-WW1mutant(q)	D3	Vav c-term(o)	E3	Grb2 c-term(o)	F3	Shc (o)
A4	FBP30AB (f)	B4	Ca150 WWI (a)	C4	HYAP-WW2(q)	D4	Yrk(q)	E4	Nck(o)	F4	G-lck SH3/SH2(o)
A5	FBP30A (f)	B5	Pin 1(f)	C5	Fe65 w.t.(q)	D5	PRMT2(f)	E5	Spectrin(o)	F5	Vav F.L. SH3/SH2(o)
A6	HYPC (f)	B6	YJQ8(q)	C6	HYAP WW1 w.t.(q)	D6	Vav n-term(o)	E6		F6	Grb2(o)
A7	YAP (f)	B7	Ca150 WW2(a)	C7	YAP chicken(q)	D7	Us yes(q)	E7	Abl(o)	F7	p59 fyn SH3/SH2(o)
A8	FBP28(f)	B8	Itck B(f)	C8	Fe65 WW mutant(q)	D8	Yes(f)	E8	SLAP SH3/SH2(o)	F8	Vav SH2-GST(o)
A9	PDGFR Short(f)	B9	RSP II(q)	C9		D9	G-lck SH3/SH2(o)	E9	Ras GAP(o)	F9	PI3K c-term(o)
A10	FBP21(f)	B10	Itck A(f)	C10	DYS WW(q)	D10	Abl(f)	E10	Fyn-2(o)	F10	PI3K(f)
A11	PDGFR Long(f)	B11	RSP I(q)	C11		D11	Vav F.L. SH3/SH2(o)	E11	PLC(o)	F11	Lck(o)
A12	FBP11B(f)	B12	WWOX 1&2(f)	C12	DYS WW mutant(q)	D12	Fyn-1(f)	E12	PLC gamma(o)	F12	Yes(f)

SH2		PH		PH		14.3.3		PDZ		PDZ	
G1	PLC SH2/SH2(o)	H1	Dynamin-2(n)	I1	hSos(n)	J1	14.3.3 Zeta(f)	K1	NHERF PDZ2(e)	L1	PDZ CAP1(g)
G2	PI3K n-term(o)	H2	Dynamin-1(n)	I2	Ras GAP(n)	J2	14.3.3 Beta(f)	K2	NHERF PDZ1(e)	L2	PDZ3 CAP1(g)
G3		H3	Akt(n)	I3	PLC (209-572)(b)	J3		K3	Alfa-syn(m)	L3	PDZ8 (z)
G4		H4	Spectrin(n)	I4	PLC (5-180)(b)	J4	14.3.3 Eta(f)	K4	Gamma-syn(m)	L4	PDZ9 (z)
G5		H5	KIAA 0053(n)	I5	PLC (5-295)(b)	J5	14.3.3 Epsilon(e)	K5	Densin PDZ2(i)	L5	PDZ10 (z)
G6		H6	PLC gamma(n)	I6	PLC (570-848)(b)	J6		K6	GST PDZ3(I)	L6	PDZ 11/12 (z)
G7		H7	Pleckstrin(n)	I7	PLC (5-150)(b)	J7		K7	Beta-syn(m)	L7	PDZ 12 (z)
G8	Fyn(o)	H8	PLC delta(n)	I8	PLC (849-1234)(b)	J8	14.3.3 Gamma(e)	K8	AF6 PDZ1(i)	L8	PDZ 13 (z)
G9		H9	DAGK delta(n)	I9		J9		K9	NHERF2(g)	L9	KIT (z)
G10	Abl(o)	H10	DAPP1(n)	I10	Itk(n)	J10	14.3.3 Tau(f)	K10	SAP 97 3x PDZ(f)	L10	KIT delta 976 (z)
G11		H11	Beta ARK-1(n)	I11	PLC (5-208)(b)	J11		K11	NHERF1(g)	L11	
G12	SLAP SH3/SH2(o)	H12	Grp-1(n)	I12	PLC Beta3(n)	J12	14.3.3Sigma(e)	K12	SAP 97 F.L.(f)	L12	PDZ CAP2(g)

FHA/PTB/KH		FF		OTHER		OTHER		OTHER		OTHER	
M1	RAD 53 FHA2(e)	N1	FBP 11(1)(f)	W1	GST-CTD(w)	X1	HDAC1 (v)	Y1	PRMT1 (f)	Z1	EHR(f)
M2	RAD 53 FHA1(e)	N2	CA 150 (1-3)(f)	W2	Cox2(f)	X2	HDAC2 (v)	Y2	PRMT2 (f)	Z2	EH(f)
M3	PTB2 FE 65 L2(s)	N3	CA 150 (4-6)(f)	W3	MRE11 c-term (f)	X3	HDAC3 (v)	Y3	PRMT3 (f)	Z3	Xiap (124-240)(p)
M4	PTB1 FE 65(s)	N4	FBP 11 (3-4) (f)	W4	TARPP c-term (f)	X4	HDAC4c1 (u)	Y4	CARM (f)	Z4	GAR(f)
M5	PTB DAB2(k)	N5	PRP (1-2)(f)	W5	TARPP F.L. (f)	X5	HDAC4C6 (u)	Y5	PRMT6 (f)	Z5	SMN(f)
M6	QKI KH(o)	N6		W6	FKHRL w.t.(x)	X6	HDAC5R4 (u)	Y6	PABP F.L.(f)	Z6	BARDI 13-638(d)
M7	PTB2 FE 65(s)	N7	p190B(3-4)(f)	W7	FKHRL m(x)	X7	HDAC1 (t)	Y7	PABP hind III(f)	Z7	Histone 3(f)
M8	PTB DAB1(k)	N8	p190B(1-2)(f)	W8	FKHRL N-term(x)	X8	HDAC2 (t)	Y8	PABP c-term (f)	Z8	EVL(f)
M9		N9	p190B(1-4)(f)	W9	hDM2 (x)	X9	HDAC3 (t)	Y9	PABP w.t.(f)	Z9	HPI B F.L.(j)
M10	KI FHA (1-168)(r)	N10		W10	hER alfa(x)	X10	HDAC10 (f)	Y10	PABP R-1(f)	Z10	MENA(f)
M11	SAM 68 KH(o)	N11		W11	Tubby c-term (z)	X11		Y11	PABP R-1.2 (f)	Z11	BARDI 13-777(d)
M12	CHN2 FHA(e)	N12		W12	Tulp3 c-term (z)	X12		Y12		Z12	Profilin(f)

OTHER		OTHER	
AA1	p53 F.L.(q)	AB1	WBP 11-5'(f)
AA2	U1A(f)	AB2	SRPK 2-5'(f)
AA3		AB3	Id 28(f)
AA4	DP-1(c)	AB4	Id 10(f)
AA5	p73 beta(h)	AB5	Sap 62 3x motif(f)
AA6	p53 BP-2(q)	AB6	B dystroglican(f)
AA7	E2F4(c)	AB7	3 BPI(f)
AA8	p73 alfa(h)	AB8	11 PGM(f)
AA9	p53 BP-2mt2(q)	AB9	SMB Full(f)
AA10	E2F1(c)	AB10	SmB'(f)
AA11	p53 BP-2mt1(q)	AB11	U1C(f)
AA12	p53 W.T(h)	AB12	WBP 11-3'(f)

VECTOR SOURCE:

a- Gracia Blanco, Mariano.
b- Snyderman, Ralph.
c- Johnson, David.
d- Manley, James.
e- Yaffe, Michael B.
f- Bedford, Mark.
g- Biber, Jürg.
h- Blandino, Giovanni.
i- Borg, Jean-Paul.
j- Kouzarides, Tony.
k- Cooper, Jonathan A.
l- Dobrosotskaya, Irina.
m- Gee, Stephen.
n- Lemmon, Mark A.
o- Richard, Stéphane.
p- Shi, Yigong.
q- Sudol, Marius.
r- Takagi, Masatoshi.
s- Zambrano, Nicola.
t- Chang, Kun-Sang.
u- Jianrong, Lu.
v- Yang, Wen-Ming.
w- Dynan, William.
x- Hu, Mickey.
y- Shapiro, Lawrence.
z- Mancini, Annalisa

Figure 1 List of 212 proteins, protein domains and protein motifs that have been purified as GST fusion proteins and arrayed on nitrocellulose-coated slides

The upper panel shows the design of the array. Slides (right) feature a 5 × 5 grid pattern. Proteins are arrayed in duplicate, and each block harbours 12 different fusion proteins. The middle position (M) contains GST alone as a negative marker and background indicator. The lower panel lists all the arrayed fusion proteins and their position on the slide. The domain type is indicated at the top of each box (WW, SH3, SH2, PH, 14.3.3, PDZ, FHA/PTB/KH and FF). 'Other' refers to GST fusion proteins that do not contain domains but are of general interest. Workers that have contributed to this array are listed.

duplicate on to the nitrocellulose film of FAST™ Slides, using a robot. Each spot contained approx. 250 ng of fusion protein. The arrayed proteins were allowed to dry on the nitrocellulose surface and the slide was subsequently stored at 4 °C.

Peptide motifs bind specifically to immobilized protein domains

To test the integrity of the arrayed domains, we synthesized biotinylated peptides that are known to bind specifically to certain protein domains. These peptides were conjugated to streptavidin–Cy3 and then used to probe the protein-domain array. The peptides used included the proline-rich P3 motif of Sam68 (which binds SH3 and WW domains [16]) (Figure 2F), the proline/glycine/methionine (PGM) motif of the splicing factor SmB′ (which binds group III WW domains [26]) (Figure 2B), the PPYP motif of WBP1 (WW-domain-binding protein 1, which binds group I WW domains [18]) (Figure 2D) and the C-terminus of Kv1.4 (which binds a subset of PDZ domains [28]) (Figure 2C). Unique binding profiles were detected for each of these peptides in the predicted regions of the array, demonstrating that these arrayed domains are functional and that their binding specificity is intact.

Post-translational modification of a peptide changes the binding profile

Post-translational modifications are key events associated with the initiation or the redirection of signalling pathways. We focused on arginine methylation to assess the ability of the array to distinguish between unmodified and modified peptides. We used the P3 motif of Sam68, which we have shown to display reduced binding to SH3 but not WW domains when arginine-methylated [16]. The unmethylated peptides were conjugated to streptavidin–Cy3 and the methylated peptides were conjugated to streptavidin–Cy5. A mixture of the labelled peptides was then used to probe a single protein-domain array. The unmethylated P3 peptide (Figure 2F) bound two WW domains and six SH3 domains. The methylated P3 peptide (Figure 2G) bound the same two WW domains, but only three SH3 domains. The SH3 domains that are sensitive to arginine methylation are most obvious when both the Cy3 and Cy5 signals are read simultaneously. Domains that bind the P3 motif regardless of its methylation state are shown in yellow, whereas those domains that are sensitive to methylation are shown in green (Figure 2H). A similar approach can be taken to screen for phosphorylation-regulated protein–protein interactions.

Detection of the binding profiles of individual proteins from a total-cell lysate

Next we attempted to determine the binding profiles of endogenous proteins. To do this we established an ELISA-based method. First, a total-cell lysate from MCF7 cells was used to probe a protein-domain array. The array was then washed with PBST and re-probed with an antibody raised against a protein of interest. Finally, the primary antibody was detected with a FITC-labelled secondary antibody. We chose to look at the binding profiles of two endogenous proteins, Sam68 and SmB′. We have obtained distinct binding profiles for these two proteins using short peptides (Figures 2B and 2F) that represent just a fraction of their entire length (Figure 3B). Previous studies had identified the P3 [16,29] and PGM [26] motifs as the dominant protein-interacting regions within Sam68 and SmB′ respectively. We thus reasoned that the binding profile of the peptide should be indicative of how the full-length protein would bind to the domain array, with perhaps a slightly broader binding spectrum for the endogenous protein, as it may harbour additional interacting motifs.

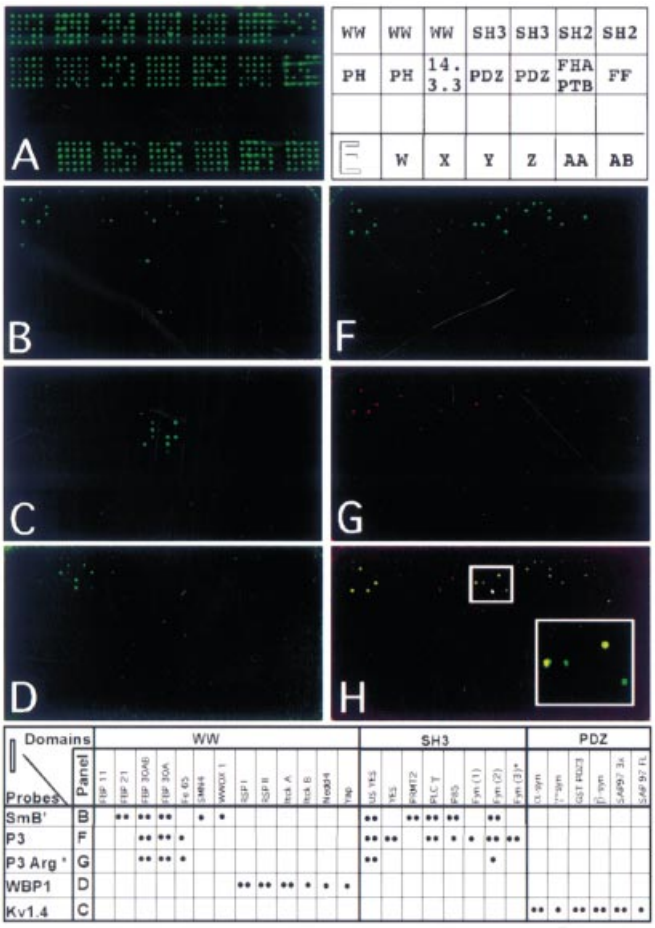


Figure 2 GST fusion proteins containing protein-interacting domains retain their binding specificity in a microarrayed format

GST fusion proteins (212 in total) were arrayed in duplicate on to a nitrocellulose slide. (**A**) The array was probed with an anti-GST primary antibody and detected with a FITC-conjugated secondary antibody. (**B**) The array was probed with a Cy3-labelled SmB′ peptide (biotin–PPGMRPPPPGMRRGPPPPGMRPPRP). (**C**) The array was probed with a Cy3-labelled PDZ ligand peptide from Kv1.4 (biotin–SGSGSNAKAVETDV-CO_2H). (**D**) The array was probed with a Cy3-labelled WBP1 peptide (biotin–SGSGGTPPPPYTVG). (**E**) Key to the arrayed domains. The identity of the interacting domains can be extrapolated from Figure 1. In (**F**)–(**H**), the same protein array was probed with a Cy3-labelled P3 peptide of Sam68 (biotin–GVSVRGRGAAPPPP-PVPRGRGVGP) and with a Cy5-labelled P3 arginine-methylated peptide of Sam68 (biotin–GVSVR*GR*GAAPPPPPVPR*GR*GVGP; asterisks denote asymmetrically dimethylated arginine residues). (**F**) Detection of Cy3 signal; (**G**) detection of Cy5 signal; (**H**) Cy3 and Cy5 signals are superimposed. The yellow signal indicates protein interactions that are insensitive to arginine methylation, and the green signals mark protein interactions that are sensitive to arginine methylation (see inset). (**I**) Diagrammatic representation of the signals seen in (**A**)–(**H**). A single dot indicates a signal of low intensity, a double dot indicates a signal of high intensity, and a single asterisk denotes a GST fusion protein that contains both SH3 and SH2 domains.

Using this ELISA approach, we obtained distinct signals with antibodies to Sam68 and SmB′ (Figures 3A and 3C). Both of these molecules are proline-rich, and as such bind SH3 and WW domains. The proline-rich sequences of Sam68 and SmB′ are different (Figure 3B) and they display binding profiles that are distinct, with some overlap. The SH3 and WW domain regions of the probed arrays were enlarged and compared. The patterns

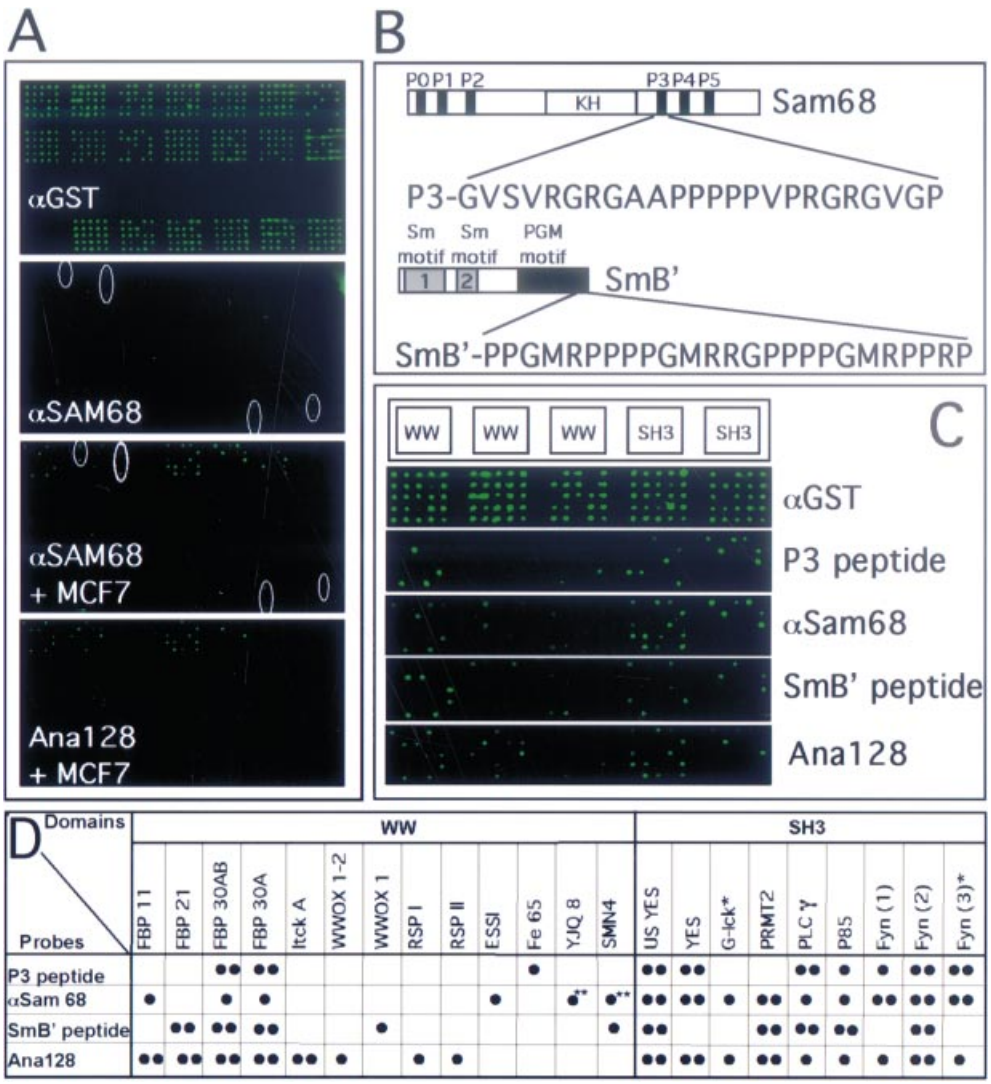


Figure 3 The protein-domain array detects endogenous protein binding profiles

(**A**) Top panel: the array was probed with an anti-GST primary antibody (αGST) and detected with a FITC-conjugated secondary antibody. Second panel: the array was probed with an antibody to Sam68 and detected with a FITC-conjugated secondary antibody. Third panel: the array was first incubated with 2 mg/ml MCF7 cell lysate, and then probed with an antibody to Sam68, followed with a FITC-conjugated secondary antibody. Bottom panel: the array was first incubated with 2 mg/ml MCF7 cell lysate, and then probed with an antibody to SmB′ (Ana128), followed with a FITC-conjugated secondary antibody. The circled signals represent non-specific interactions detected by the Sam68 primary antibody. No non-specific binding was detected with the Ana128 antibody. (**B**) Schematic representation of Sam68 and SmB′, and positions of the biotinylated peptides used as probes. (**C**) Alignment of the WW- and SH3-domain section of arrays probed with a total cell lysate, followed by a specific antibody (anti-Sam68 or Ana128), compared with the binding seen with a nested peptide from these two proteins (see Figures 2B and 2F). (**D**) Diagrammatic representation of the signals seen in (**C**). A single dot indicates a signal of low intensity, a double dot indicates a signal of high intensity, a single asterisk denotes a GST fusion protein that contains both SH3 and SH2 domains, and a double asterisk marks background signals generated by cross-reactivity of the primary antibody.

of binding observed for the cellular proteins were very similar to those seen with the respective peptides (Figures 3C and 3D). Thus binding of the synthetic peptide reflects the binding profile of the endogenous protein.

DISCUSSION

In the present paper we show that a protein-domain array can be used to detect interactions between peptides and arrayed proteins, as well as between endogenous cellular proteins and the array. Of paramount importance is the fact that this array not only will detect the binding profiles of cellular proteins, but also will identify those post-translational modifications that create or prevent protein–protein interactions. Using this protein-domain array, we have identified the binding profiles of two proline-rich proteins, Sam68 and SmB′. Sam68 binds both SH3 and WW domains [16,29], and thus the observed binding profile was expected (Figures 3A and 3C). The propensity of the core splicing factor, SmB′, for WW domain binding has been reported [26], but the degree of SH3 domain binding demonstrated by SmB′ was rather unexpected. Splicing factors are localized in nuclear speckle domains *in vivo* [30], and both SmB′ and WW-domain-containing proteins have been shown to co-localize with SC35 speckles [26]. Recently, two different SH3-domain-containing proteins were also shown to co-localize with speckles [31,32]. It is thus possible that SmB′ forms an attachment scaffold for WW- and SH3-domain-containing spliceosome components.

Traditional methods of domain-interaction mapping rely on blot overlay experiments [20,33–35]. Using this approach, protein interactions are examined by electrophoretically separating fusion proteins containing regions of interest on denaturing SDS/PAGE gels, transferring them to nitrocellulose and incubating them with radiolabelled domain-containing fusion protein probes. Such blot overlay experiments have been used to obtain profiles of cellular tyrosine phosphorylation states using SH2 domains [36]. The protein-domain chip can be used for similar mapping and profiling experiments. This chip has the following advantages over the blot overlay approach: (1) a large number of protein domains are screened at a time; (2) conditions are mild (SDS/PAGE is not used), so proteins are more likely to be in their native conformation; (3) the screens are fast (3–4 h) once the arrays have been generated; (4) different fluorescent labels can be used to evaluate the consequence of post-translational modifications on domain binding; and (5) the ELISA-based detection of endogenous proteins is sensitive.

In summary, the protein-domain array described in the present work could be very effective when used hand-in-hand with a motif-based searching algorithm such as Scansite [37] (http://scansite.mit.edu/). Short linear sequence motifs that are predicted to bind domains could be identified within query proteins and verified experimentally using this type of array.

M.T.B. is supported by the Damon Runyon Cancer Research Foundation Scholar Award DRS-28-02 and NIEHS Center Grant ES07784. We thank all those researchers that contributed to the GST collection, as listed at the bottom of Figure 1.

REFERENCES

1 Bussow, K., Cahill, D., Nietfeld, W., Bancroft, D., Scherzinger, E., Lehrach, H. and Walter, G. (1998) A method for global protein expression and antibody screening on high-density filters of an arrayed cDNA library. Nucleic Acids Res. **26**, 5007–5008
2 Lee, J. and Bedford, M. T. (2002) PABP1 identified as an arginine methyltransferase substrate using high-density protein arrays. EMBO Rep. **3**, 268–273
3 Bussow, K., Nordhoff, E., Lubbert, C., Lehrach, H. and Walter, G. (2000) A human cDNA library for high-throughput protein expression screening. Genomics **65**, 1–8
4 Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T. et al. (2001) Global analysis of protein activities using proteome chips. Science **293**, 2101–2105
5 MacBeath, G. and Schreiber, S. L. (2000) Printing proteins as microarrays for high-throughput function determination. Science **289**, 1760–1763
6 Ziauddin, J. and Sabatini, D. M. (2001) Microarrays of cells expressing defined cDNAs. Nature (London) **411**, 107–110
7 Haab, B. B., Dunham, M. J. and Brown, P. O. (2001) Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. Genome Biol. **2**, research0004.0001–0004.0013
8 Pawson, T. and Nash, P. (2000) Protein–protein interactions define specificity in signal transduction. Genes Dev. **14**, 1027–1047
9 Pawson, T., Raina, M. and Nash, P. (2002) Interaction domains: from simple binding events to complex cellular behavior. FEBS Lett. **513**, 2–10
10 Tong, A. H., Drees, B., Nardelli, G., Bader, G. D., Brannetti, B., Castagnoli, L., Evangelista, M., Ferracuti, S., Nelson, B., Paoluzi, S. et al. (2002) A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. Science **295**, 321–324
11 Yaffe, M. B. and Elia, A. E. (2001) Phosphoserine/threonine-binding domains. Curr. Opin. Cell Biol. **13**, 131–138
12 Shoelson, S. E. (1997) SH2 and PTB domain interactions in tyrosine kinase signal transduction. Curr. Opin. Chem. Biol. **1**, 227–234
13 Zeng, L. and Zhou, M. M. (2002) Bromodomain: an acetyl-lysine binding domain. FEBS Lett. **513**, 124–128
14 Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C. and Kouzarides, T. (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature (London) **410**, 120–124
15 Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. and Jenuwein, T. (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature (London) **410**, 116–120
16 Bedford, M. T., Frankel, A., Yaffe, M. B., Clarke, S., Leder, P. and Richard, S. (2000) Arginine methylation inhibits the binding of proline-rich ligands to Src homology 3, but not WW, domains. J. Biol. Chem. **275**, 16030–16036
17 Friesen, W. J., Massenet, S., Paushkin, S., Wyce, A. and Dreyfuss, G. (2001) SMN, the product of the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets. Mol. Cell **7**, 1111–1117
18 Chen, H. I. and Sudol, M. (1995) The WW domain of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology 3-binding modules. Proc. Natl. Acad. Sci. U.S.A. **92**, 7819–7823
19 Bedford, M. T., Sarbassova, D., Xu, J., Leder, P. and Yaffe, M. B. (2000) A novel pro-Arg motif recognized by WW domains. J. Biol. Chem. **275**, 10359–10369
20 Ren, R., Mayer, B. J., Cicchetti, P. and Baltimore, D. (1993) Identification of a ten-amino acid proline-rich SH3 binding site. Science **259**, 1157–1161
21 Yaffe, M. B. (2002) How do 14-3-3 proteins work? – Gatekeeper phosphorylation and the molecular anvil hypothesis. FEBS Lett. **513**, 53–57
22 Durocher, D. and Jackson, S. P. (2002) The FHA domain. FEBS Lett. **513**, 58–66
23 Harris, B. Z. and Lim, W. A. (2001) Mechanism and role of PDZ domains in signaling complex assembly. J. Cell Sci. **114**, 3219–3231
24 Lemmon, M. A. and Ferguson, K. M. (2000) Signal-dependent membrane targeting by pleckstrin homology (PH) domains. Biochem. J. **350**, 1–18
25 Bedford, M. T. and Leder, P. (1999) The FF domain: a novel motif that often accompanies WW domains. Trends Biochem. Sci. **24**, 264–265
26 Bedford, M. T., Reed, R. and Leder, P. (1998) WW domain-mediated interactions reveal a spliceosome-associated protein that binds a third class of proline-rich motif: the proline glycine and methionine-rich motif. Proc. Natl. Acad. Sci. U.S.A. **95**, 10602–10607
27 Chen, T., Boisvert, F. M., Bazett-Jones, D. P. and Richard, S. (1999) A role for the GSG domain in localizing Sam68 to novel nuclear structures in cancer cell lines. Mol. Biol. Cell **10**, 3015–3033
28 Gee, S. H., Quenneville, S., Lombardo, C. R. and Chabot, J. (2000) Single-amino acid substitutions alter the specificity and affinity of PDZ domains for their ligands. Biochemistry **39**, 14638–14646
29 Richard, S., Yu, D., Blumer, K. J., Hausladen, D., Olszowy, M. W., Connelly, P. A. and Shaw, A. S. (1995) Association of p62, a multifunctional SH2- and SH3-domain-binding protein, with src family tyrosine kinases, Grb2, and phospholipase C gamma-1. Mol. Cell. Biol. **15**, 186–197
30 Spector, D. L. (1993) Nuclear organization of pre-mRNA processing. Curr. Opin. Cell Biol. **5**, 442–447
31 Craggs, G., Finan, P. M., Lawson, D., Wingfield, J., Perera, T., Gadher, S., Totty, N. F. and Kellie, S. (2001) A nuclear SH3 domain-binding protein that colocalizes with mRNA splicing factors and intermediate filament-containing perinuclear networks. J. Biol. Chem. **276**, 30552–30560
32 Ko, T. K., Kelly, E. and Pines, J. (2001) CrkRS: a novel conserved Cdc2-related protein kinase that colocalises with SC35 speckles. J. Cell Sci. **114**, 2591–2603
33 Cicchetti, P., Mayer, B. J., Thiel, G. and Baltimore, D. (1992) Identification of a protein that binds to the SH3 region of Abl and is similar to Bcr and GAP-rho. Science **257**, 803–806
34 Chan, D. C., Bedford, M. T. and Leder, P. (1996) Formin binding proteins bear WWP/WW domains that bind proline-rich peptides and functionally resemble SH3 domains. EMBO J. **15**, 1045–1054
35 Bedford, M. T., Chan, D. C. and Leder, P. (1997) FBP WW domains and the Abl SH3 domain bind to a specific class of proline-rich ligands. EMBO J. **16**, 2376–2383
36 Nollau, P. and Mayer, B. J. (2001) Profiling the global tyrosine phosphorylation state by Src homology 2 domain binding. Proc. Natl. Acad. Sci. U.S.A. **98**, 13531–13536
37 Yaffe, M. B., Leparc, G. G., Lai, J., Obata, T., Volinia, S. and Cantley, L. C. (2001) A motif-based profile scanning approach for genome-wide prediction of signaling pathways. Nat. Biotechnol. **19**, 348–353

Received 4 June 2002/2 July 2002; accepted 23 July 2002
Published as BJ Immediate Publication 23 July 2002, DOI 10.1042/BJ20020860