CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Elsevier - Publisher Connector

ELSEVIER
Chemistry & Biology 8 (2001) 951–965
Chemistry & Biology
www.elsevier.com/locate/chembiol

Research Paper

A new plasmid display technology for the in vitro selection of functional phenotype–genotype linked proteins

Robert E. Speight [a, 1], Darren J. Hart [a], John D. Sutherland [b], Jonathan M. Blackburn [a, *]

[a] Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 2GA, UK
[b] Department of Chemistry, University of Manchester, Oxford Road, Manchester M13 9PL, UK

Received 15 March 2001; revisions requested 18 May 2001; revisions received 11 July 2001; accepted 12 July 2001
First published online 10 August 2001

Abstract

Background: Display technologies which allow peptides or proteins to be physically associated with the encoding DNA are central to procedures which involve screening of protein libraries in vitro for new or altered function. Here we describe a new system designed specifically for the display of libraries of diverse, functional proteins which utilises the DNA binding protein nuclear factor κB (NF-κB) p50 to establish a phenotype–genotype link between the displayed protein and the encoding gene.

Results: A range of model fusion proteins to either the amino- or carboxy-terminus of NF-κB p50 have been constructed and shown to retain the picomolar affinity and DNA specificity of wild-type NF-κB p50. Through use of an optimal combination of binding buffer and DNA target sequence, the half-life of p50–DNA complexes could be increased to over 47 h, enabling the competitive selection of a variety of protein–plasmid complexes with enrichment factors of up to 6000-fold per round. The p50-based plasmid display system was used to enrich a maltose binding protein complex to homogeneity in only three rounds from a binary mixture with a starting ratio of 1:10^8 and to enrich to near homogeneity a single functional protein from a phenotype–genotype linked *Escherichia coli* genomic library using in vitro functional selections.

Conclusions: A new display technology is described which addresses the challenge of functional protein display. The results demonstrate that plasmid display is sufficiently sensitive to select a functional protein from large libraries and that it therefore represents a useful addition to the repertoire of display technologies. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Maltose binding protein; Nuclear factor κB; p50; Protein display

1. Introduction

A prerequisite for in vitro selections from large libraries of proteins is the ability to subsequently identify those members of the library which have the desired properties, yet direct analysis of the selected protein requires much larger amounts of material than are typically recovered in such experiments. One solution to this problem involves the creation of a physical association between the encoding gene and the functional protein which is maintained throughout the selection process such that the selected protein can be amplified and characterised via the encoding DNA or RNA. For such a process to be viable, a number of methodological issues need to be addressed. Firstly, the association of gene and protein must be stable to avoid either complete loss of the genetic information or false positive selection through scrambling of the phenotype–genotype link. Secondly, the protein must maintain its native activity when brought into association with the encoding gene and there are degrees of complexity here depending on the nature of the polypeptide being displayed. Thus, whilst the display of short peptides is generally straightforward, the display of functional proteins is rather more challenging due to individual folding and possible multimerisation requirements and often needs to be optimised for each case. The display of libraries of unrelated proteins is thus a multifaceted problem yet the cre-

[1] Present address: Department of Chemistry, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JJ, UK.

* Corresponding author.
E-mail address: jmb50@cam.ac.uk (J.M. Blackburn).

1074-5521/01/$ – see front matter © 2001 Elsevier Science Ltd. All rights reserved.
PII: S1074-5521(01)00066-7

ation of phenotype–genotype linked cDNA or genomic libraries is an area with considerable potential in functional genomics/proteomics and so is the subject of much current interest.

There are currently a number of technologies available for creating phenotype–genotype linkages, each of which has much merit, yet none of which is suitable for every application. The most widely used methodology at present is display on the surface of filamentous bacteriophage through fusion to the pIII or pVIII coat proteins [1]. Such systems have been widely used to great effect in the display of antibody and peptide libraries, yet examples involving libraries constructed from other classes of proteins, particularly proteins which are not normally secreted, remain relatively scarce. This perhaps reflects difficulties associated with the obligatory secretion of any fusion proteins to the relatively oxidising periplasmic space prior to phage assembly, since incorrect folding or proteolysis can often occur during this process, particularly when proteins which are not normally secreted are involved. Display on the surface of bacteria falls into a broadly similar category as phage display since it is subject to many of the same constraints, but here too there have been spectacular successes, most notably the recently reported directed evolution of an antibody with a femtomolar dissociation constant [2]. Other display systems which avoid the need for secretion have been developed more recently based upon λ [3] and T7 phage [4] and some impressive results have been obtained [5] but these have yet to receive widespread use, perhaps due to the increased complexity of these systems compared to filamentous phage.

Ribosome display [6] and oil–water emulsions [7] offer a number of attractions compared to other display methods since they intrinsically avoid the need for transformation of bacteria, thereby allowing much larger library sizes to be accessed. There have been some high profile successes, particularly with ribosome display [8], but in general, as with phage display, most examples here have involved short peptides or antibodies and examples of display of usually cytoplasmic proteins are still limited. It remains to be seen whether technical challenges such as accurately controlling the oxidation potential of the environment during in vitro translation will affect the fraction of functional proteins that can be obtained in a library format using entirely cell-free display technologies; this will be particularly relevant where the individual members of the library are from a range of different protein families, as would be the case in a cDNA or genomic library, and therefore potentially have quite different folding requirements.

Plasmid display is a conceptually simple alternative approach to the methods described above and neatly avoids potential difficulties with secretion, in vitro translation, or RNA stability by fusing polypeptides directly to a DNA binding protein. The fusion proteins are expressed and folded in the relatively reducing environment of the *Escherichia coli* cytoplasm and the requisite phenotype–genotype linkage is created by the fusion proteins binding in vivo to DNA sequences on the encoding plasmids whilst still compartmentalised from other members of the library. Providing that the specific protein–DNA interaction is maintained upon cell lysis, in vitro selection from a protein library can be carried out, after which the plasmid DNA encoding the enriched members of the protein library can be recovered and used to directly re-transform fresh cells prior to characterisation or further rounds of selection.

Plasmid display was first validated experimentally by Cull et al. who fused short peptides to the carboxyl-terminus of the *lac* repressor protein [9], but that original system apparently suffered from difficulties encountered in controlling both the soluble expression of fusion proteins and the polyvalency of the display. A simplified system using an evolved, monomeric *lac* 'headpiece dimer' was subsequently developed [10] but to date this modified system has only been validated for the display of short peptides and seemingly relies heavily on polyvalent, non-specific and relatively low affinity binding of the headpiece dimer to the plasmid DNA. Thus, although the basic concept has been demonstrated, the full potential of plasmid display to provide phenotype–genotype links for a diverse repertoire of functional proteins has yet to be realised.

Here we describe our efforts to extend the general concept of plasmid display using an alternate DNA binding protein, the nuclear factor κB (NF-κB) p50 homodimer (termed p50 herein), such that the basic system becomes compatible with display of highly diverse libraries of proteins. The key areas in which our new system appears to differ from the earlier *lac* repressor system are the intrinsic solubility of the DNA binding component of the system, the stability and specificity of the protein–DNA complexes formed, and the ability to make viable fusions to either terminus of NF-κB p50. We report the quantitative characterisation of the DNA binding properties of model p50 fusion proteins, demonstrate the p50-based plasmid display system through in vitro selections from model protein mixtures, and use this technology to isolate and identify a protein from a genomic library via a functional selection.

2. Results

2.1. *In vitro DNA binding affinity of p50 fusion proteins*

The success of plasmid display is critically linked to the maintenance of the protein–DNA interaction during the selection procedure (Fig. 1). Although it has previously been reported that native p50 binds to specific DNA sequences with a picomolar dissociation constant [11–13], we felt it necessary to confirm that this high affinity was not perturbed by the fusion of proteins to one or another

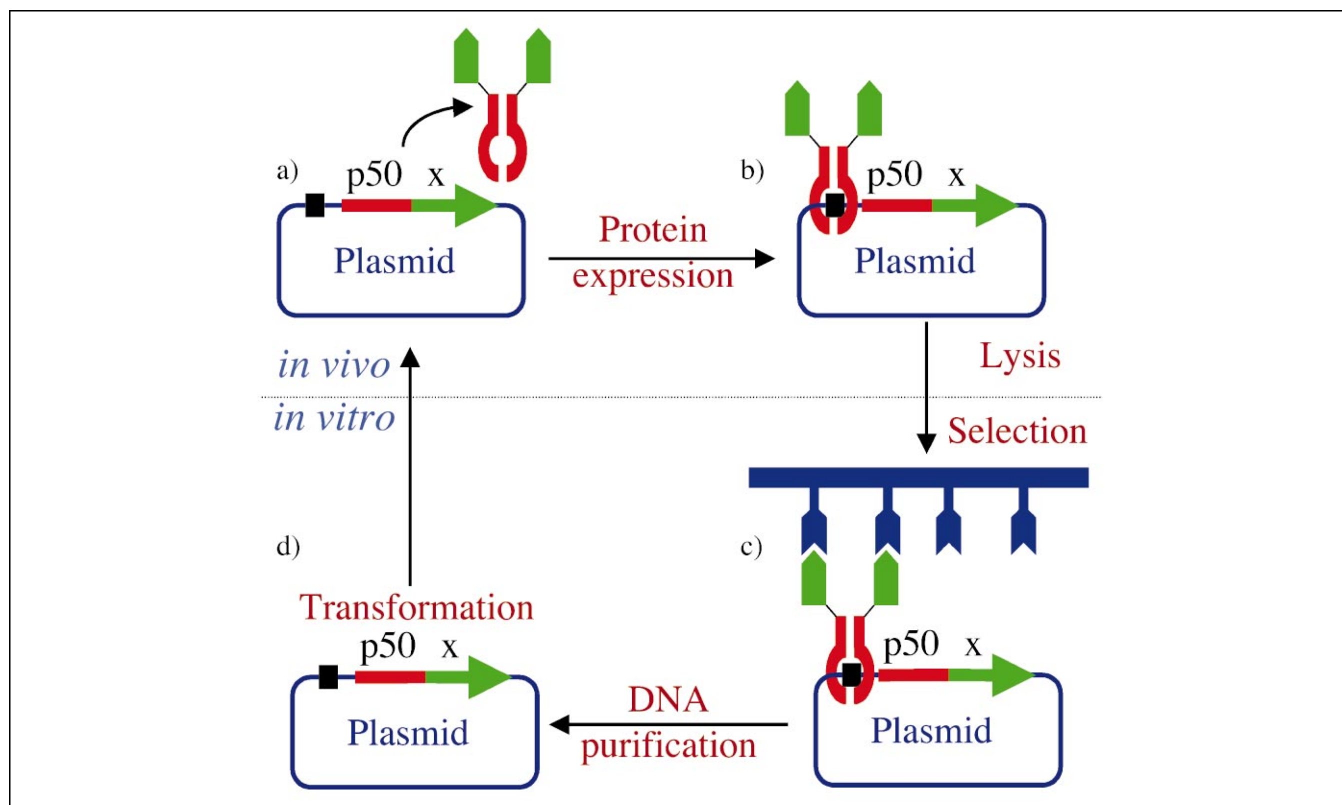


Fig. 1. Schematic representation of one cycle of plasmid display enrichment. a: Fusion proteins are synthesised in the cytoplasm of *E. coli*. b: These bind to the p50 recognition sequence on the plasmid. c: Upon cell lysis, fusion protein–plasmid complexes may be selected, based on the p50 fusion partner binding to an immobilised ligand. d: Once non-bound complexes have been removed by extensive washing, the plasmids are purified and used to transform bacteria for further cycles of enrichment or identification.

terminus of p50. The plasmid vectors pRES101 and pRES201 were constructed to contain 5′- or 3′-genetic fusions of the human *p50* gene to the *malE* gene. The *p50* gene encoded residues 40–366 of the wild-type p50 protein whilst *malE* encoded the *E. coli* maltose binding protein (MBP); the individual domains of the fusion proteins were connected by a 10 residue polyasparagine linker peptide and the resultant fusion proteins were denoted MBP–p50 and p50–MBP respectively. The vector pRES101 was a derivative of pMAL-c2 (New England Biolabs) and the fusion gene was expressed under the control of an inducible *tac* promoter. The vector pRES201 was derived from pLM1-p50 [14] and expressed p50–MBP under the control of the T7 promoter. MBP activity was verified in each case by the purification of the p50 fusions on amylose resin (data not shown).

Electrophoretic mobility shift assays (EMSAs) were used to compare the affinity of wild-type p50, MBP–p50 and p50–MBP for the sequences *target-κB* and *Ig-κB*, which contained the specific binding sites 5′-GGG-AATTCCC-3′ and 5′-GGGACTTTCC-3′ respectively. Fig. 2 shows the EMSA and resulting Scatchard analysis for the interaction of MBP–p50 with the *target-κB* duplex. These initial experiments demonstrated that both MBP–p50 and p50–MBP fusion proteins had similar equilibrium binding constants to native p50, the absolute values obtained being consistent with previously reported values for p50 ([11–13]). However, in the context of a plasmid display system, the in vitro dissociation rate is more important than the overall affinity since dissociation must be minimised during successive washes of the affinity matrix. The association rate is of lesser importance since the initial p50 fusion–DNA interaction will occur in vivo when individual plasmid and encoded protein species are still compartmentalised. The separate association and dissociation kinetic rate constants of the protein–DNA interactions were therefore determined using surface plasmon resonance (SPR) with the aim of establishing binding conditions which maximised the in vitro stability of the protein–DNA complexes.

SPR data were acquired for p50, MBP–p50 and p50–MBP, each binding to the *target-κB*, *Ig-κB* and *PRDII* duplexes, the latter being identical to *Ig-κB* except that the specific p50 binding sequence was 5′-GGGAAA-TTCC-3′. Data analysis confirmed that both fusion proteins bound in a similar way to native p50 and also showed that in each case the dissociation rates from *target-κB* were greatly reduced compared to those from *PRDII* and *Ig-κB* (Fig. 3). The association phases of the various sensorgrams were found to be very rapid and significantly mass transport-limited so manual data processing was used to identify the limited number of data points

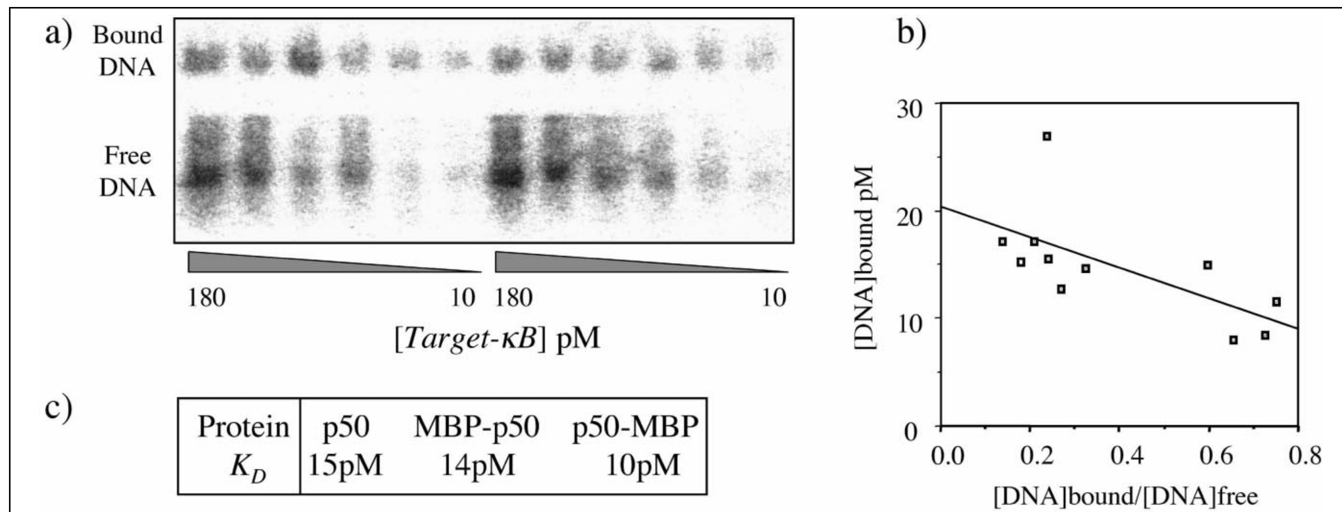


Fig. 2. Gel shift (a) and Scatchard (b) analysis of the fusion protein MBP–p50 binding to an oligonucleotide containing the sequence *target-κB*. c: Calculated equilibrium binding constants for each protein binding to the *target-κB* sequence.

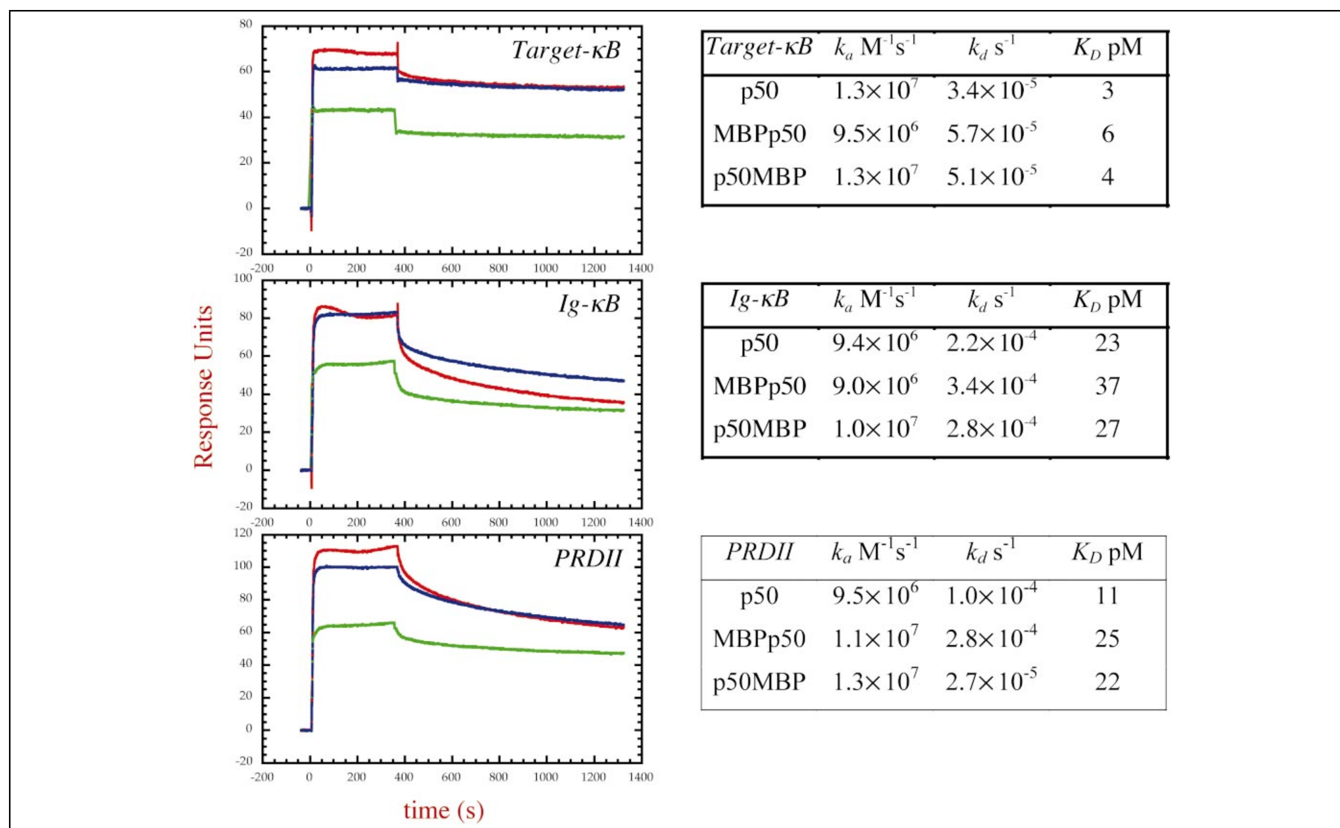
near the end of each association event which could be used with reasonable confidence in calculating the rate association constants [15]. The association rate constants shown in Fig. 3 should perhaps therefore be regarded as close approximations rather than highly accurate figures, although it is worth noting that the absolute values for K_D calculated by EMSA (Fig. 2) and SPR (Fig. 3) are close enough to suggest the SPR association rate constants are relatively accurate. No such constraints, however, applied to the dissociative phases and our dissociation model [15] fitted the data well across a 200 s period (data not shown), implying that the dissociation rate constants obtained are likely to be accurate. The half-lives of the fusion protein–*target κB* complexes were thus determined to be > 3.4 h under the SPR conditions used, immediately suggesting that specific co-purification of plasmid DNA during affinity purification of the fusion proteins should be possible.

2.2. *A phenotype–genotype linkage system incorporating a feedback repression element*

In order to create a phenotype–genotype linkage for individual fusion proteins, we chose to incorporate a single *target-κB* site into the expression plasmids. As with any display system it is preferable that each protein molecule has a genotypic linkage; as such we decided to limit protein expression levels in our plasmid display system through design of a feedback repression system which would more closely link plasmid and encoded protein copy numbers in vivo. We had previously found that insertion of a *target-κB* site at a specific position within the −10 region of the *lac* promoter did not significantly affect the intrinsic strength of the promoter yet resulted in 200-fold repression of reporter gene expression when p50 was co-expressed [16]. Based on this knowledge, a new set of expression plasmids were created in which this modified *lac* promoter was used to deliberately down-regulate the expression of the various p50 fusion proteins. Placement of the *target-κB* site in the −10 region of the promoter disrupted the existing *lac* repressor binding site, thus removing the need for IPTG induction and making p50 fusion protein expression levels dependent instead on the intracellular p50 fusion protein concentration. Levels of fusion protein expression controlled by this modified promoter were deduced to decrease significantly since the fusion proteins were only detectable by Western blot and not by Coomassie blue-stained SDS–PAGE analysis of cell lysates, whereas over-expressed proteins could readily be observed by both methods from pRES101 and pRES201 (data not shown). This seemed to indicate qualitatively that a feedback repression system had been created, although no attempt was made to quantify this effect here.

2.3. *Model systems – comparative enrichments*

To demonstrate effective functional protein display in a p50-based system, the selective enrichment of one fusion protein–plasmid complex over another was studied. The fusion proteins glutathione *S*-transferase (GST)–p50, p50–MBP and p50–His_6 were chosen as model systems since the appropriate affinity matrices were well-characterised and readily available. GST was originally derived from the plasmid pGEX-2T [17] whilst His_6 denotes a hexahistidine peptide. The *gst–p50*, *p50–malE* and *p50–his*$_6$ genes were expressed from the pUC-derived plasmids pRES108, pRES306 and pRES314 respectively under the control of the feedback repression-modified *lac* promoter; pRES108 and pRES314 conferred resistance to chloramphenicol whilst pRES306 conferred resistance to ampicillin. Recovery of plasmid DNA pre- or post-selection, transformation of *E. coli* cells and plating identical ali-



Target-κB	k_a M^{-1}s^{-1}	k_d s^{-1}	K_D pM
p50	1.3×10^{7}	3.4×10^{-5}	3
MBPp50	9.5×10^{6}	5.7×10^{-5}	6
p50MBP	1.3×10^{7}	5.1×10^{-5}	4

Ig-κB	k_a M^{-1}s^{-1}	k_d s^{-1}	K_D pM
p50	9.4×10^{6}	2.2×10^{-4}	23
MBPp50	9.0×10^{6}	3.4×10^{-4}	37
p50MBP	1.0×10^{7}	2.8×10^{-4}	27

PRDII	k_a M^{-1}s^{-1}	k_d s^{-1}	K_D pM
p50	9.5×10^{6}	1.0×10^{-4}	11
MBPp50	1.1×10^{7}	2.8×10^{-4}	25
p50MBP	1.3×10^{7}	2.7×10^{-5}	22

Fig. 3. Sensorgrams and data for the binding comparison of p50 and each MBP fusion protein to the three recognition sequences in buffer containing 100 mM KCl. Traces for p50 (green), MBP–p50 (red) and p50–MBP (blue) along with the data show that the fusion proteins have a similar but slightly reduced affinity for each sequence compared with p50 and that reduced dissociation rates are observed in all cases from the sequence *target-κB*.

quots of pre- or post-selection transformants on LB plates containing either ampicillin or chloramphenicol thus allowed the relative level of enrichment of one specific plasmid–protein complex over another to be assessed by the simple process of counting colonies.

In order to perform the enrichment experiments, two cultures, each expressing a different fusion protein, were grown to OD_{600} (optical density at 600 nm) = 0.4 and 10 ml of the competing culture was mixed with various dilutions of cells containing the protein–plasmid complex to be selected. In order to maximise the stability of the protein–plasmid complexes, the following gentle procedure was used to lyse the cells: spheroplasts were prepared from the cell mixtures and burst by freeze–thaw/osmotic shock in water and the resultant lysate immediately diluted into the enrichment buffer, after which cellular debris was removed by centrifugation. The buffer in these enrichment experiments was the same as for the SPR binding studies except that, based on the SPR data, the KCl concentration was reduced to 75 mM in order to minimise dissociation of the p50 fusion–plasmid complex during selection. In addition, sonicated herring sperm DNA was included as a non-specific competitor for any excess, non-DNA-bound or dissociating p50 fusion proteins that may have been present. The lysate was then applied to the relevant affinity matrix and washed extensively with binding buffer. Protein–plasmid complexes that remained bound to the matrix were then eluted either by addition of the appropriate soluble affinity ligand or by a high salt wash, after which plasmids were purified and used to transform *E. coli* cells. Enrichment factors were calculated from the relative numbers of ampicillin or chloramphenicol resistant colonies pre- and post-selection.

A number of experiments were performed in which p50–MBP complexes were enriched from an excess of GST–p50 complexes on amylose resin, GST–p50 complexes were enriched from p50–MBP complexes on glutathione Sepharose, and p50–His_6 complexes were enriched from p50–MBP complexes on a Co^{2+}-NTA affinity resin. Enrichment factors as high as 5900-fold, together with reproducible yields, were achieved from increasing dilutions (Table 1) and whilst the best results were obtained when enriching for amylose binding, relatively high yields and enrichment factors were consistently observed in each case.

2.4. Library simulation

In order to test the sensitivity of the system and to quantify the enrichment factors that could be obtained

Table 1
Comparative enrichments of three different fusion proteins from binary mixtures

Pre-enrichment	Post-enrichment	Enrichment factor	Yield (%)
Ratio of p50–MBP to GST–p50			
1:1.4	4200:1	5900	6.3
1:22	220:1	4800	4.8
1:330	4:1	1300	17.4
Ratio of GST–p50 to p50–MBP			
1:300	1.2:1	375	2.6
1:1300	0.2:1	260	2.2
Ratio of p50–His_6 to p50–MBP			
1:46	1.5:1	69	1.9
1:247	1:1	247	2.6
1:31300	0.015:1	470	2.8

Comparative enrichments of three different fusion proteins showing the ratios of the relevant fusion protein producing plasmids before (Pre) and after (Post) enrichment, the enrichment factors (the 'Post' ratio divided by the 'Pre' ratio) and overall yield in terms of the number of colonies counted after transformation of *E. coli* with the post-enrichment fraction compared to the pre-enrichment fraction.

under more challenging circumstances, high dilutions of cells expressing MBP–p50 fusion proteins in an excess of cells expressing the GST–p50 equivalent were prepared to simulate the selection of a single member from a library. Since multiple rounds of selection were now necessary to enrich selected plasmid species to detectable levels, a number of subtle modifications to the model system were carried out: Plasmids named pRES109 and pRES110, encoding MBP–p50 and GST–p50 respectively, were designed such that the fusion-encoding genes were expressed under the control of the p50-repressible *lac* promoter as before, but now both plasmids encoded the ampicillin resistance marker to enable growth of all colonies from each round of selection on the same agar plates; the size of the vector backbone in each case was minimised to increase bacterial transformation efficiencies since we reasoned that plasmid recovery might be a critical parameter at high dilution; and plate-based growth rather than liquid culture was used at each stage of the enrichment process to reduce the potential bias arising from differential growth rates of cells containing different plasmids. In addition, it had previously been reported that use of potassium glutamate in place of potassium chloride in a DNA binding buffer had increased the affinity of certain protein–DNA interactions [18]. We therefore determined the effect of glutamate on the specific p50–DNA interaction by SPR, the study being performed at salt concentrations of 125 mM to enable us to observe variations in protein–plasmid dissociation rates between the different salts. The results showed that whilst there was not a dramatic effect on overall stability of the p50–*target-κB* complexes, the off-rate was reduced by ~20-fold in potassium glutamate compared to the same concentration of potassium chloride (Table 2); the absolute dissociation rate for the p50–*target-κB* complex measured at 125 mM potassium glutamate corresponded to a half-life of 47 h. Extrapolating from this and earlier SPR data [15], we therefore chose to use 50 mM potassium glutamate in place of 75 mM potassium chloride in subsequent enrichment experiments in order to maximise the stability of the plasmid–p50 complexes during selection.

Initially we simulated a modest library by using a 1 in 10^5 dilution of cells expressing MBP–p50 in a GST–p50 background. We carried out three rounds of enrichment and analysed a subset of the population by PCR colony screening using vector-specific primers to amplify either the *malE* or *gst* genes. The different sizes of the two possible PCR products allowed the relative ratio of plasmids in the enriched mixtures to be determined and showed that the plasmid encoding the MBP–p50 fusion was in significant excess after round 3 (Table 3). It is interesting to note that as the positively selected complex became dominant

Table 2
Comparison of the binding kinetics of p50 with the DNA binding motifs *target-κB*, *Ig-κB* and *PRDII* in the presence of 125 mM potassium chloride (KCl) or 125 mM potassium glutamate (KGlu)

125 mM KCl	k_a (M^{-1} s^{-1})	k_d (s^{-1})	$t_{1/2}$ (h)	K_D (pM)
Target-κB	1.3×10^7	9.6×10^{-5}	2.0	7
Ig-κB	9.0×10^6	4.7×10^{-4}	0.4	52
PRDII	8.0×10^6	6.5×10^{-4}	0.3	81
125 mM KGlu				
Target-κB	1.0×10^6	4.1×10^{-6}	47.0	4
Ig-κB	2.0×10^6	2.1×10^{-5}	9.2	11
PRDII	1.0×10^6	1.3×10^{-5}	14.8	13

Data calculated from SPR experiments for the comparison of the binding kinetics of p50 with the DNA binding motifs *target-κB*, *Ig-κB* and *PRDII* in the presence of 125 mM potassium chloride (KCl) or 125 mM potassium glutamate (KGlu).

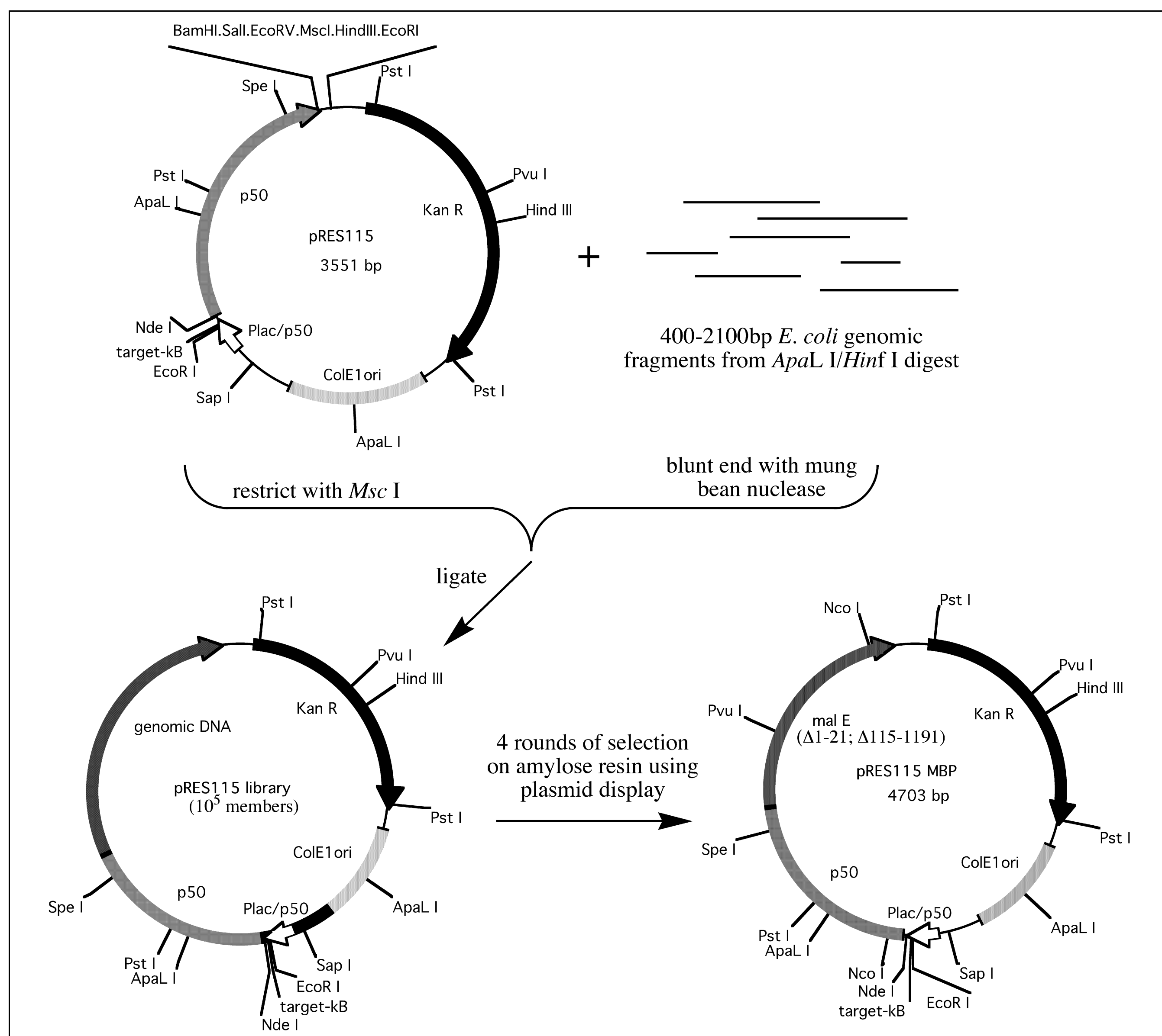


Fig. 4. Schematic representation of the construction of a directed *E. coli* genomic library and selection of an individual member of that library based on function.

in successive rounds of selection, the absolute plasmid yields greatly increased accordingly, as might be expected.

In a subsequent experiment, we simulated a library as large as transformation efficiencies of *E. coli* typically allow by preparing a 1 in 10^8 dilution of MBP–p50-expressing cells in a GST–p50 background, corresponding to 1.3 nl (approximately 600 cells) of MBP–p50 culture in 92 ml of GST–p50 producing culture. Iterative rounds of enrichment and analysis were carried out and after round 3, the ratio of MBP–p50 to GST–p50-expressing cells was found to be approximately 1:1 (Table 3) whilst a fourth round of enrichment resulted in the MBP–p50 clones becoming the major component of the mixture. The results of these library simulation experiments strongly suggested that should an affinity-based selection exist for a protein then this plasmid display system would be sufficiently sensitive and selective to isolate a single member from a library of at least 10^8 different clones based on protein function.

2.5. Library screening

The previous experiments simulated enrichment from a library by dilution of one fusion protein in an excess of another. To extend the validation of our system further we created a true library by fusing *E. coli* genomic DNA fragments to the 3′-terminus of the *p50* gene. To facilitate library work, plasmid pRES115 was constructed from the library simulation vectors such that it contained a multiple cloning site at the 3′-terminus of the *p50* gene in place of the stop codon or any fusion gene (Fig. 4). The resistance

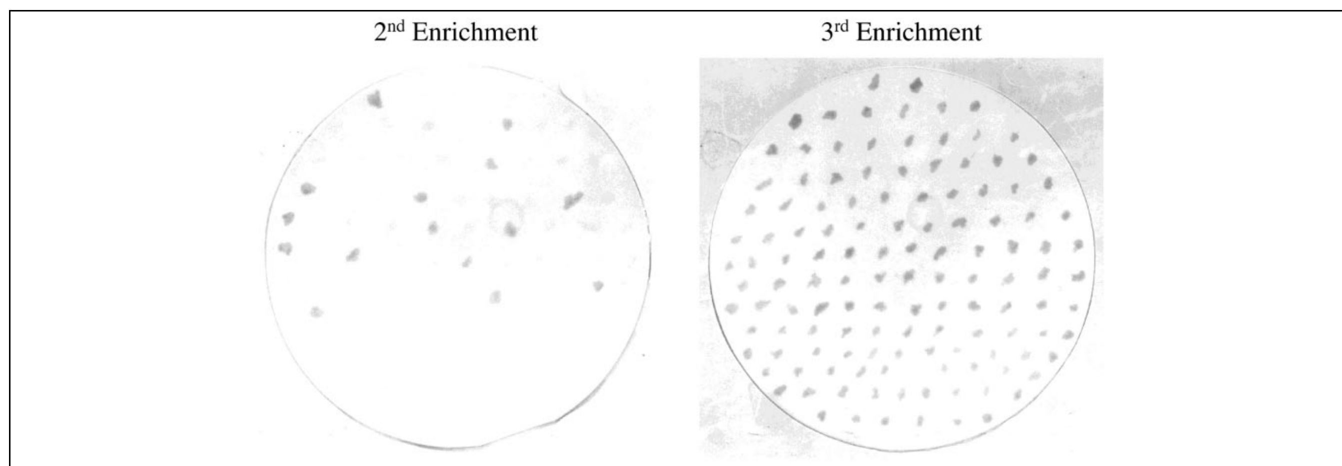


Fig. 5. Western blot analysis of selected colonies from the second and third rounds of enrichment of the *E. coli* genomic library. Each spot corresponds to a single colony that contained detectable levels of MBP. Native expression of chromosomal MBP was not sufficient to be detectable, as shown by the negative colonies where the p50–MBP fusion was presumably absent.

marker in pRES115 was changed to kanamycin to avoid the risk of selecting contaminating fusion proteins which had previously been cloned into the plasmid display system during the library simulation work.

Initially we considered using a restriction endonuclease with a 4 bp recognition sequence to fragment the *E. coli* genome but we decided instead to adopt a more directed approach towards the selection of a chromosomally derived p50–MBP–plasmid complex which would allow us to work with a library that could realistically be fully sampled during in vitro selections. Analysis of the nucleotide sequence of the *E. coli* genome [19] allowed us to identify the restriction enzymes *Hin*fI and *Apa*LI which would together digest the genome into approximately 200–3000 bp fragments. Within the *E. coli* genome there are 10 743 *Hin*fI sites and 575 *Apa*LI sites, equating to a predicted mean fragment size of 410 bp. Detailed analysis of the genomic region around the *malE* gene confirmed that *Hin*fI would digest significantly less frequently than every 256 bp, despite having a recognition sequence of only four specific bases. Complete digestion of the *E. coli* genome by a combination of *Hin*fI and *Apa*LI was therefore predicted to generate, amongst many others, a 1155 bp chromosomal *malE* gene fragment, encoding a MBP truncated by seven residues at the N-terminus and five residues at the C-terminus, which could be cloned in-frame with the 3′-end of the p50 gene in pRES115 as one member of the genomic library.

DNA fragments between 400 and 2100 bp produced by over-digestion of *E. coli* genomic DNA with *Hin*fI and *Apa*LI were purified and the single-stranded overhangs removed with mung bean nuclease. The genomic fragments were then ligated into the unique *Msc*I restriction site in the multiple cloning site of pRES115 (Fig. 4), yielding a library of 110 000 members; PCR screens on clones picked at random from the library showed a good insert size distribution within the expected range (data not shown). Theoretical analysis showed that a complete *Hin*fI/*Apa*LI digestion would result in 4142 fragments in the size range 400–2100 bp so this represents the minimum number of unique inserts likely to be present in the genomic library created here. However, the true size of the library here may be significantly larger than 4142 since it seems unlikely that the *Hin*fI/*Apa*LI digestion of the genomic DNA would have gone absolutely to completion, making the total number of fragments in this size range resulting from complete and partial digests difficult to estimate. Assuming though a theoretical population of 4142 members, statistically there is a 99.99% probability that every member of the library will be represented in a physical library of 110 000 clones (using $p=1-V\times\exp(-N/V)$, where N = sample size, V = library size), thus ensuring that each possible clone within the 400–2100 bp size range would be sampled in subsequent in vitro selections.

We prepared protein–plasmid complexes from cells transformed by the library as before and carried out iter-

Table 3
The total number of colonies and the proportion which contained the plasmid encoding MBP–p50 in each round of selection from the library simulations

Enrichment round	Number of recovered colonies	Total colonies encoding MBP–p50 (%)
(a) Initial dilution $1:10^5$		
1	1 930 000	none detected
2	174 500	2
3	483 000	92
(b) Initial dilution $1:10^8$		
1	19 500	none detected
2	6 000	none detected
3	14 000	58
4	648 000	87

The total number of colonies and the proportion which contained the plasmid encoding MBP–p50 in each round of selection in (a) the 1 in 10^5 dilution experiment and (b) the 1 in 10^8 dilution experiment.

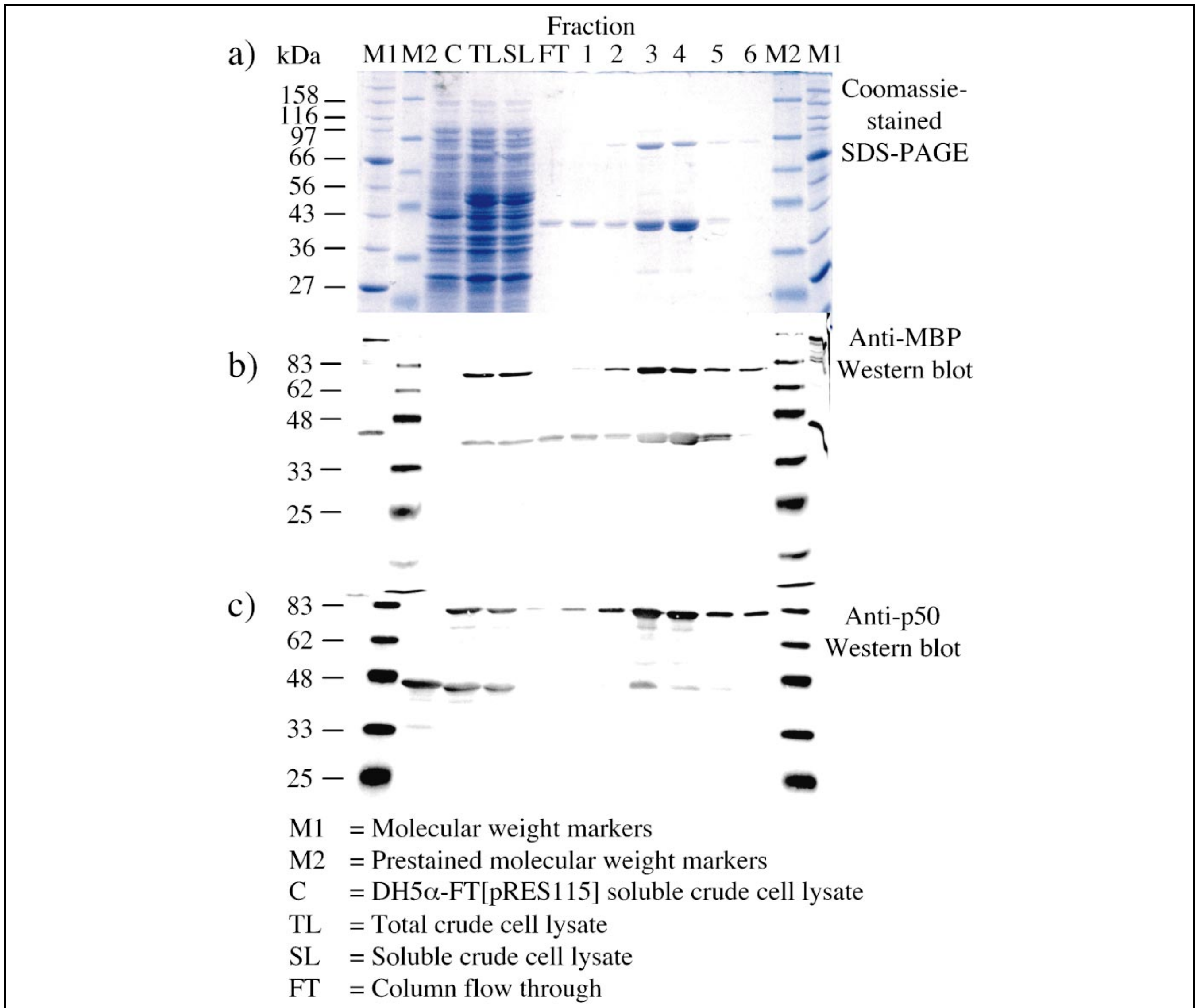


Fig. 6. SDS–PAGE analysis of the purification of a selected library clone on an amylose-agarose column visualised by (a) Coomassie stain, (b) anti-MBP Western blot and (c) anti-p50 Western blot. Although some proteolysis of the fusion protein is observed, the full-length protein is also visible in each case at the calculated mass of 81 633 Da

ative rounds of selection on the plasmid-displayed genomic library, using amylose resin to enrich members of the library that bound this matrix. PCR screening using primers that bound either side of the *Msc*I cloning site showed a good distribution of genomic insert sizes after the first and second rounds of enrichment (data not shown). Following the third round of selection, over 98% of the clones contained an insert of ~1200 bp, the expected size for the *Hin*fI/*Apa*LI *malE* fragment (data not shown). The likely identity of this insert was confirmed by PCR screening of individual and pooled third round plasmids with a vector/*malE*-specific primer pair which generated the predicted 1245 bp fragment in each case (data not shown). To characterise the second and third round populations further, colony Western blots were performed with an anti-MBP antibody (Fig. 5) and this showed that approximately 20% of the colonies were positive for MBP after the second round whilst over 98% of the colonies were MBP-positive after the third round of enrichment. Plasmids from three positive clones from round 3 were sequenced and all were found to be identical, each containing the precise chromosomal *malE* gene fragment predicted to arise from *Hin*fI and *Apa*LI digestion of the genomic DNA.

To demonstrate that the selected clones expressed protein with the desired amylose binding property, a single clone was grown overnight in 100 ml liquid culture, cells were harvested and lysed, crude protein was applied to an amylose resin, and specifically bound protein eluted with a maltose solution. The purification profile was analysed by SDS–PAGE and Western blot using anti-MBP and anti-p50 antibodies, confirming that the selected clone ex-

pressed an anti-MBP reactive NF-κB p50 fusion protein possessing amylose resin binding activity (Fig. 6).

3. Discussion

The burgeoning fields of proteomics and directed evolution have the potential to benefit greatly through access to robust, massively parallel experimental methods for identifying proteins based on function. In vitro selection technologies such as phage display have begun to fulfil this role but a significant hurdle to wider exploitation remains the creation of a phenotype–genotype link for functional proteins, and in particular for libraries of diverse proteins. Here we describe the creation and validation of a new in vitro display technology in which proteins are physically associated with their encoding genes via fusion to the DNA binding protein NF-κB p50, which itself binds to a specific site on the parental plasmid with high affinity. In this system, the direct protein–DNA association occurs in the cytoplasm of the host organism, thus avoiding the potentially problematic need for export to the periplasmic space which is obligatory in filamentous phage display and bacterial surface display.

In designing a plasmid display system, a key consideration is that the integrity of the protein–DNA interaction must be maintained during selection processes since any dissociation of the protein–plasmid complex after cell lysis could lead to scrambling or loss of the phenotype–genotype link. The DNA binding protein component of the system therefore needs to have an intrinsically high binding affinity for specific DNA sequences, whilst being able to tolerate fusion of a wide variety of proteins to one or another terminus without compromising that DNA binding affinity. Clearly it is also essential that the DNA binding proteins, and any fusions there-to, are expressed in a soluble form such that they are capable of binding DNA in vivo.

We decided to base our de novo plasmid display system around the eukaryotic transcription factor, NF-κB p50, for a number of reasons. The crystal structures of human and murine p50 homodimer bound to DNA had been determined at high resolution [20,21] and revealed that in both cases, the N- and C-terminal regions of p50 were disordered with respect to the DNA binding *rel* homology domain. In the case of the recombinant murine p50 protein, only residues 39–364 comprising the core DNA binding domain were expressed and crystallised, implying that the N- or C-terminal polypeptides beyond these limits were not critical for DNA binding activity. Closer inspection of the structure of the human protein showed that Pro-43 and Glu-353 effectively delimit the human p50 DNA binding domain and are essentially in free solution with respect to the core structure. In addition, human p50 (residues 2–366) had been highly expressed in *E. coli* as a soluble protein and, as a final point, NF-κB p50 had been reported to bind to DNA with picomolar affinity, several orders of magnitude tighter than most other DNA binding proteins. Taken together, these observations suggested that fusion of alternative polypeptide chains to the N- or C-termini of the core human p50 DNA binding domain might be tolerated without affecting DNA binding affinity and that the intrinsic solubility of the DNA binding protein might even help to solubilise fusion proteins which otherwise had tendencies to aggregate. NF-κB p50 thus seemed a prime choice to fulfil the DNA binding role in our putative plasmid display system.

In order to validate this choice experimentally, we used EMSA to show that NF-κB p50 did indeed tolerate fusions to either the amino- or carboxyl-terminus without significant effect on affinity or specificity of DNA binding, the fusion proteins retaining picomolar dissociation constants in all cases. These EMSA studies were carried out in a buffer which had previously been shown to promote high affinity DNA binding and low equilibrium dissociation constants for the p50–DNA interaction [11]. However, for the purposes of a plasmid display system, we thought the equilibrium position less important than the absolute dissociation rates of the protein–DNA complexes. We therefore determined the individual rate constants for the p50–DNA interaction under a variety of conditions by SPR and observed significant stabilisation of the specific protein–DNA complexes at lower salt concentrations and also observed that replacement of chloride by glutamate in the binding buffers resulted in further increases in stability. Using this information, we were able to design a plasmid display buffer which minimised the dissociation rate of the protein–DNA complexes such that, by extrapolation, the half-life for dissociation of the complexes was in excess of 2 days; we thought this likely to be sufficiently stable to enable selections to be carried out on libraries of plasmid-displayed proteins without loss of the phenotype–genotype linkage.

Whilst analysing in detail the SPR data, we observed a significant variation in the binding characteristics of p50 to the three different NF-κB recognition sequences, with the p50–*target-κB* complex showing markedly slower dissociation rates than the p50–*PRDII* or p50–*Ig-κB* complexes under all conditions. The difference in stability observed for the various complexes probably reflects the fact that the palindromic *target-κB* motif consists of two optimal half-sites for p50, whereas the non-palindromic *Ig-κB* and *PRDII* motifs each contain one optimal and one suboptimal p50 half-site. In a cellular environment, the *Ig-κB* and *PRDII* motifs are usually bound within enhanceosome complexes by the commonly observed, transcriptionally active p50/p65 heterodimer, whereas the *target-κB* motif has not been observed in nature. Our kinetic data suggest a possible reason for this since if a *target-κB* motif did occur within an enhancer region, occlusion of that site by formation of highly stable complexes with the transcriptionally inactive p50 homodimer might present a signifi-

cant barrier to gene activation. For the purposes of plasmid display, however, it was clear that these same properties made the *target-κB* motif the binding site of choice to mediate the required phenotype–genotype linkage.

In the context of a plasmid display system, we thought that expression of fusion proteins at levels in excess of those required to bind the plasmids in vivo would be undesirable since either saturation of the selective affinity matrix or binding to non-self plasmids upon cell lysis might occur, either of which might interfere with selection procedures. In order to regulate protein expression, we therefore devised a negative feedback repression system which exploited the DNA binding activity of the p50 domain of a given fusion protein such that it repressed its own expression. Incorporation of the *target-κB* binding site into the −10 region of the promoter driving expression of the p50 fusion proteins thus created the desired feedback repressed plasmid display system in which the stability of the p50–DNA complexes was maximised.

Putting together the individual optimised components of binding buffer, binding site, and feedback repression into a plasmid display system, we have been able to achieve single-cycle enrichment factors of up to 6000-fold with our plasmid display system, both in model studies and in a library scenario. This showed, by implication, that the quantitative data generated by SPR and EMSA translated well to the more complex plasmid display environment and, moreover, confirmed that any slow dissociation of the protein–plasmid complexes which may occur during the in vitro steps does not interfere with the functional selections carried out.

Comparative enrichments with different fusion partners demonstrated that our plasmid display system is versatile and that high enrichment factors can be obtained, given the availability of a suitable affinity-based selection. Extension of the model experiments to conditions of high dilution and very low absolute amounts of selectable material showed that the p50-based plasmid display system exhibits high enrichment factors together with selectivity since we were able to enrich a maltose binding complex in the presence of competitors from a starting ratio of 1 in 10^8 to near homogeneity in only three rounds; calculation suggested that enrichment factors of roughly 1000-fold per cycle had been achieved here, consistent with the model studies.

It could be argued that two-component model library simulations are subtly different to selections from a ‘real life’ library containing either many variants of a single protein or many diverse proteins. Therefore as a further validation of the suitability of this new plasmid display system for library screening, we carried out a functional selection for amylose binding from an *E. coli* genomic library containing 110 000 members and were able to enrich a single member of the genomic library, a truncated *malE* fragment, to near homogeneity in only three rounds of selection. We confirmed the identify of the selected clone by restriction mapping and DNA sequencing and in addition, we confirmed that the encoded p50 fusion protein had amylose binding activity and showed cross-reactivity with anti-MBP as well as anti-p50 antibodies. This established beyond doubt that the plasmid display system had indeed successfully enriched the chromosomally encoded MBP through functional selection.

The genomic library created and screened here was deliberately biased during construction through combined use of specific restriction endonucleases with 4 and 6 bp recognition sequences to ensure that an active fragment of the chromosomal *malE* gene could potentially be represented as a fusion to p50 in a library encoding ≥4100 unique inserts. It is perhaps relevant then to consider briefly the scalability of our plasmid display system since it should not be automatically assumed that selections from libraries with much greater diversity will meet with the same degree of success as selections from smaller, more targeted libraries; this is for a number of reasons which relate principally to the relative ligand binding affinities of the different proteins in the library. In model two-component libraries, the relative affinities of the two proteins that are each competing for a given immobilised ligand are likely to be different by many orders of magnitude so the selection efficiency should in principle be high. In much larger libraries composed of unrelated proteins, the vast majority of the proteins are still likely to have little or no affinity for the specific immobilised ligand so selection of high affinity ligand binders should again be relatively efficient although it is always possible that more than one protein in the library will bind the ligand tightly – in other words, you will get what you select for. Selections from libraries composed of many variants on a single protein scaffold perhaps represent the most stringent type since as the size of library increases, the number of individual variants with similar ligand binding affinities is also likely to increase; this has the effect of making selection of the highest affinity binder in the presence of other, lower affinity binders more difficult and is further complicated by any avidity affects. An additional consideration is that in all in vitro selections, there will presumably be some minimum ligand binding affinity below which it is not possible to carry out enrichments because the signal-to-noise ratio becomes too low. Even where the desired interaction affinity is above this threshold, non-specific binding interactions with the matrix or with the immobilised ligand itself could significantly reduce signal-to-noise ratios although with the right choice of solid support for the ligand it should be possible to minimise this effect [22].

Here we have shown that using our plasmid display system we were able to readily enrich an active MBP clone from a diverse, albeit relatively small, genomic library and have also demonstrated the ease with which individual members of model libraries of up to 10^8 could be enriched via functional selection for a range of protein–ligand interactions with differing affinities. Thus, notwithstanding

the difficulties in selecting from larger libraries discussed above, we believe that the data presented here clearly demonstrate that our new plasmid display system has the potential to select functional proteins from large, highly diverse libraries on the basis of activity. Indeed we have recently used this system to successfully enrich and identify individual members of a library of $>10^7$ variants of a single α/β-barrel protein using in vitro functional selections (W. Patrick, J.M.B., manuscript in preparation) and our data suggest that selections from still larger libraries should be possible.

An issue which frequently crops up in respect of all in vitro display technologies concerns the valency of the displayed proteins. In filamentous phage systems, for example, vastly differing numbers of copies of the displayed proteins can be achieved depending on whether fusion is made to gene VIII or gene III on a phage vector, or gene III on a phagemid vector and it is not uncommon to switch from polyvalent to monovalent display in later rounds of selection in order to enrich the tightest binders in a given library. In the plasmid display system described here, we have considered the possibility that, at the low salt concentrations used during selections, tight binding of p50 fusion proteins to a multitude of cryptic or non-specific sites on the plasmid might give rise to an uncharacterised polyvalent display effect. We have previously shown by SPR that p50 has nanomolar affinity for non-specific DNA sequences in buffers identical to the binding buffers used here except containing 75 mM KCl [15] and this might argue in favour of a polyvalent effect operating in this system. However, the SPR data also show that such non-specifically bound p50–DNA complexes typically have relatively rapid off-rates, with half-lives of only 70 s under those conditions [15]. It seems likely then that under the selection conditions employed here, any p50 fusion proteins bound non-specifically to plasmid DNA immediately prior to cell lysis would rapidly dissociate after cell lysis and statistically would then re-bind to the excess sonicated herring sperm DNA we deliberately included in the binding buffers for this purpose. Thus, whilst we have not specifically quantified the exact valency of display in the system described here, it seems likely that it is predominantly bivalent since there is only a single high affinity p50 binding site on the encoding plasmids and p50 binds this site as a dimer, resulting in two copies of each fusion partner being displayed per high affinity p50 site on the plasmid. We are presently constructing further variants of this system into which we are incorporating five *target-κB* sites in order that, when working with lower affinity protein–ligand interactions, we can deliberately take advantage of an avidity effect in early rounds of selection.

The observation that p50 tolerates fusion to either its N- or C-terminus without affecting DNA binding activity is significant since it is known that most proteins can tolerate a fusion to one terminus or the other, yet in a library of diverse proteins it seems unlikely that all proteins will tolerate fusions to the same terminus without affecting folding or function. A potentially important additional feature of our p50-based plasmid display system then is that it could readily be configured such that genetic fusions of libraries of diverse proteins could be made in parallel to the N- or C-termini of p50, thus theoretically increasing the chance of producing correctly folded, active proteins ready for in vitro selection. Thus, whilst not as well-established as phage display, nor allowing access to the very large library sizes of ribosome display, our plasmid display system offers some potential advantages, especially where the correct folding, and hence function, of the displayed proteins is not compatible with export to the periplasmic space of *E. coli* or with in vitro translation. In this regard it is interesting to note that whilst our system was validated in *E. coli* it could conceivably be modified so that it becomes compatible with other organisms such as yeast to promote correct folding, post-translational modification and hence functional display of eukaryotic proteins.

In conclusion, we have characterised and demonstrated the versatility and advantages of a new plasmid display system which should prove a useful addition to the current repertoire of display technologies.

4. Significance

In principle, protein display technologies have the capacity to play a major role in protein library screening, for both proteomic analysis and directed evolution experiments. Here, we have described a new approach to plasmid display in which functional proteins may be fused to either the amino- or carboxyl-terminus of the DNA binding protein NF-κB p50. We have shown by EMSA and SPR that the fusion of proteins to either terminus of NF-κB p50 does not significantly perturb the DNA binding affinity of p50, with picomolar dissociation constants being observed with specific DNA sequences in vitro. Furthermore, we have demonstrated by means of in vitro selections that the high affinity interaction of such recombinant fusion proteins in vivo with a specific binding site on the encoding plasmid is maintained after cell lysis. In model systems using fusions of NF-κB p50 to MBP, GST, or a hexahistidine peptide, enrichment factors of up to 6000-fold were obtained in a single round of selection on an appropriate affinity support. The sensitivity of this technology was demonstrated by successfully enriching a plasmid-bound MBP–NF-κB p50 fusion protein complex to near homogeneity in only four rounds of selection from an initial cellular dilution of $1{:}10^8$. In order to demonstrate the viability of the plasmid display system in a genuine library format, a single member of an *E. coli* genomic library was enriched by functional selection to near homogeneity in only three rounds of selection. The results pre-

sented here show that this new protein display methodology permits rapid and facile in vitro selection, amplification and identification of proteins based on their activity and thus has considerable potential for enabling a wide range of selections from libraries of functional proteins. We anticipate that our plasmid display system will prove to be a useful addition to the existing protein display repertoire, especially where functional display of a diverse range of proteins is required.

5. Materials and methods

5.1. Plasmid constructions

All plasmids were constructed using standard protocols and details are available on request. The characteristics of the individual plasmids that were used in this study are described in Section 2.

5.2. Protein purification for DNA binding studies

The proteins p50, MBP–p50 and p50–MBP were expressed and purified from *E. coli* cultures BL21(DE3) [pLM1-p50], NM554 [pRES101] and BL21(DE3) [pRES201] respectively as described previously for p50 [15]. The MBP fusion proteins underwent an additional affinity purification step on amylose resin (New England Biolabs), according to the manufacturer's protocol, prior to purification on a Mono S HR 16/10 column (Pharmacia).

5.3. EMSA

The complementary oligonucleotides 5′-AGCTTCAGAGGGGACTTTCCGAGAGTACTG-3′ and 5′-GATCCAGTACTCTCGGAAAGTCCCCTCTGA-3′ were annealed resulting in double-stranded probe *Ig-κB* where the 10 bp p50 binding sequence is underlined. The duplex *target-κB* was produced by annealing the oligonucleotides 5′-TGACGGCGGTGGTAGGGAATTCCCCAGGAAACAGTAAGGATC-3′ and 5′-CTTACTGTTTCCTGGGGAATTCCCTACCACCGCCGTCAACTA-3′. The duplexes were labelled with ^{32}P using T4 polynucleotide kinase and [γ-^{32}P]ATP and were purified into water using a Biospin 6 column (Bio-Rad).

Binding reactions were performed with 100 pM purified protein and the radiolabelled probe was titrated between 10 pM and 180 pM in 10 μl of buffer KB (10 mM Tris–HCl pH 7.4, 100 mM KCl, 3 mM dithiothreitol, 0.02% (v/v) Triton X-100, 10% (v/v) glycerol) plus 0.2 mM EDTA and 400 μg/ml bovine serum albumin for 20 min at 25°C. The reaction mixtures were loaded onto 5% non-denaturing polyacrylamide gels containing 0.25×TBE buffer. Gels were run at 45 mV for 30 min and dried. Bands were quantified on a phosphorimager (Molecular Dynamics) and the dissociation constants were calculated by Scatchard analysis.

5.4. SPR analysis

The duplex DNA molecules were identical to those used in the EMSA studies with an additional duplex, *PRDII*, which contained the binding sequence 5′-GGGAAATTCC-3′ but was otherwise identical to *Ig-κB*. One of each pair of oligonucleotides was 5′-biotinylated during synthesis. Annealing was performed with biotinylated and non-biotinylated oligonucleotides at concentrations of 5 μM and 6 μM respectively in 10 mM Tris–HCl pH 7.4, 1 mM EDTA, 100 mM NaCl by cooling from 95°C to 35°C over 30 min.

The SPR instrument was a BIAcore 2000 optical biosensor (BIAcore AB). The sensor chip SA (BIAcore AB) consisted of a thin gold film coated with a carboxymethyldextran hydrogel matrix to which streptavidin was cross-linked and contained four flow cells. The biotinylated DNA duplexes were injected at 500 pM in buffer KB supplemented with 0.2 mM EDTA at a flow rate of 20 μl/min. The individual flow cells were loaded with 17–19 response units (RU) where 1 RU corresponds to a surface density of DNA of approximately 1 pg/mm^2 [23]. The first flow cell was left underivatised to control for non-specific protein binding to the sensor chip matrix, bulk refractive index changes between the injected solution and the running buffer and baseline drift.

For each experiment, purified p50 was diluted to 10, 20, 30 and 40 nM in buffer KB supplemented with EDTA to 0.2 mM and injected for 360 s at a flow rate of 20 μl/min over each flow cell. The protein sample was then replaced by the buffer alone at the same flow rate and the protein–DNA complex was allowed to dissociate for 1200 s. The chip surface was then regenerated with an injection of the buffer containing 2 M NaCl for 10 s. All solutions were filtered through a 0.22 μm membrane and degassed before use. Each assay was carried out at 25.0°C. Sensorgrams were analysed and kinetic constants calculated as previously described [15].

5.5. Comparative enrichments of fusion proteins using plasmid display

E. coli strains XL1-blue [pRES108], NM554 [pRES311] and NM554 [pRES314] were grown to approximately $OD_{600} = 0.5$ in LB at 37°C, 250 rpm. Although the promoters in these plasmids were derived from the *lac* promoter, the *lac* repressor binding site had been disrupted during incorporation of a *target-κB* site into the −10 region of the promoter. Expression of the relevant fusion proteins was therefore no longer controlled by the *lac* repressor and IPTG induction was not necessary. Dilutions of 1 ml, 100 μl and 1 μl of the culture containing the protein–plasmid complexes to be selected in 10 ml of a competing culture were made and the cells were pelleted by centrifugation (4100 rpm, 15 min). The cells were re-suspended in *x* ml of 0.1 M Tris–HCl pH 8.0 and 20% (v/v) sucrose (where $x = (\text{culture volume (ml)} \times OD_{600})/50$), lysozyme was added to 0.2 mg/ml and the suspension incubated at 37°C, 250 rpm for 20 min. The resulting spheroplasts were pelleted by centrifugation (1300 rpm, 2 min) and frozen at −20°C.

A 1 ml bed volume of affinity resin was pre-equilibrated by washing with 2×15 ml H_2O and 1×15 ml buffer KB containing 75 mM KCl supplemented with 0.1 mg/ml sonicated herring sperm DNA (buffer KBD). Spheroplasts were lysed by re-suspension in 1 ml H_2O and immediately transferred into 12 ml buffer KBD. If necessary, the excess cell debris was pelleted by centrifugation (4100 rpm, 5 min) and the supernatant was gently shaken with the affinity resin for 20 min at room temperature. The resin was pelleted by centrifugation (4100 rpm, 1.5 min), the supernatant decanted and 12 ml of fresh buffer KBD was added.

This washing procedure was repeated five times before bound protein–plasmid complexes were eluted by re-suspension of the resin in 3 ml of specific elution buffer with agitation for 1 h at room temperature. The elution buffer was 10 mM maltose, 20 mM Tris–HCl pH 7.4 for amylose resin, 10 mM glutathione, 50 mM Tris–HCl pH 8.0, 0.5 M NaCl for glutathione Sepharose and 100 mM imidazole, 100 mM NaCl, 20 mM Tris–HCl pH 8.0 for the metal affinity resin (TALON Co^{2+} support, Clontech).

Plasmids from the crude cell lysates and final eluate were purified on a QIAprep spin miniprep column (QIAgen) by adding 0.35 ml of buffer N3 to 0.3 ml of the crude solution prior to addition to the spin column. The spin columns were washed with 0.75 ml buffer PE and eluted in 10 mM Tris–HCl pH 8.5 (30 μl). The purified plasmid mixtures (5 μl) were used to transform *E. coli* and the resulting transformation mix plated in parallel on LB agar plates containing either ampicillin (50 μg/ml) or chloramphenicol (50 μg/ml). After overnight incubation at 37°C, colonies were counted to assess the relative numbers of each plasmid in the crude lysate and enriched samples.

5.6. Library simulation

In the first experiment 100 ml LB (ampicillin 50 μg/ml) cultures containing *E. coli* strain XL1-blue [pRES109] and XL1-blue [pRES110] were grown to early log phase and diluted 1 μl XL1-blue [pRES109] in 100 ml XL1-blue [pRES110] ($1{:}10^5$ dilution) and spheroplasts were prepared as described above. Enrichment on amylose resin was as described above with the exception that 50 mM potassium glutamate was used in place of 75 mM KCl in buffer KBD and the entire 3 ml of the enriched eluant was purified on a QIAprep spin miniprep column (QIAgen), eluting in 30 μl of 10 mM Tris–HCl pH 8.5. The purified eluant (5 μl) was used to transform 100 μl aliquots of *E. coli* DH5α-FT Ultramax cells (Life Technologies) by heat shock according to the manufacturer's protocol and the whole transformation mix incubated overnight at 37°C on LB agar containing ampicillin (50 μg/ml).

To recover the cells, LB media (2 ml) were added to each agar plate and a cell suspension was produced by mechanical agitation of the agar surface. Spheroplasts were then produced as described above and several rounds of enrichment were performed by repeating this protocol. The relative proportion of each plasmid in the enriched DNA pools after each round of selection was analysed by PCR screening colonies with primers that bound in the vector backbone, either side of the *malE* or *gst* genes. The different sizes of the genes allowed for unambiguous identification of the proportion of each plasmid within the population whilst colony counting revealed the absolute number of plasmids at each enrichment step.

In the second experiment, the same cultures were grown to approximately $OD_{600} = 0.5$ and by serial dilution in LB, the equivalent of 1 nl of XL1-blue [pRES109] (shown to be 600 cells by plating an aliquot of the dilution on LB agar containing 50 μg/ml ampicillin) was mixed with 92 ml XL1-blue [pRES110] to give a dilution of $1{:}10^8$. Several rounds of enrichment and analysis were performed as described above for the $1{:}10^5$ dilution experiment.

5.7. Library construction and screening

Genomic DNA was purified from *E. coli* strain TG1 using Genomic-tip 100/G columns (QIAgen). This was fully digested with the restriction enzymes *Hin*fI and *Apa*LI followed by removal of single-stranded overhang DNA with mung bean nuclease (New England Biolabs). DNA fragments between 400 and 2100 bp were gel-purified from a 1% agarose TBE gel. The purified genomic fragments were ligated with *Msc*I-linearised pRES115 plasmid DNA which had been dephosphorylated with shrimp alkaline phosphatase (United States Biochemicals). Subsequent transformation of *E. coli* strain DH5α-FT (Life Technologies) by heat shock produced a library of approximately 1.3×10^5 clones when plated on LB agar plates containing kanamycin (50 μg/ml). Spheroplasts were produced and several rounds of selection on amylose resin were performed as described for the library simulation experiments. Single colonies from each round of selection were analysed by PCR with primers that annealed within the *p50* gene and the transcriptional terminators to assess the size distribution of the inserts in the library. Western blot analysis of specific colonies for the presence of MBP was achieved by growing colonies of bacteria on nitrocellulose membranes supported on LB agar plates at 37°C, overnight. Colonies were then lysed on the membrane by freeze–thaw, then incubation with lysozyme (4 mg/ml) and SDS (1% (w/v)) followed by extensive washing with phosphate-buffered saline to remove cell debris. The presence of MBP was then assessed using standard Western blot procedures using an anti-MBP antibody (New England Biolabs).

Acknowledgements

We thank Professor S.C. Harrison (Harvard, MA, USA) for the gift of pLM1-p50 and Dr G.I. Johnston (Pfizer) for helpful discussions. R.E.S. thanks the BBSRC and Pfizer Global Research and Development for a CASE studentship. D.J.H. was funded by the BBSRC and J.M.B. thanks the Royal Society for a University Research Fellowship.

References

[1] B.K. Kay, J. Winter, J. McCafferty, Phage Display of Peptides and Proteins: A Laboratory Manual, Academic Press, San Diego, CA, 1996.

[2] E.T. Boder, K.S. Midelfort, K.D. Wittrup, Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity, Proc. Natl. Acad. Sci. USA 97 (2000) 10701–10705.

[3] I.N. Maruyama, H.I. Maruyama, S. Brenner, Lambda-Foo – a lambda-phage vector for the expression of foreign proteins, Proc. Natl. Acad. Sci. USA 91 (1994) 8273–8277.

[4] A. Rosenberg, K. Griffin, F.W. Studier, M. McCormick, J. Berg, R. Novy, R. Mierendorf, T7Select phage display system: A powerful new protein display system based on bacteriophage T7, inNovations 6 (1996) 1–6.

[5] P.P. Sche, K.M. McKenzie, J.D. White, D.J. Austin, Display cloning: functional identification of natural product receptors using cDNA-phage display, Chem. Biol. 6 (1999) 707–716.

[6] R.W. Roberts, Totally in vitro protein selection using mRNA–protein fusions and ribosome display, Curr. Opin. Chem. Biol. 3 (1999) 268–273.

[7] D.S. Tawfik, A.D. Griffiths, Man-made cell-like compartments for molecular evolution, Nat. Biotechnol. 16 (1998) 652–656.

[8] J. Hanes, C. Schaffitzel, A. Knappik, A. Pluckthun, Picomolar affin-

ity antibodies from a fully synthetic naive library selected and evolved by ribosome display, Nat. Biotechnol. 18 (2000) 1287–1292.

[9] M.G. Cull, J.F. Miller, P.J. Schatz, Screening for receptor ligands using large libraries of peptides linked to the C-terminus of the *lac* repressor, Proc. Natl. Acad. Sci. USA 89 (1992) 1865–1869.

[10] C.M. Gates, W.P.C. Stemmer, R. Kaptein, P.J. Schatz, Affinity selective isolation of ligands from peptide libraries through display on a *lac* repressor 'headpiece dimer', J. Mol. Biol. 255 (1996) 373–386.

[11] M. Kretzschmar, M. Meisterernst, C. Scheidereit, G. Li, R.G. Roeder, Transcriptional regulation of the HIV-1 promoter by NF-κB in vitro, Genes Dev. 6 (1992) 744–761.

[12] T. Fujita, G.P. Nolan, S. Ghosh, D. Baltimore, Independent modes of transcriptional activation by the p50 and p65 subunits of NF-κB, Genes Dev. 6 (1992) 775–787.

[13] U. Zabel, R. Schreck, P.A. Baeuerle, DNA binding of purified transcription factor NF-κB, J. Biol. Chem. 266 (1991) 252–260.

[14] M. Sodeoka, C. Larson, L. Chen, K. LeClair, G. Verdine, A multifunctional plasmid for protein expression by ECPCR: overproduction of the p50 subunit of NF-κB, Biomed. Chem. Lett. 3 (1993) 1089–1094.

[15] D.J. Hart, R.E. Speight, M.A. Cooper, J.D. Sutherland, J.M. Blackburn, The salt dependence of DNA recognition by NF-κB p50: a detailed kinetic analysis of the effects on affinity and specificity, Nucleic Acids Res. 27 (1999) 1063–1069.

[16] D.J. Hart, R.E. Speight, J.D. Sutherland, J.M. Blackburn, Analysis of the NF-κB p50 dimer interface by diversity screening, J. Mol. Biol. 310 (2001) 565–577.

[17] D.B. Smith, K.S. Johnson, Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione *S*-transferase, Gene 67 (1988) 31–40.

[18] S. Leirmo, C. Harrison, D.S. Cayley, R.R. Burgess, M.T. Record, Replacement of KCl by potassium glutamate dramatically enhances protein–DNA interactions in vitro, Biochemistry 26 (1987) 2095–2101.

[19] F.R. Blattner, G. Plunkett, C.A. Bloch, N.T. Perna, V. Burland, M. Riley, J. ColladoVides, J.D. Glasner, C.K. Rode, G.F. Mayhew, J. Gregor, N.W. Davis, H.A. Kirkpatrick, M.A. Goeden, D.J. Rose, B. Mau, Y. Shao, The complete genome sequence of *Escherichia coli* K-12, Science 277 (1997) 1453–1462.

[20] G. Ghosh, G.V. Duyne, S. Ghosh, P.B. Sigler, Structure of NF-κB p50 homodimer bound to a κB site, Nature 373 (1995) 303–310.

[21] C.W. Müller, F.A. Rey, M. Sodeoka, G.L. Verdine, S.C. Harrison, Structure of the NF-κB p50 homodimer bound to DNA, Nature 373 (1995) 311–317.

[22] L.A. Ruiz-Taylor, T.L. Martin, F.G. Zaugg, K. Witte, P. Indermuhle, S. Nock, P. Wagner, Monolayers of derivatized poly(L-lysine)-grafted poly(ethylene glycol) on metal oxides as a class of biomolecular interfaces, Proc. Natl. Acad. Sci. USA 98 (2001) 852–857.

[23] E. Stenberg, B. Persson, H. Roos, C. Urbaniczky, Quantitative determination of surface concentration of protein with surface-plasmon resonance using radiolabelled proteins, J. Colloid Interface Sci. 143 (1991) 513–526.