J. Mol. Biol. (1997) **267**, 118–131

JMB

Single-chain Repressors Containing Engineered DNA-binding Domains of the Phage 434 Repressor Recognize Symmetric or Asymmetric DNA Operators

András Simoncsits*, Jinqiu Chen, Piergiorgio Percipalle Shenglun Wang, Imre Törö and Sándor Pongor*

International Centre for Genetic Engineering and Biotechnology (ICGEB), Area Science Park Padriciano 99, I-34012, Trieste Italy

*Corresponding authors

Single-chain (sc) DNA-binding proteins containing covalently dimerized N-terminal domains of the bacteriophage 434 repressor cI have been constructed. The DNA-binding domains (amino acid residues 1 to 69) were connected in a head-to-tail arrangement with a part of the natural linker sequence that connects the N and C-terminal domains of the intact repressor. Compared to the isolated N-terminal DNA-binding domain, the sc molecule showed at least 100-fold higher binding affinity *in vitro* and a slightly stronger repression *in vivo*. The recognition of the symmetric O_R1 operator sequence by this sc homodimer was indistinguishable from that of the naturally dimerized repressor in terms of binding affinity, DNase I protection pattern and *in vivo* repressor function. Using the new, sc framework, mutant proteins with altered DNA-binding specificity have also been constructed. Substitution of the DNA-contacting amino acid residues of the recognition helix in one of the domains with the corresponding residues of the *Salmonella* phage P22 repressor c2 resulted in a sc heterodimer of altered specificity. This new heterodimeric molecule recognized an asymmetric, artificial 434-P22 chimeric operator with high affinity. Similar substitutions in both 434 domains have led to a new sc homodimer which showed high affinity binding to a natural, symmetric P22 operator. These findings, supported by both *in vitro* and *in vivo* experiments, show that the sc architecture allows for the introduction of independent changes in the binding domains and suggest that this new protein framework could be used to generate new specificities in protein-DNA interaction.

Keywords: single-chain proteins; 434 repressor; helix-turn helix motif; protein-DNA interaction; altered recognition specificity

Introduction

Engineering DNA-binding proteins with altered recognition specificity and affinity to DNA targets is an important research area with potential practical applications. DNA-binding proteins of novel functions can be obtained by rational design or by selection from randomized, combinatorial libraries. These approaches may work best if applied within a natural protein framework which may provide good folding and stability characteristics to the mutant proteins. Redesigning the natural topology of a protein to obtain a simplified stable framework, whilst maintaining the essential functions, is of theoretical and practical importance as a simplified framework may allow new approaches to be applied to structure-function studies.

Most of the DNA-binding proteins involved in gene regulation are of oligomeric nature in their active form (see Wong & Lohman, (1995) for examples and references) and recognize inverted or direct repeats as target sequences. However, the modular nature of DNA-binding proteins (Pabo & Sauer, 1992) allows identical or different DNA-binding domains to be combined either with the help of natural dimerization domains (for a review,

Present address: P. Percipalle, MRC Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH, UK.

Abbreviations used: sc, single-chain; HTH, helix-turn-helix; PCR, polymerase chain reaction; IPTG, isopropyl-β-D-thio-galactoside; EMSA, electrophoretic mobility shift assay.

0022–2836/97/110118–14 $25.00/0/mb960832

see Hu, 1995) or by covalently joining them through designed or selected peptide linker sequences. The latter approach has been shown to generate extended recognition surfaces with enhanced affinity and/or altered specificity (Pomerantz *et al*., 1995; Percipalle *et al*., 1995a; Robinson & Sauer, 1996; Gates *et al*., 1996). An intriguing question is which of these strategies can be best used to join DNA-binding domains and thereby to combine the subsite recognition properties of the individual domains. Different binding domains, supplied with natural dimerization domains, may form a mixture of two homodimers and the expected heterodimer. This makes specificity studies with the heterodimer complicated, although the detection of new specificity toward a chimeric operator has been documented, both *in vitro* (Hollis *et al*., 1988) and *in vivo* (Webster *et al*., 1992). Preferential heterodimer formation may be achieved by using two different dimerization domains with a high tendency to form heterodimers (Kim & Hu, 1995). A covalent attachment of the binding domains, however, may prove to be superior to the use of natural dimerization domains, as this strategy provides smaller molecules which are more amenable to physico-chemical studies (Percipalle *et al*., 1995a). The dimerization is permanent and therefore concentration-independent. The binding domains are brought to close proximity by the linker, resulting in an arrangement which is energetically more favourable than the non-covalent active dimer formation through monomer-dimer equilibrium. Using the covalent attachment approach, two binding domains with identical or different recognition properties can be combined with strictly defined stoichiometry leading to the specific recognition of combined subsites which, in turn, can be symmetric (for a sc homodimer) or asymmetric (for a sc heterodimer).

We have chosen the DNA-binding domain of bacteriophage 434 repressor (cI) for artificial, covalent dimerization to obtain a new protein framework in which systematic or library-based combinatorial changes can be introduced at positions which are mainly responsible for specific DNA recognition. This repressor is one of the best studied members of the helix-turn-helix (HTH) family of DNA-binding proteins (see Ptashne (1992) for a review and references). Amino acid substitutions in the α3 recognition helix of this protein resulted in stable mutants with altered recognition properties (Wharton *et al*., 1984; Wharton & Ptashne, 1985, 1987). It was also shown that the equilibrium mixture of the wild-type and a redesigned 434 repressor, both containing the natural dimerization domain, resulted in a heterodimer with new DNA-binding specificity (Hollis *et al*., 1988).

Our approach to obtain altered recognition specificity mutants in a simplified framework of the 434 repressor is based on previous observations that covalent dimerization of the DNA-binding domain, in head-to-tail or tail-to-tail arrangements, provided sc proteins with high binding affinity for 434 phage operator sites (Percipalle *et al*., 1995a; 1995b). The head-to-tail arrangement was obtained by expression of a recombinant gene coding for tandem copies of the N-terminal 1 to 69 amino acid residues and a linker segment of the intact 434 repressor. Here we describe the *in vitro* and *in vivo* DNA-binding properties of this sc molecule and show that this architecture provides a simple framework for changing and combining DNA-binding specificities of recognition helices. By introducing amino acid changes in one or both of the domains, hetero- or homodimeric sc repressor analogs with extended high affinity recognition surfaces were constructed, which specifically interacted with asymmetric or symmetric DNA operator sequences.

Results

Construction of sc repressor analogs

The gene encoding two direct repeats of the N-terminal domain of the phage 434 repressor was constructed from two DNA fragments obtained by independent PCR amplifications on a λgt10 template (Huynh *et al*., 1984). This construct can be described as $R_{1\text{-}69}L_{70\text{-}89}R_{1\text{-}69}$, where R refers to the repressor function of the N-terminal DNA-binding domain, L refers to a linker sequence and the subscript numbers identify the amino acid residues of the natural repressor. For the sake of simplicity, we use the abbreviation RR69 for this homodimeric sc repressor. The $L_{70\text{-}89}$ sequence was chosen because it is part of the natural linker that connects the N and C-terminal domains of the intact repressor (Carlson & Koudelka, 1994). Computer modelling based on the $R_{1\text{-}69}$/operator crystal structure (Aggarwal *et al*., 1988) showed that, although substantially shorter peptide sequences could span the distance between the C and N termini of the two subunits, the chosen long linker should also be suitable as it is likely to adopt a flexible conformation (Percipalle *et al*., 1995a). Silent mutations were then introduced into the RR69 coding gene near the borders of the α3 recognition helix of the second domain to generate unique restriction sites for *Kpn*I and *Xho*I. This enabled us to perform a ''helix swap'' experiment similar to that described (Wharton & Ptashne, 1985) but which is restricted to the second domain of RR69. Amino acid residue replacements at positions, −1, 1, 2 and 5 of the α3 helix of the 434 repressor with the corresponding residues of the P22 phage c2 repressor resulted in a heterodimeric sc repressor, abbreviated as RR*69. Identical amino acid replacements were then performed in the first domain of RR*69 to obtain a homodimeric sc repressor R*R*69, which contains two DNA-binding domains of the 434 repressor with the P22 repressor amino acid residues at the DNA-binding positions. In order to evaluate the *in vivo* function of these sc repressor analogs, the control genes coding for the natural 434 repressor cI, its N-terminal 1-69 domain R69, and a

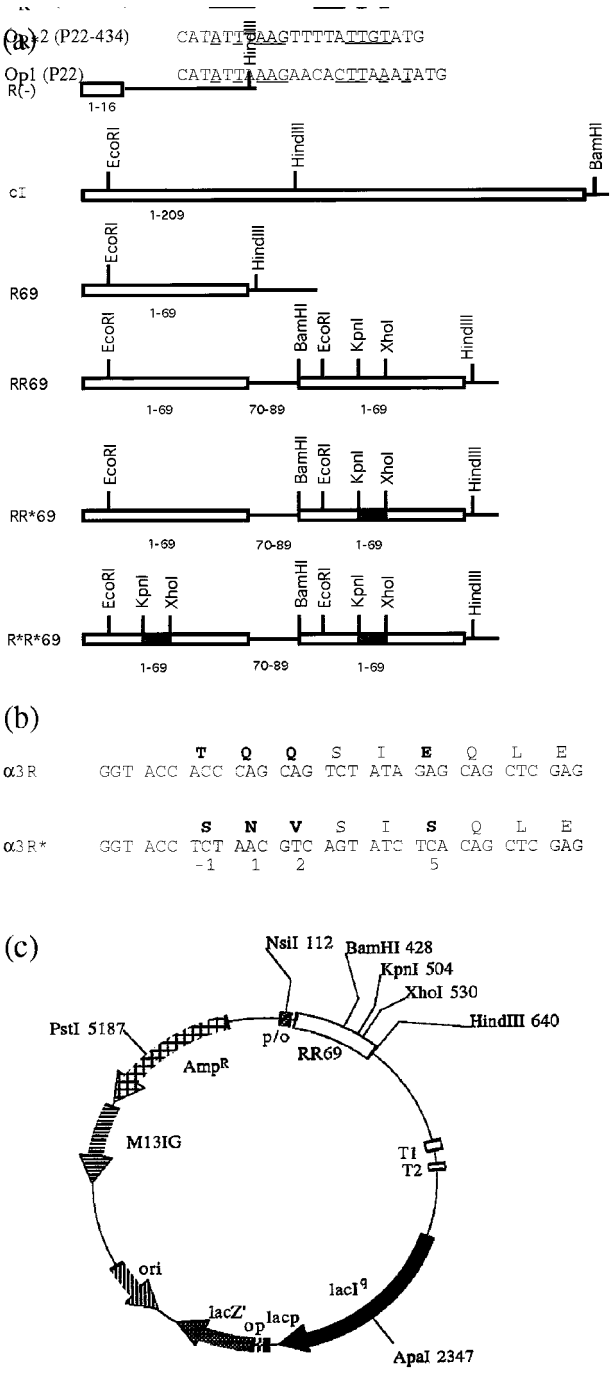

Figure 1. Components of the pRIZ′ vectors used to detect repressor-operator interaction *in vivo*. (a) Linear maps of the genes coding for the repressor analogs. Open boxes represent coding regions for 434 repressor parts, the linker of the sc repressors is shown as a continuous line between boxes and is labelled 70-89. Shaded areas between *Kpn*I and *Xho*I sites represent altered α3 recognition helices. (b) Nucleotide and amino acid sequences of α3 helix regions: α3R, α3 helix of 434 repressor; α3R*, altered α3 helix in R* domain after replacement of amino acid residues −1, 1, 2 and 5 of α3 with the corresponding residues of the P22 repressor. (c) Map of the pRIZ′ vectors exemplified by pRIZ′O_{lac}RR69. Repressor analog genes (a) are located as shown for RR69 and are placed downstream of the *rrnB* P2 promoter-*lac* operator region (p/o). T1 and T2 represent tandem transcriptional terminators of the *rnnB* operon. (d) Operator sequences at the op site of (c), listed as upper strand sequences in *lac* promoter-op-*lacZ*′ arrangement. Lower case letters show operator flanking sequences. The consensus bases of the 434 and P22 operators are underlined.

frameshifted derivative of the latter, R(−), coding for 16 residues only, were also constructed and cloned into the expression/detection vectors. The repressor analogs used in this study are listed in Figure 1(a) and the altered nucleotide and amino acid sequences are shown in Figure 1(b).

Construction of a system to detect *in vivo* repressor-operator interaction

The system used to detect *in vivo* repressor-operator interaction in this work is conceptually similar to other systems (Wharton & Ptashne, 1987; Lehming *et al*., 1987), with the major difference that the repressor gene and the operator-reporter gene fusion are on the same replicon. The repressor analog genes, devoid of any regulatory element of the imm^{434} region, were cloned into derivatives of the pRIZ′ *Escherichia coli* expression vector (Simoncsits *et al*., 1994). This vector contains the *lac* operator-controlled *rrnB* P2 promoter (Lukacsovich *et al*., 1990; Simoncsits *et al*., 1988) combined with an improved ribosomal binding site (Simoncsits *et al*., 1994) to establish high level expression of cloned genes. It also contains the *lacI*^{q}-*lac*pro-*lacZ*′(1-146) region and it is abbreviated here as pRIZ′O_{lac}. To replace O_{lac}, first it was deleted by a simultaneous creation of a unique *Nde*I site to obtain pRIZ′O(−) (Figure 1(d)). Natural or designed operators, as listed in Figure 1(d), were then cloned into this site between the *lac* promoter and the *lacZ*′ gene. The repressor analogs (Figure 1(a)) and operators (Figure 1(d)) can be shuffled, by using unique restriction sites (Figure 1(c)), to obtain any combination of them in the same vector. Thereby a simplified system is created in which a single expression vector carries all the necessary elements (a regulated repressor gene and its putative operator target placed upstream of a reporter gene) to detect *in vivo* repressor-operator interactions. The *lacZ*′(1-146) reporter gene on the plasmid gives rise to β-galactosidase activity only through α-complementation (Ullmann *et al*., 1967) with a defective β-galactosidase encoded by the episome of the host strain XL1-Blue (Bullock *et al*., 1987). The principle

of the detection is based on parallel and consecutive derepression-repression mechanisms taking place at three operator sites. The lactose analog IPTG (isopropyl-β-D-thio-galactoside) derepresses the *rrnB* P2 promoter-*lac* operator-controlled repressor and concomitantly the *lac* promoter-operator controlled *lac*ZΔM15 synthesis. Since the expression of the *lacZ′* reporter gene is not under *lac*I control, it is constitutive unless the derepressed overproduced repressor analog occupies the created operator site. Thus, the level of the observed β-galactosidase activity depends mainly on the interaction of the repressor analog with the operator of the reporter gene.

In vivo interaction of sc repressors with cognate, half-cognate and non-cognate operators

The *in vivo* detection system described above was applied to most of the repressor analog-operator pairs shown in Figure 1, and the results are summarized in Table 1. As it was observed that different operator constructs gave rise to different β-galactosidase activities under non-repressed conditions, relative β-galactosidase activities were calculated and are shown in Table 1. This presentation of the data makes it easier to compare the repression levels, obtained by binding of different repressor analogs to a given operator or by binding of a given repressor analog to different operators.

The most important observation based on the data of Table 1 is that the homodimeric sc repressor RR69 is as active *in vivo* as is the natural counterpart cI when tested on the natural O_R1 operator (27 and 25% relative β-galactosidase activities, respectively, which represent about fourfold repression). The N-terminal domain R69 shows lower, but significant, repression (33% relative value, threefold repression). The mutant sc repressors show significantly lower but clearly detectable repression. These latter observations suggest that the intracellular concentrations of the repressor analogs are high enough to give rise to interaction of R69 with O_R1 as well as interactions of the sc repressor analogs with half-cognate and non-cognate operators. *In vitro* studies (see below) showed that these interactions do take place at high protein concentrations. Nevertheless, the repression observed on O_R1 with different repressor analogs follows the expected order (see the O_R1 column of Table 1). Comparison of the different sc repressor analogs on O_R1 shows that the cognate RR69-O_R1 interaction (where two operator subsites may interact with two cognate DNA-binding domains) is the strongest, the half-cognate RR*69-O_R1 interaction (where one binding domain of RR*69 is cognate to either subsites of O_R1) is medium and the non-cognate R*R*69-O_R1 interaction is the weakest.

A similar order of cognate and half-cognate interactions on the hybrid, asymmetric operators $O_{R*}1$ and $O_{R*}2$, and on the symmetric operator O_P1 (O_R1 of P22) can be detected, as shown in the respective columns of Table 1. It is interesting to note that on these operators the non-cognate RR69 and the R69 do not show any repression and surprisingly, the β-galactosidase activities are 20 to 50% higher than those observed with the non-functional R(−) control. Our experimental data do not provide a rational explanation for this observation.

We mention here that the pRIZ′ vectors contain the *lacL8* mutation which makes the *lac* promoter insensitive to positive regulation by CAP (Ebright *et al*., 1984). This implicates that the β-galactosidase activities (Miller, 1972) may be too low in our system to detect repression. This is indeed true for the constructs containing the *lac* operator, but substantially higher activities were observed when the *lac* operator was replaced with the operators used in this study. The β-galactosidase activities obtained in a typical experiment were 4.8, 164, 235, 284 and 245 units for O_{lac}, O_R1, $O_{R*}1$, $O_{R*}2$ and O_P1 operators, respectively, when they were tested under similar non-repressed conditions. Compared to O_{lac}, even its deletion derivative, O(−) showed four to five times higher β-galactosidase activity.

Table 1. Recognition of operators by different repressors *in vivo*

	β-Galactosidase activity[b] observed with operators[c]			
Repressor[a]	O_R1	$O_{R*}1$	$O_{R*}2$	O_P1
$R^{(-)}$	100 ± 14	100 ± 12	100 ± 9	100 ± 16
R69	33 ± 4	115 ± 6	120 ± 4	130 ± 10
RR69	**27** ± 3[d]	120 ± 11	138 ± 12	152 ± 9
cI	**25** ± 3	ND[e]	ND	ND
RR*69	55 ± 5	**49** ± 4	**30** ± 4	53 ± 3
R*R*69	67 ± 3	64 ± 2	43 ± 2	**43** ± 4

[a] Repressors or repressor analogs are shown in Figure 1(a) and 1(b).

[b] Data measured in Miller units (Miller, 1972) are expressed in relative β-galactosidase activity as a percentage of the units obtained with R(-) non-functional polypeptide (non-repressed condition). A 100% was given for all non-repressed operators although the absolute units were different for different operators (see text for data). Data shown are based on four to six independent assays (mean ± standard deviation).

[c] Operator sequences are listed in Figure 1(d).

[d] Data for cognate or putative cognate interactions are shown in bold.

[e] Not determined. Vectors containing these repressor-operator combinations were not constructed.

This observation, together with the fact that different synthetic operator sequences cloned in different orientations all gave rise to elevated β-galactosidase activity, makes it unlikely that an insertion of a fortuitous promoter element is responsible for the higher activities.

Expression and purification of sc repressors

The sc repressors were initially expressed using the pRIZ′ expression vectors in XL1-Blue cells in relatively low levels (less than 1% of the total cell protein). For large scale expressions, T7 promoter-based expression systems were chosen (Studier *et al.*, 1990). Vector construction, expression and purification are described in Materials and Methods. Figure 2 shows the purities of the sc repressors used in the *in vitro* studies.

In vitro DNA-binding properties of sc repressors

Electrophoretic mobility shift assay (EMSA) was used to determine the protein concentration required for half-maximal binding, which represents the apparent equilibrium dissociation constant (K_d), if the DNA probe concentration is negligible compared to the protein concentration and the binding equilibrium is not perturbed under the assay conditions (Carey, 1991; Fairall *et al.*, 1992). Small increments in protein concentration were used in these titration experiments to be able to define, even by visual inspection of the autoradiograms, a narrow concentration range in which half-maximal binding takes place. Figure 3 shows representative experiments performed at the same time with different operator-repressor cognate pairs. DNA probes containing the cloned operator sequences were about 160 bp long. Based on these and on many other experiments, RR69 and the

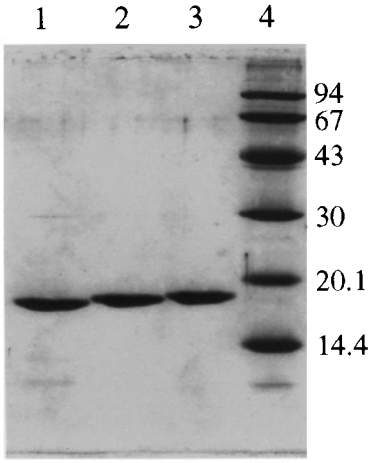

Figure 2. SDS-PAGE analysis of the purified sc repressors. Lane 1, R*R*69; lane 2, RR*69; lane 3, RR69; lane 4, molecular mass standards (kDa); 100 pmol proteins were analysed and bands were detected by Coomassie blue staining.

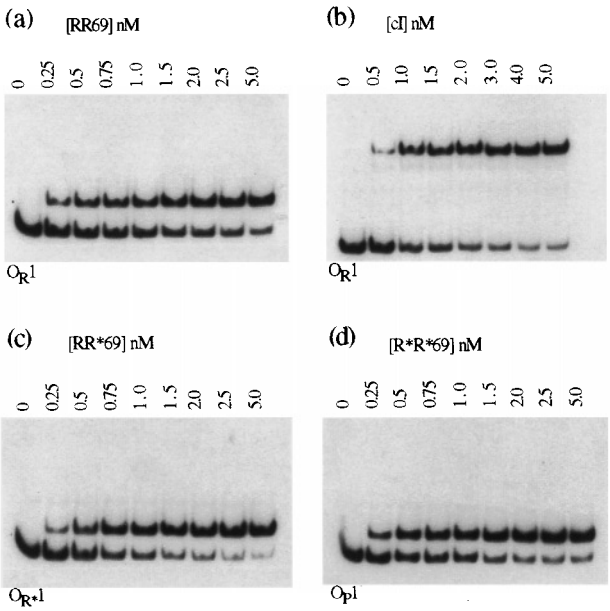

Figure 3. Determination of half-maximal binding for cognate protein-DNA interactions by EMSA. The ^{32}P end-labelled DNA probes (157 to 160 bp) were generated by PCR. The binding reactions and electrophoretic separations were performed as described in Materials and Methods. (a) Interaction of RR69 with O_R1 (O_R1 of 434). (b) Interaction of cI repressor with O_R1. (c) Interaction of the heterodimer RR*69 with $O_{R*}1$ (434-P22 hybrid operator containing the consensus boxes of 434 O_R1 and P22 O_R1). (d) Interaction of the double-mutant homodimer R*R*69 with O_P1 (O_R1 of P22). The DNA probe concentrations were less than 20 pM. Protein concentrations are shown over the respective lanes of the gels. For easy comparison of the affinities of RR69 and cI to O_R1, concentrations of cI in (b) were given as dimer equivalents.

natural cI repressor bind to O_R1 with approximately the same affinity ($K_d \sim 1 \times 10^{-9}$ M) under our assay conditions. At the same time, half-maximal binding with the isolated DNA-binding domain R69 was observed between 100 and 200 nM concentration (expressed in dimer equivalent for better comparison). For the $O_{R*}1$-RR*69 and O_P1-R*R*69 interactions the estimated K_d value is in the range of 5×10^{-10} to 8×10^{-10} M. These data are in close agreement with those reported for the naturally dimerized 434, P22 and hybrid repressors and for R69 (Hollis *et al.*, 1988; Wharton & Ptashne, 1985; Koudelka & Lam, 1993; Bell & Koudelka, 1993; Carlson & Koudelka, 1994).

Substantially higher sc repressor concentrations than those shown in Figure 3 were also tested with the three operators (O_R1, $O_{R*}1$ and O_P1) in all possible operator-repressor combinations in order to detect non-specific binding or cross-reactivity. The appearance of shifted bands in non-cognate interactions may indicate either non-specific binding or lower affinity specific binding due to sequence homology between different operator subsites. The appearance of multiple shifted bands in both cog-

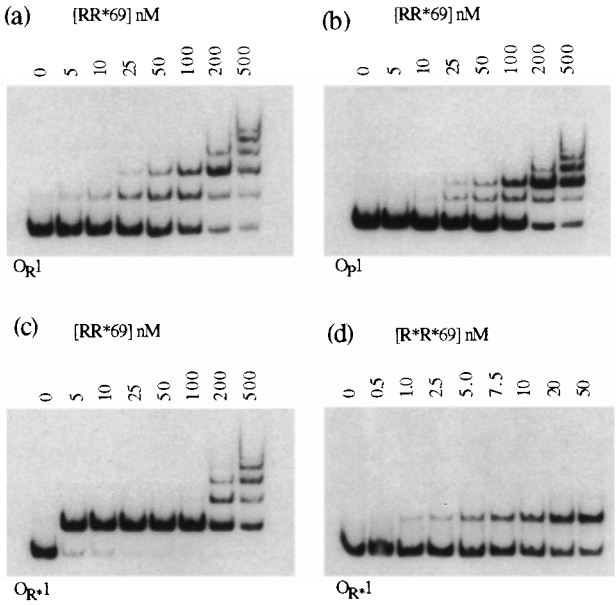

Figure 4. Detection of half-cognate interactions by EMSA. (a) RR*69-O_R1 interaction, the R domain of RR*69 is cognate to O_R1 half-sites. (b) RR*69-O_P1 interaction, the R* domain is half-cognate to O_P1 half-sites. (c) Interaction of RR*69 with its cognate operator $O_{R*}1$ under the same conditions as in (a) and (b), showing a significantly higher affinity as seen from the disappearance of the free DNA. See also Figure 3(c) for lower RR*69 concentrations. (d) Interaction of R*R*69 homodimer with the half-cognate $O_{R*}1$ operator, showing a half-maximal binding with single stoichiometry at or slightly under 20 nM protein concentration. Only trace amounts of higher complexes could be detected at even higher (100 nM) concentration (not shown). For comparison with the cognate operator interaction see also Figure 3(d).

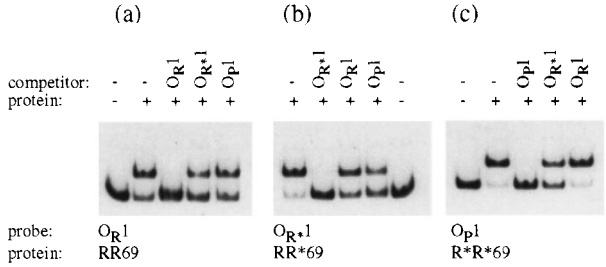

Figure 5. Specificity of the cognate interactions as studied by qualitative competition assays. Complexes formed between operator probes and repressors such as O_R1-RR69 (a), $O_{R*}1$-RR*69 (b), and O_P1-R*R*69 (c) pairs were treated with large excesses of unlabelled double-stranded oligonucleotide competitors and the reaction mixtures were analyzed by EMSA. Operator probes were 157 to 160 bp long, ^{32}P end-labelled PCR fragments. The competitor operators were 17 bp (O_R1), 18 bp ($O_{R*}1$) and 20 bp (O_P1) synthetic double-stranded oligonucleotides. The final concentrations were: <20 pM probe DNA, 25 nM non-specific carrier DNA, 5 nM sc repressor and 100 nM competitor oligonucleotides.

nate and non-cognate interactions may be due either to non-specific binding outside the operator region or to protein aggregation at the operator site. At 200 nM concentration, RR69 did not give detectable bandshift with $O_{R*}1$ and O_P1, and gave only a single specific shift with its cognate O_R1 (not shown). RR*69 gave a single specific bandshift with its cognate $O_{R*}1$ operator in a broad concentration range and only weak bands of slower mobilities were seen at 200 nM concentration (Figure 4(c)). With the half-cognate O_R1 and O_P1 operators, however, RR*69 was shown to interact at relatively low concentrations. In these cases, both single and double-shifted bands could be detected at 10 to 25 nM and increasing protein concentration gave rise to a predominantly double-shifted band at around 200 nM concentration (Figure 4(a) and 4(b)). It is possible that at higher concentrations, RR*69 forms protein dimer (or oligomer) at the half-cognate binding site of the long DNA probe. Similar binding and protein aggregation under identical conditions was not detected on a DNA probe which did not contain an operator site (see O(−) of Figure 1(d); data not shown), indicating that the initial DNA binding by RR*69 is due mainly to the presence of half-cognate operator sites and not to non-specific interactions. These observations are in agreement with the *in vivo* behaviour of RR*69 which showed nearly twofold repression on the half-cognate operators (see Table 1). R*R*69 was shown to bind with single stoichiometry to both its cognate O_P1, and to the half-cognate $O_{R*}1$, up to about 100 nM concentration, but the protein concentrations at half-maximal binding were substantially different: 20 nM for $O_{R*}1$ (Figure 4(d)) and 0.5 to 0.8 nM for O_P1, indicating about 30-fold higher affinity to the cognate O_P1 site over the half-cognate site. On the non-cognate O_R1 site, R*R*69 did not show significant binding under 200 nM concentration.

Competition assays were also performed in order to support the results of the direct binding experiments, which suggested that the operator sites of the long (160 bp) DNA probes were the targets of the specific binding. In these assays, the radioactively labelled long DNA probes were allowed to form complexes with the respective cognate repressors, then large excess of short double-stranded DNA competitors containing only the operator sequences were added. Figure 5 shows that the cognate operator competitors caused complete dissociation of the preformed complexes, in contrast to the half-cognate or non-cognate competitors which caused only partial or no dissociation. These results are in qualitative agreement with those obtained in the direct binding experiments.

DNase I protection experiments, performed in all possible operator-repressor combinations are shown in Figure 6. Relatively high repressor concentrations (100 and 200 nM) were used in order to be able to detect both the specific and the possible non-specific or non-cognate interactions. As ex-

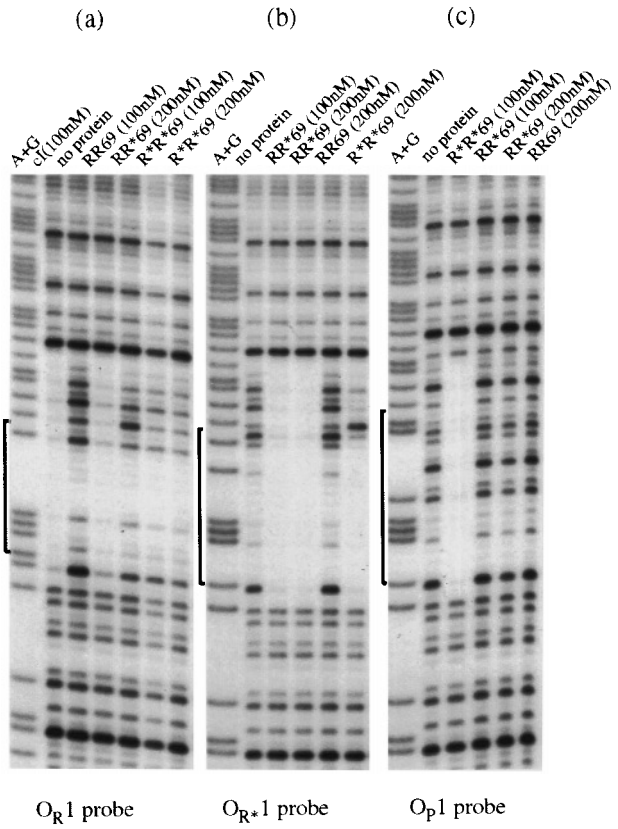

Figure 6. Analysis of sc repressor-operator interactions by DNase I protection assay. Operator probes were 157 to 160 bp long PCR products and the 5′ end of the lower strands (complementary to the sequences listed in Figure 1(d)) were ^{32}P end-labelled. Protection of the O_R1 region (a), the $O_{R*}1$ region (b) and the O_P1 region (c) by different repressors are shown. The operator regions are marked by vertical lines and aligned with the A + G control lanes.

pected, strong protection could be detected only at the operator sites with the corresponding cognate repressors. The natural cI and the sc RR69 repressors showed identical footprints at the O_R1 site. RR69 protected neither $O_{R*}1$ nor O_P1. RR*69 and R*R*69 showed strong protection at their respective cognate operator sites $O_{R*}1$ and O_P1. RR*69 showed very weak protection at 200 nM concentration on its half-cognate operators, O_R1 and O_P1 (see Figure 6(a) and 6(c), respectively). Under the same conditions, EMSA showed the disappearance of over 50% of the free O_R1 and O_P1 probes and mainly double-shifted bands were detected (Figure 4(a) and 4(b)). The assumption based on the EMSA results that, at high protein concentrations, RR*69 could dimerize at the half-cognate operator sites could be supported by the weak DNase I protection patterns which were restricted to the operator regions of the long DNA probes. The difference between the extent of the DNase I protection and the extent of binding observed by EMSA could be due to different kinetic stabilities of the complexes under the conditions of these two different methods. R*R*69 showed partial protection of the O_R1 non-cognate operator, and strong protection of the $O_{R*}1$ half-cognate operator. In this latter case, however, the strongly protected area of the footprint was approximately six nucleotides shorter (see lane R*R*69 of Figure 6(b)) than that obtained with the cognate RR*69, indicating an asymmetric occupancy of the half-sites of the hybrid $O_{R*}1$ operator. For the same interaction, EMSA showed a single stoichiometry binding with a K_d value of $\sim 2 \times 10^{-8}$ M. It is possible that one of the binding domains of R*R*69 may initially approach and interact with the consensus CTT.A.T P22 operator half-site (Poteete *et al*., 1980), causing a local increase in the effective concentration of the covalently joined second domain, which can then recognize with lower affinity the homologous aTTgtaT sequence of the other, 434 specific half of the $O_{R*}1$ operator. This mode of binding is somewhat similar to that observed, for example in the glucocorticoid receptor-DNA complex which shows that one monomer of the DNA-induced homodimer binds to a specific half-site and the other one binds to a non-specific sequence (Luisi *et al*., 1991). The Arc repressor provides a further example for such binding, detected by EMSA, to a DNA fragment containing only one specific half-site (Brown & Sauer, 1993). It is likely, however, that the second domain of R*R*69 interacts with a correctly spaced and homologous half-site sequence, as shown above.

The sc architecture of the covalently dimerized N-terminal domains may, in principle, allow the recognition of operator half-sites with different juxtapositions assuming that the linker is really flexible and that it allows the independent free movement of the individual binding domains within the constraint of the linker length. We have tested this possibility with mutant O_R1 operators in which the consensus recognition boxes were spaced by five, seven, eight and ten nucleotides instead of the natural six nucleotide spacing. As detected by EMSA, no binding took place between RR69 and the mutant operators at a protein concentration which is 100-fold higher than that required for half-maximal binding to O_R1 (not shown). A similar result was obtained with RR*69, which did not show significant binding at 50 nM concentration to an $O_{R*}1$ hybrid operator analog which contained six nucleotides between the consensus operator boxes of 434 (ACAA) and P22 (CTT.A.T) instead of the five nucleotides spacing of the designed $O_{R*}1$ operator (not shown). These findings and binding site selection experiments performed with RR69 and RR*69 (J. C., S. P. & A. S., unpublished results) show that a strict spacing of the contacted half-operator boxes is required for high affinity binding to take place. The spacing preferred by the sc repressors is the same as that observed for the naturally dimerized repressor counterparts: see RR69 *versus* 434 repressor (Harrison & Aggarwal, 1990, and references cited therein) and RR*69 *versus* the

434 + 434R[α3(P22R)] heterodimer complex (Hollis *et al*., 1988). Therefore, it is likely that the covalently joined binding domains of the DNA-bound sc repressors are in the same spatial arrangements as the $R_{1\text{-}69}$ domains in the complexes with DNA. X-ray crystallography studies for the latter complexes revealed significant interdomain contacts at the dimer interface (Anderson *et al*., 1987; Aggarwal *et al*., 1988), which are believed to be the same in the intact repressor (Aggarwal *et al*., 1988). We assume that the same dimer contacts primarily determine the orientation of one monomer with respect to the other in the DNA-bound sc repressors, and therefore the linker plays no role in the juxtapositioning of the binding domains. In other words, the natural linker seems to be flexible enough to undergo conformational changes, probably enforced by the contacts between the two domains.

Discussion

We have constructed sc proteins containing two linked DNA-binding domains of the phage 434 repressor, which functions as a homodimer. For covalent dimerization an arbitrarily chosen, 20 amino acid residues long natural linker, present in the same intact repressor was used and proved to be suitable. It is likely that designed, flexible peptide linkers of suitable length such as those used, for example in sc antibodies (Huston *et al*., 1988) or in the P22 Arc repressor dimer (Robinson & Sauer, 1996) could also have been used in our constructs. Nevertheless, we chose the natural linker because practical considerations in the gene construction favoured the use of this domain-contiguous linker, and previous modelling studies indicated that it could undergo flexible conformational changes (Percipalle *et al*., 1995a). Using this linker sequence, covalent dimerizations were also performed between the wild-type and an engineered DNA-binding domain and between two identically altered domains to obtain a heterodimeric (RR*69) and a new homodimeric (R*R*69) sc repressor, respectively. To the best of our knowledge, these are the first examples of covalent dimerization of engineered DNA-binding domains of a transcription factor, effected by a recombinant linker. The mutant zinc finger domains of the Cys2His2 type, obtained by rational or random mutagenesis (for reviews, see Berg & Shi, 1996; Rhodes *et al*., 1996), also contain engineered covalently linked DNA-binding modules. These are, however, naturally linked in an existing framework. The difference in this work is that, for similar mutagenesis studies, a new sc framework has been created by covalent attachment of DNA-binding domains that naturally function in non-covalent dimers.

We have shown that covalent joining of the DNA-binding domains of the 434 repressor by different strategies gave rise to an increased DNA-binding affinity compared to that of the isolated domain, and that the sc molecules could be used to study conformational changes upon interaction with DNA (Percipalle *et al*., 1995a). It is shown here that the sc RR69 and the natural cI bind to O_R1 with the same affinity, therefore the covalent linker has the same binding enhancement effect as the non-covalent dimerization domain. RR69 recognizes operators only with the natural 6 bp spacing between the consensus half-sites. This suggests that the relative orientation of the two domains in the DNA-bound RR69 is the same as in the natural repressor and it is primarily determined by the interdomain contacts observed between the R69 monomers in the complexes with DNA (Aggarwal *et al*., 1988). DNase I footprinting, *in vivo* repression assay and binding site selection experiments (not shown here) further confirmed that RR69 and cI recognize DNA in a similar manner. The functional replacement of the non-covalent dimerization domain with the covalent linker resulted in a sc molecule with significantly reduced (by over 60%) molecular mass but with the DNA-binding properties of the parent molecule.

The sc repressors of this study selectively recognized their respective cognate operators with high affinity. Half-maximal binding was observed around or slightly under 1 nM repressor concentrations. Significant binding to non-cognate or to half-cognate sites took place at least at a 100-fold higher concentrations, with the exception that R*R*69 showed only 30-fold discrimination between the cognate O_P1 and the half-cognate $O_{R*}1$ operators. The interesting observation that, at higher repressor concentration (200 nM), the sc heterodimer RR*69 binds to the half-cognate operators O_R1 and O_P1 with predominantly double stoichiometry may serve as a starting point to elucidate a mechanism of DNA recognition by the sc repressors. Based on our experimental data, we assume that at high protein concentrations and at symmetric operator sites of long DNA probes the heterodimeric sc repressors may dimerize. Our hypothesis is that in these cases an asymmetric interaction with single stoichiometry can initially take place: one domain makes specific strong contacts with one half-site and the other one interacts weakly with the other half-site of the operator. This is probably the case since a single-shifted band could be detected by EMSA at relatively low protein concentrations. At higher protein concentrations, this complex may rearrange by the replacement of the weakly interacting domain with the specific domain of a second molecule, which approaches the operator site by a sliding mechanism. The new complex can be stabilized by the stronger protein-DNA contacts and, possibly, by the formation of intermolecular protein-protein contacts. The involvement of the sliding mechanism in this hypothetical pathway seems to be essential, since the domain replacement, as above, by direct entry is very unlikely. By using short DNA probes containing only the operator sequences, dimerization could not be observed at even higher protein con-

centrations. However, the dimerization observed *in vitro* on longer DNA is likely to take place *in vivo*, and this may explain the discrepancies occasionally observed *in vivo*.

The *in vivo* interaction of the sc repressors with operator DNA was studied in a simplified detection system. The one-plasmid system used here may be advantageous over the two-plasmid systems as no selective loss of either the repressor gene or the operator-reporter gene fusion can take place under the assay conditions. In this system the basic sc repressor RR69 was just as active as the natural 434 repressor. The fact that the isolated DNA-binding domain R69 caused only a slightly lower repression than either the covalent or the natural dimer does not contradict our *in vitro* binding data. Overexpression of the isolated N-terminal domain of lambdoid repressors was shown to give rise to detectable activity *in vivo* (Sauer *et al*., 1979; Bushman, 1993; Dallman *et al*., 1991). The mutant sc repressors RR*69 and R*R*69 were most effective when they interacted with their respective cognate operators. This indicates that our system is capable of detecting specific interactions, although the accurate data evaluation is somewhat obscured by the high level repressor expression in this system, and consequently, by the possible repressor dimerization at closely related or half-cognate operators, as discussed above. The level of repression in this system was generally about two- to four-fold. This low value may be due to ''titration out'' of the repressors by the multicopy operators. However, the level of repression in artificial assay systems seems to depend mainly on the repressor itself: the *lac* repressor showed a minimum of 200-fold repression in a two-plasmid system (Lehming *et al*., 1987), while the 434 repressor (Wharton & Ptashne, 1987), the Arc repressor and its covalent dimer (Robinson & Sauer, 1996) showed only a modest few-fold repression of single-copy reporter genes. Nevertheless, such a low repression level proved to be sufficient to isolate a new specificity 434 repressor mutant showing at least 50 to 150 fold operator discrimination *in vitro* (Wharton & Ptashne, 1987). The sc repressors of this study showed a similar degree of *in vitro* discrimination.

The major finding of this work is that both the wild-type and engineered DNA-binding domains of the 434 repressor can be covalently dimerized to form functional sc repressors which show selective high affinity recognition of DNA operators composed of the respective subsites of the joined domains. To obtain engineered DNA-binding domains in the sc framework we used the principle of the α-helix redesign experiment (Wharton & Ptashne, 1985) and showed that the obtained homo- and heterodimeric sc repressors behaved similarly to their naturally dimerized counterparts (Wharton & Ptashne, 1985; Hollis *et al*., 1988; Webster *et al*., 1992). A potential advantage of the sc heterodimer is that unlike the non-covalent heterodimer, which exists in a mixture with two homodimers, it is homogeneous. This permits a better study of DNA-binding properties, for example by binding site selection from randomized DNA pools (J. C., S. P. & A. S., unpublished results). Moreover, the sc framework allows new approaches to be applied in the search for altered recognition specificities. A previous study showed that a combinatorial mutant library of the natural 434 repressor selected against natural or symmetrically altered mutant operators did not provide mutant repressors with new high affinity recognition properties (Hu *et al*., 1994). It is possible that weaker, but specific interactions are overlooked in this system, as simultaneous binding of two weakly interacting subunits may form an unstable complex with DNA. This problem may be overcome by using the sc framework, as it provides the possibility of altering only one of the domains whilst keeping the other one unchanged. The unchanged domain can provide a supporting interaction with operators containing a half-site cognate to it. Selection of a sc repressor library containing one unchanged domain and one partially randomized domain against operators containing a cognate half-site to the unchanged domain can identify sc repressor mutants with high overall affinity, even if the intrinsic affinity of the isolated mutant domain toward the target half-site is not very high. For the selection of such libraries it is possible to use the *in vivo* detection system described here to detect phenotypic differences on indicator plates (A. S., S. W., I. T., J. C. & S. P., unpublished results). Alternatively, more straightforward, direct genetic selection techniques (Elledge *et al*., 1989; Mossing *et al*., 1991) could be applied. The possibility to create combinatorial libraries with amino acid residue changes in one of the domains of the sc dimer, and the availability of *in vivo* selection methods together may provide a system which, similarly to the zinc-finger phage display systems (Rebar & Pabo, 1994; Jamieson *et al*., 1994; Choo & Klug, 1994a,b; Wu *et al*., 1995), could be used to derive further recognition rules and to generate new DNA-binding specificities in a given domain framework.

Recently, the covalent dimerization approach has also been applied to the full-length subunits of the Arc repressor and even an enhanced DNA-binding affinity as well as *in vivo* repressor activity compared to the non-covalent repressor were observed (Robinson & Sauer, 1996). The Arc and 434 repressors, and consequently the corresponding sc molecules, are unrelated DNA-binding proteins in terms of molecular architecture and DNA recognition mechanism. One major difference is that the covalent Arc dimers, as binding units, interact with operator half-sites forming dimers of dimers on a full operator, while the domains of the covalent 434 dimer interact each with one half-site, in a bidentate manner. Despite the differences both sc proteins may provide new experimental approaches in studying the roles of protein-protein and protein-DNA contacts in protein-DNA complexes.

Materials and Methods

Materials

Restriction endonucleases and DNA modifying enzymes were obtained from New England Biolabs, Inc., Boehringer Mannheim, Pharmacia Biotech, GIBCO BRL (Life Technologies, Inc.), Promega and Perkin Elmer. Oligonucleotides were prepared by the ICGEB oligonucleotide synthesis service or by Primm s.r.l. (Milan, Italy) and were used without purification. Radiochemicals were from Amersham, the T7 sequencing kit from Pharmacia Biotech. Protein and DNA molecular weight markers were from Pharmacia Biotech. All other chemicals were from Merck, Sigma, Aldrich or Boehringer Mannheim.

Bacterial strains and vectors

XL1-Blue (from Stratagene) was used as host during the vector construction steps and in the *in vivo* repressor-operator interaction studies, the genotype is *sup*E44, *hsd*R17, *rec*A1, *end*A1, *gyr*A96, *thi*, *rel*A1, *lac*, F′[*pro*AB, *lac*I^{q}ZΔM15, Tn*10*(*tet*^{r})] (Bullock *et al.*, 1987). The BL21(DE3)pLysS expression host (Studier *et al.*, 1990) was obtained from Novagen. CJ236, used to isolate uracil-containing DNA template for site-directed mutagenesis was from Bio-Rad, helper phage VCSM13 was from Stratagene. The pRIZ′ (named pRIZ′O$_{lac}$ in this work) key vector for the construction of repressor and operator clones was constructed by Simoncsits *et al.*, (1994). The polycloning region of this vector is *Nsi*I (at the −10 box of the *rrn*B P2 promoter)-*Nco*I (containing translational initiation codon)-*Bam*HI-*Sal*I-*Eco*RI-*Cla*I-*Pst*I-*Xba*I-*Bgl*II-*Hin*dIII, the sites used for repressor gene construction and cloning are underlined. The *Nsi*I site was used for repressor shuffling between different operator clones. M13mp18 (Yanisch-Perron *et al.*, 1985) was from Pharmacia Biotech. Phagemid pKZ152 (Tjörnhammar & Simoncsits, 1991) was used for cloning and mutagenesis of the *lac*I^q-*lac*pro-*lacZ*′(1-146) cluster at the *lac* operator site. The source of this cluster was pMC9 (Miller *et al.*, 1984) isolated from strain Y1089 (Huynh *et al.*, 1984), which was obtained from Stratagene. pRSET expression vectors were donated by R. Schoepfer (Schoepfer, 1993) and pET expression vectors (Studier *et al.*, 1990) were from Novagen. The source of the 434 repressor gene was the λgt10 vector (Boehringer Mannheim), which contains the *imm*^{434} region (Huynh *et al.*, 1984).

General techniques

The guidelines of published methods: recombinant DNA techniques (Sambrook *et al.*, 1989), site-directed mutagenesis (Kunkel *et al.*, 1987), DNA sequencing (Sanger *et al.*, 1977), β-galactosidase assay (Miller, 1972), EMSA or gel retardation (Carey, 1991), DNase I protection assay (Johnson *et al.*, 1979) and base-specific chemical cleavage of DNA (Maxam & Gilbert, 1980) were followed.

Vector constructions for *in vivo* repressor-operator interaction studies: operator cloning and construction of repressor genes in pRIZ′ vectors

Operator cloning

First, the *lac* operator region of pRIZ′O$_{lac}$ was replaced by a *Nde*I site to obtain pRIZ′O(−), which was then used to clone synthetic oligonucleotides containing operator sequences. The pRIZ′O$_{lac}$ phagemid contains two *lac* operator sites, therefore it could not be converted to pRIZ′O(−) by site-directed mutagenesis. pRIZ′O(−) was therefore obtained as follows. The 1.72 kb *lac*I^q-*lac*pro-*lacZ*′(1-146) region of pMC9 (Miller *et al.*, 1984) was cloned as an *Eco*RI fragment into pKZ152 phagemid, which does not contain *lac* regions (Tjörnhammar & Simoncsits, 1991), then the *lac* operator region was replaced by the *Nde*I site using the mutagenic oligonucleotide: TTCCGGCTCGTATGTTGCATATGAGGAAA-CAGCTATGACCAT (the *Nde*I site is underlined). The altered cluster was then cloned as a blunt-end fragment (obtained after *Eco*RI cleavage and Klenow polymerase plus dNTP treatment) into the precursor of the pRIZ′ vector as described (Simoncsits *et al.*, 1994) to obtain pRIZ′O(−), in which the gene cluster was in the same orientation as in pRIZ′O$_{lac}$ (Figure 1(c)). The following synthetic operators with *Nde*I compatible cohesive ends were cloned into the unique *Nde*I site of pRIZ′O(−):

```
TACAAGAAAGTTTGT          O_R1 (of 434)
  GTTCTTTCAAACAAT

TACAATAAAACTTAAA         O_R*1 (434-P22) or
  GTTATTTTGAATTTAT       O_R*2 (P22-434), in opposite orientation

TATTAAAGAACACTTAAA       O_P1 (O_R1 of P22)
  AATTTCTTGTGAATTTAT
```

where underlined letters show the respective consensus operator boxes. The vectors obtained are abbreviated as pRIZ′O$_R$1, pRIZ′O$_{R*}$1, pRIZ′O$_{R*}$2 and pRIZ′O$_P$1. Altered spacer operator mutants were cloned similarly, the spacer sequences of the O$_R$1 mutants (shown for upper strand) were GAAAG (O$_R$1, 5 bp), GAAAGAT (O$_R$1, 7 bp), GAAAGTAT (O$_R$1, 8 bp) and GAAAGTATAT (O$_R$1, 10 bp), the spacer sequence of the O$_{R*}$1 mutant was TAAAAT (O$_{R*}$1, 6 bp). The operator clones were verified by sequencing; in most cases both possible orientations with respect to the *lac* promoter were obtained. The orientations shown in Figure 1(d) or those corresponding to the above listed mutant spacer sequences were used in this study.

Construction of repressor genes in pRIZ′ vectors

In the initial stages of this work R$_{1-69}$L$_{70-89}$R$_{1-90}$ (or RR90) and R$_{1-90}$ (or R90) were constructed. These elongated proteins showed similar DNA-binding properties to that of the shorter counterparts but they showed enhanced proteolytic sensitivity during expression and purification and were not used in this study. Their genes, however, were used as intermedieres during the construction of the shortened versions as described below.

For gene construction and cloning, regions of the cI gene of the λgt10 vector (Huynh *et al.*, 1984) were amplified by PCR, using primers designed according to the published nucleotide sequence (Nikolnikov *et al.*, 1984). PCR products, when blunt-end cloning was used, were treated with Klenow polymerase and dNTP. Primer sequences are written in the 5′ to 3′ direction. The R90 region was amplified by using two different pairs of primers. To obtain the first copy of the tandem repeats, primers TCCTTTCATGAGTATTTCTTCCAGGGT (AT405, *Rca*I site underlined) and **TCA**GGATCCAGCTCTAAC-CATGCTAAT (AT406, *Bam*HI site underlined, TCA stop codon complementer in bold) were used and the PCR product was cloned as a *Rca*I-blunt-end fragment into the *Nco*I-blunt-end(*Bam*HI) pRIZ′O$_R$1 to obtain pRIZ′O$_R$1R90. The second copy was obtained by PCR using primers TACTTGGATCCATTCTTCCAGGG-

TAAAAAGC (AT407, *Bam*HI site underlined) and CTGCTCAAGCTTCACGAACCAGCTCTAACCAT (AT408, *Hin*dIII site underlined, stop codon complementer in bold), which was cloned as *Bam*HI-*Hin*dIII fragment into M13mp18. Site-directed mutagenesis of the mp18 clone was performed with primers CTTAGTTTTACCGTTCTCGAGCTGCTCT (AT409, *Xho*I site underlined, mismatched base in bold) and TAGACTGCTGGGTGGTACCCACCTTTTGAG (AT410, *Kpn*I site is underlined, mismatched bases in bold) in one step as these two primers could anneal contiguously to the template. This second mutant copy was then cloned as a *Bam*HI-*Hin*dIII fragment into pRIZ′O_R1R90 to obtain pRIZ′O_R1RR90, which contains *Kpn*I and *Xho*I sites in the second copy, near the borders of the α3 helix coding region (at the *Bam*HI fusion site between the repeats, Ser90 of the first R90 coincides with Ser1 of the second R90, therefore the RR90 abbreviation stands for $R_{1\text{-}89}R_{1\text{-}90}$). The R69 coding region of the mutated mp18 clone was PCR amplified with primers AT407 and TCATCTAACATTCGAATCAGAGGT (AT414, stop codon complementer in bold), the PCR product was cleaved with *Bam*HI and cloned into the *Bam*HI-blunt-end (*Hin*dIII) pRIZ′O_R1R90 vector to obtain pRIZ′O_R1RR69, in which the *Hin*dIII site downstream of the stop codon is regenerated. The α3 helix coding region in the second repeats of pRIZ′O_R1RR90 and pRIZ′O_R1RR69 was replaced with a *Kpn*I-*Xho*I linker, (coding for amino acid substitutions as in Figure 1(b).)

CTCTAACGTCAGTATCTCACAGC
CATGGAGATTGCAGTCATAGAGTGTCGAGCT

to obtain pRIZ′O_R1RR*90 and pRIZ′O_R1RR*69, respectively. To obtain the R*R*69 gene, the second repeat of RR*90 was used as follows. Complete *Eco*RI cleavage (the *Eco*RI site is present in both repeats at amino acid residues 10 to 12, see Figure 1(a)) of pRIZ′O_R1RR*90 followed by religation eliminated the first repeat and resulted in pRIZ′O_R1R*90. This vector was used as PCR template with primers TGTAGCGGGAAGGCGTATTAT (AS107, vector-specific primer at the *rrnB* P2 promoter overlapping the unique *Nsi*I site of pRIZ′) and AT406, then the *Nsi*I-*Bam*HI-cleaved PCR product was cloned into pRIZ′O_R1RR*69, replacing the first domain coding region of RR*69 with R* and providing pRIZ′O_R1R*R*69. The pRIZ′O_R1R69 vector was obtained from pRIZ′O_R1RR69 after complete *Eco*RI cleavage and religation. pRIZ′O_R1R(−) was also obtained from *Eco*RI cleaved pRIZ′O_R1RR69, but Klenow polymerase treatment in the presence of dNTP preceded the religation. The full 434 repressor coding region was obtained from λgt10 by PCR performed with primers AT405 and TCTCTGGATCCTCATACGAATTTTACCCTCGCT (AT453, *Bam*HI site underlined, stop codon complementer in bold), cleaved with *Rca*I and *Bam*HI and cloned into *Nco*I-*Bam*HI-cleaved pRIZ′ to obtain pRIZ′O_R1cI. All repressor coding regions obtained by PCR were verified by nucleotide sequencing, using AS107 and GGCAGTTTCCCAGACATTACTC (AT419, backward vector-specific primer located downstream of the *Hin*dIII site) flanking primers and sometimes internal primers. The desired repressor-operator combinations in the same pRIZ′ vector were obtained by subcloning the repressor coding regions into other operator vectors (shuffling) by using the unique restriction site of the pRIZ′ vectors (see Figure 1(c), usually *Nsi*I-*Hin*dIII or sometimes *Pst*I-*Hin*dIII sites).

In vivo detection of repressor-operator interactions: β-galactosidase assay

XL1-Blue *E. coli* cells containing pRIZ′ vectors were grown in LB medium containing 100 mg/l ampicillin and 10 mg/l tetracycline at 37°C with shaking. Overnight cultures were diluted 50-fold into fresh medium and grown for two hours (A_{600} is approximately 0.3 to 0.4). IPTG was then added to a final concentration of 1 mM and the cultures were grown for 2.5 hours. β-Galactosidase assays were performed as described (Miller, 1972).

Construction of T7 promoter-based expression vectors

The pRSETRR69 vector was described by Percipalle *et al.*, (1995a). Large scale expression proved to be difficult with this vector, but an even more serious problem was encountered in the case of the pRSETRR*69 vector, which did not provide productive colonies upon transformation into BL21(DE3)pLysS or pLysE expression strains. Therefore we constructed an improved vector, called pSET5a from pRSET5a (Schoepfer, 1993) and pET16b (Novagen) vectors by cloning the *Sca*I-*Xba*I fragment of pRSET5a into *Sca*I-*Xba*I cleaved pET16b. The new pSET5a vector combines the advantageous properties of the parents pET16b (lower copy number and tightly controlled T7*lac* promoter) and pRSET5a (versatile multiple cloning site and phagemid properties). The RR69 region of pRSETRR69 was cloned as an *Xba*I-*Hin*dIII fragment into pSET5a to obtain pSETRR69. pSETRR*69 was obtained by replacing R69 of pSETRR69 with R*69 in *Bam*HI-*Hin*dIII cloning. To obtain pSETR*R*69, similar intermediate clones were used as described for the corresponding pRIZ′ construct. Briefly, the R*90 coding region from pRIZ′O_R1RR*90 replaced the R69 part of pSETRR69 by *Bam*HI-*Hin*αdIII cloning to provide pSETRR*90, which was converted to pSETR*90 by complete *Eco*RI cleavage followed by religation. PCR was then performed on the pSETR*90 template using the T7 promoter primer (Novagen) and AT406, then the *Xba*I-*Bam*HI-cleaved PCR product was cloned into pSETRR*69 to obtain pSETR*R*69. The pSETR69 vector was obtained from pSETRR69 after complete *Eco*RI cleavage and religation. The full-length 434 repressor gene in the pET vector was obtained by PCR performed on the λgt10 vector with primers AT405 and AT453, followed by cloning the *Rca*I-*Bam*HI cleaved PCR product into *Nco*I-*Bam*HI-cleaved pET16b.

Expression and purification of sc repressors

pSETRR69 was freshly transformed into the BL21(DE3)-pLysS strain to obtain about 1000 to 2000 small colonies on LB plates containing 75 mg/l ampicillin and 25 mg/l chloramphenicol after 12 to 14 hours incubation at 37°C. The colonies were suspended and grown in 3.6 l of LB medium containing antibiotics as described above. Induction with IPTG was performed as described (Studier *et al.*, 1990). After two hours of induction, cells were harvested by centrifugation, resuspended in 120 ml of TE buffer (10 mM Tris-HCl, 2 mM EDTA (pH 8.0)), frozen at −80°C and thawed. The suspension was sonicated briefly to reduce the viscosity, centrifuged and batch adsorption was performed on the supernatant by adding 25 ml of 50 % suspension of SP-Sepharose in TE followed by gentle shaking of the suspension for 20 minutes. A short column was then prepared which was washed

with 150 mM KCl in TE followed by 350 mM KCl in TE to elute the highly enriched RR69. The eluate was diluted threefold with TE and purified (in two portions) on a Mono S HR 10/10 column (Pharmacia Biotech) using a linear gradient of KCl in TE. The isolated yield was 40 mg per liter culture. RR*69 and R*R*69 were similarly purified using a Resource S column in 20 and 6 mg per liter culture yields, respectively. In these latter cases the expression levels were lower and a fraction of the repressors was found to be insoluble. Only the soluble fractions were used in the further purification steps.

The 434 repressor and its DNA-binding domain R69 were also purified as described above with minor modifications.

Electrophoretic mobility shift assay

Radioactively labelled DNA probes of 157 (O_R1), 158 ($O_{R*}1$) and 160 (O_P1) bp long containing operator sites were generated by PCR on the corresponding pRIZ′ operator vectors with primers TAGCTCACTCATTAGGCACC (AT404, located upstream of the *lac* promoter −35 box) and GTAACGCCAGGGTTTTCCCAGT (AS181, backward primer, located in *lacZ*) of which one (usually AS181) was ^{32}P end-labelled. Fifteen cycles (94°C, 58°C and 72°C, 30 seconds each) were performed and the radioactive PCR products were purified using the QIAquick PCR purification kit (Qiagen). The maximal DNA probe concentrations used in EMSA and DNase I protection experiments were calculated on the basis of the molar amounts of PCR primers, assuming 100% incorporation and purification yields.

Binding reactions were performed in 1× binding buffer (200 mM KCl, 2.5 mM $MgCl_2$, 1 mM $CaCl_2$, 0.1 mM EDTA, 25 mM Tris-HCl (pH 7.0), 6% (v/v) glycerol) containing 2.5 μg/ml sonicated salmon sperm DNA, 100 μg/ml bovine serum albumin, <20 pM labelled DNA probe and bromophenol blue at the minimal visible concentration, and repressors in different concentrations. For titration experiments, large volumes of binding stocks were prepared (without repressor), and to 45 μl aliquots of this stock, 5 μl of sc repressor (diluted into 1× binding buffer to obtain ten times higher protein concentrations than required for each step) was added. In case of the natural repressor, however, the highest dilution was 50 nM and different volumes were added to the binding stocks to obtain the indicated final concentrations given in Figure 3(b). The mixtures were incubated at room temperature for one hour, then identical volumes were loaded onto a running 8% (w/v) polyacrylamide gel (29:1 acrylamide/bisacrylamide), prerun at 4°C in 0.5× TBE buffer (45 mM Tris-borate, 1 mM EDTA, pH 8.0) and the electrophoresis was performed with 25V/cm at 4°C for two hours. The gels were fixed in 10% (v/v) acetic acid and dried before autoradiography.

Competition experiments

Binding reactions containing 5 nM repressor were performed as described above. After one hour, aliquots were taken out and double-stranded oligonucleotide competitors containing O_R1, $O_{R*}1$ or O_P1 operator sequences were added to a final concentration of 100 nM. The reaction mixtures were incubated at room temperature for 30 minutes before analysis on gels as described above. Competitor oligonucleotide duplexes were prepared from the listed cohesive end duplexes (see above) by a Klenow polymerase fill-in reaction followed by purification on a non-denaturing polyacrylamide gel.

DNase I protection

Binding reactions containing 100 or 200 nM repressor were performed as described above in 100 μl volumes. DNase I (10 ng) was added and the mixtures were incubated at room temperature for five minutes. 100 μl of cold stop solution (4 M ammonium acetate, 40 mM EDTA containing 200 μg/ml glycogen) was added and further steps were as described (Johnson *et al*., 1979) using 6% (w/v) acrylamide/8M urea gel.

Acknowledgements

We thank Professors A.Falaschi and F.E.Baralle (ICGEB) for their support and critical comments on this work. Thanks are due to Drs S.Zakhariev and C.Guarnaccia for their help with protein purification, to Dr D. Rhodes for useful comments on an early version of the manuscript, to K.Vlahovicek for the construction of the Figures, to Ms S. Kerbavcic and Dr M.-L. Tjörnhammar for their help with the preparation of the manuscript. J.C. and P.P. were supported by predoctoral fellowships of SISSA (International School of Advanced Studies, Trieste), S.W. and I.T. were recipients of ICGEB fellowships. A.S. is on leave from the Department of Neurochemistry, The Arrhenius Laboratories for Natural Sciences, Stockholm University, Sweden.

References

Aggarwal, A. K., Rodger, D. W., Drottar, M., Ptashne, M. & Harrison, S. C. (1988). Recognition of DNA operator by the repressor of phage 434: a view at high resolution. *Science,* **242**, 899–907.

Anderson, J. E., Ptashne, M. & Harrison, S. C. (1987). Structure of the repressor-operator complex of bacteriophage 434. *Nature,* **326**, 846–852.

Bell, A. C. & Koudelka, G. B. (1993). Operator sequence context influences amino-acid–base-pair interactions in 434 repressor-operator complexes. *J. Mol. Biol.* **234**, 542–553.

Berg, J. M. & Shi, Y. (1996). Galvanization of biology: a growing appreciation for the roles of zinc. *Science,* **271**, 1081–1085.

Brown, B. M. & Sauer, R. T. (1993). Assembly of the Arc repressor-operator complex: cooperative interactions between DNA-bound dimers. *Biochemistry,* **32**, 1354–1363.

Bullock, W. O., Fernandez, J. M. & Short, J. M. (1987). XL1-Blue: a high efficiency plasmid transforming *recA Escherichia coli* strain with beta-galactosidase selection. *BioTechniques,* **5**, 376–380.

Bushman, F. D. (1993). The bacteriophage 434 right operator. Roles of O_R1, O_R2 and O_R3. *J. Mol. Biol.* **230**, 28–40.

Carey, J. (1991). Gel retardation. *Methods Enzymol.* **208**, 103–117.

Carlson, P. A. & Koudelka, G. B. (1994). Expression, purification and functional characterisation of the carboxyl-terminal domain fragment of bacteriophage 434 repressor. *J. Bacteriol.* **176**, 6907–6914.

Choo, Y. & Klug, A. (1994a). Toward a code for interaction of zinc fingers with DNA: selection of ran-

domized fingers displayed on phage. *Proc. Natl Acad. Sci. USA,* **91**, 11163–11167.

Choo, Y. & Klug, A. (1994b). Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. *Proc. Natl Acad. Sci. USA,* **91**, 11168–11172.

Dallman, G., Marincs, F., Papp, P., Gaszner, M. & Orosz, L. (1991). The isolated N-terminal domain of the c repressor of bacterophage *16-3* is functional in DNA binding *in vivo* and *in vitro*. *Mol. Gen. Genet.* **227**, 106–112.

Ebright, R. H., Cossart, P., Gicquel-Sanzey, B. & Beckwith, J. (1984). Mutations that alter the DNA sequence specificity of the catabolite gene activator protein of *E. coli*. *Nature,* **311**, 232–235.

Elledge, S. J., Sugiono, P., Guarente, L. & Davis, R. W. (1989). Genetic selection for genes encoding sequence-specific DNA-binding proteins. *Proc. Natl Acad. Sci. USA,* **86**, 3689–3693.

Fairall, L., Harrison, S. D., Travers, A. A. & Rhodes, D. (1992). Sequence-specific DNA binding by a two zinc-finger peptide from the *Drosophila melanogaster* Tramtrack protein. *J. Mol. Biol.* **226**, 349–366.

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

Harrison, S. C. & Aggarwal, A. K. (1990). DNA recognition by proteins with the helix-turn-helix motif. *Annu. Rev. Biochem.* **59**, 933–969.

Hollis, M., Valenzula, D., Pioli, D., Wharton, R. & Ptashne, M. (1988). A repressor heterodimer binds to a chimeric operator. *Proc. Natl Acad. Sci. USA,* **85**, 5834–5838.

Hu, J. C. (1995). Repressor fusions as a tool to study protein-protein interactions. *Structure,* **3**, 431–433.

Hu, L., Sera, T. & Schultz, P. G. (1994). A permutational approach toward protein-DNA recognition. *Proc. Natl Acad. Sci. USA,* **91**, 3969–3973.

Huston, J. S., Levinson, D., Mudgett-Hunter, M., Tai, M.-S., Novotny, J., Margolies, M. N., Ridge, R. J., Bruccoleri, R. E., Haber, E., Crea, R. & Opperman, H. (1988). Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in *Escherichia coli*. *Proc. Natl Acad. Sci. USA,* **85**, 5879–5883.

Huynh, T. V., Young, R. A. & Davis, R. W. (1984). Constructing and screening cDNA libraries in λgt10 and λgt11. In *DNA Cloning: A Practical Approach* (Glower, D. M., ed.), vol. 1, pp. 49–78, IRL Press, Oxford.

Jamieson, A. C., Kim, S.-H. & Wells, J. A. (1994). *In vitro* selection of zinc fingers with altered DNA-binding specificity. *Biochemistry,* **33**, 5689–5695.

Johnson, A. D., Meyer, B. J. & Ptashne, M. (1979). Interaction between DNA-bound repressors govern regulation by the λ phage repressor. *Proc. Natl Acad. Sci. USA,* **76**, 5061–5065.

Kim, Y.-I. & Hu, J. C. (1995). Operator binding by λ repressor heterodimers with one or two N-terminal arms. *Proc. Natl Acad. Sci. USA,* **92**, 7510–7514.

Koudelka, G. B. & Lam, C.-Y. (1993). Differential recognition of O_R1 and O_R3 by bacteriophage 434 repressor and Cro. *J. Biol. Chem.* **268**, 23812–23817.

Kunkel, T., Roberts, J. & Zakour, R. (1987). Rapid and efficient site specific mutagenesis without phenotypic selection. *Methods Enzymol.* **154**, 367–382.

Lehming, N., Sartorius, J., Niemöller, M., Genenger, G., von Wilcken-Bergmann, B. & Müller-Hill, B. (1987). The interaction of the recognition helix of *lac* repressor with *lac* operator. *EMBO J.* **6**, 3145–3153.

Luisi, B. F., Xu, W. X., Otwinowski, Z., Freedman, L. P., Yamamoto, K. R. & Siegler, P. B. (1991). Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature,* **352**, 497–505.

Lukacsovich, T., Orosz, A., Baliko, G. & Venetianer, P. (1990). A family of expression vectors based on the *rrnB* P2 promoter of *Escherichia coli*. *J. Biotechnol.* **16**, 49–56.

Maxam, A. M. & Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**, 499–560.

Miller, J. H. (1972). In *Experiments in Molecular Genetics,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Miller, J. H., Lebkowski, J. S., Greisen, K. S. & Calos, M. P. (1984). Specificity of mutations induced in transfected DNA by mammalian cells. *EMBO J.* **3**, 3117–3121.

Mossing, M. C., Bowie, J. U. & Sauer, R. T. (1991). A streptomycin selection for DNA-binding activity. *Methods Enzymol.* **208**, 604–619.

Nikolnikov, S., Pósfai, G. & Sain, B. (1984). The construction of a versatile plasmid vector that allows direct selection of fragments cloned into six unique sites of the *cI* gene of coliphage 434. *Gene,* **30**, 261–265.

Pabo, C. O. & Sauer, R. T. (1992). Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **61**, 1053–1095.

Percipalle, P., Simoncsits, A., Zakhariev, S., Guarnaccia, C., Sanchez, R. & Pongor, S. (1995a). Rationally designed helix-turn-helix proteins and their conformational changes upon DNA binding. *EMBO J.* **14**, 3200–3205.

Percipalle, P., Zakhariev, S., Tossi, A., Guarnaccia, C., Cserzö, M., Simoncsits, A. & Pongor, S. (1995b). Synthetic dimeric peptides based on the phage 434 *cI* repressor. In *Peptides '94* (Maia, H. L. S., ed.), pp. 391–392, ESCOM, Netherlands.

Pomerantz, J. L., Sharp, P. A. & Pabo, C. O. (1995). Structure-based design of transcription factors. *Science,* **267**, 93–96.

Poteete, A. R., Ptashne, M., Balliret, M. & Eisen, H. (1980). Operator sequences of bacteriophages P22 and 21. *J. Mol. Biol.* **137**, 81–91.

Ptashne, M. (1992). In *A Genetic Switch,* Cell Press and Blackwell Scientific Publications, Cambridge, MA.

Rebar, E. J. & Pabo, C. O. (1994). Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science,* **263**, 671–673.

Rhodes, D., Schwabe, J. W. R., Chapman, L. & Fairall, L. (1996). Towards understanding of protein-DNA recognition. *Phil. Trans. Roy. Soc. ser. B.* **351**, 501–509.

Robinson, C. R. & Sauer, R. T. (1996). Covalent attachment of Arc repressor subunits by a peptide linker enhances affinity for operator DNA. *Biochemistry,* **35**, 109–116.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). In *Molecular Cloning: A Laboratory Manual,* 2nd edit., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain terminating inhibitors. *Proc. Natl Acad. Sci. USA,* **74**, 5463–5467.

Sauer, R. T., Pabo, C. O., Meyer, B. J., Ptashne, M. & Backman, K. D. (1979). Regulatory functions of λ repressor reside in the amino-terminal domain. *Nature,* **279**, 396–400.

Schoepfer, R. (1993). The pRSET family of T7 promoter expression vectors for *Escherichia coli*. *Gene,* **124**, 83–85.

Simoncsits, A., Tjörnhammar, M. L., Kálmán, M., Cserpán, I., Gafvelin, G. & Bartfai, T. (1988). Synthesis, cloning and expression in *Escherichia coli* of artificial genes coding for biologically active elongated precursors of the vasoactive intestinal polypeptide. *Eur. J. Biochem.* **178**, 343–350.

Simoncsits, A., Bristulf, J., Tjörnhammar, M. L., Cserzö, M., Pongor, S., Rybakina, E., Gatti, S. & Bartfai, T. (1994). Deletion mutants of human interleukin 1β with significantly reduced agonist properties: search for agonist/antagonist switch in ligands to the interleukin 1 receptors. *Cytokine,* **6**, 206–214.

Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubbendorf, J. W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60–89.

Tjörnhammar, M. L. & Simoncsits, A. (1991). Direct selection vectors, promoter-active adaptors and their use for gene assembly by a single-stranded cloning method. *Nucl. Acids Res. Symp. Series,* **24**, 252.

Ullmann, A., Jacob, F. & Monod, J. (1967). Characterization by *in vitro* complementation of a peptide corresponding to an operator-proximal segment of the β-galactosidase structural gene of *Escherichia coli*. *J. Mol. Biol.* **24**, 339–343.

Webster, C., Merryweather, A. & Brammar, W. (1992). Efficient repression by a heterodimeric repressor in *Escherichia coli*. *Mol. Microbiol.* **6**, 371–377.

Wharton, R. P. & Ptashne, M. (1985). Changing the binding specificity of a repressor by redesigning an α-helix. *Nature,* **316**, 601–605.

Wharton, R. P. & Ptashne, M. (1987). A new specificity mutant of 434 repressor that defines an amino acid-base pair contact. *Nature,* **326**, 888–891.

Wharton, R. P., Brown, E. L. & Ptashne, M. (1984). Substituting an α-helix switches the sequence-specific DNA interactions of a repressor. *Cell,* **38**, 361–369.

Wong, I. & Lohman, T. M. (1995). Linkage of protein assembly to protein-DNA binding. *Methods Enzymol.* **259**, 95–127.

Wu, H., Yang, W.-P. & Barbas, C. F., III. (1995). Building zinc fingers by selection: toward a therapeutic application. *Proc. Natl Acad. Sci. USA,* **92**, 344–348.

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene,* **33**, 103–119.

Edited by T. Richmond

(Received 27 September 1996; received in revised form 29 November 1996; accepted 4 December 1996)