3474–3480 *Nucleic Acids Research, 1999, Vol. 27, No. 17*

Isolation of altered specificity mutants of the single-chain 434 repressor that recognize asymmetric DNA sequences containing the TTAA and TTAC subsites

András Simoncsits*, Marie-Louise Tjörnhammar, Shenglun Wang and Sándor Pongor

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

Received May 24, 1999; Revised and Accepted July 16, 1999

ABSTRACT

A novel single-chain (sc) protein framework containing covalently dimerized DNA-binding domains (DBD) of the phage 434 repressor was used to construct combinatorial mutant libraries in order to isolate mutant DBDs with altered specificities. The library members contain one wild-type DBD and one mutant domain with randomized amino acids in the DNA-contacting region. Based on previous studies, the mutant sc derivatives are expected to recognize a general ACAA–6 bp–NNNN sequence, where ACAA is contacted by the wild-type and NNNN by the mutant domain. In principle, any sequence can stand for NNNN and serve as a selection target. Here an *in vivo* library screening method was used to isolate mutant sc repressors that interact with an asymmetric operator containing the TTAA target. Several mutants showed high affinity *in vitro* binding to operators containing the target and strong (up to 80-fold) preference for the TTAA target over the wild-type TTGT. Specificity studies revealed that certain mutants bound with substantially higher affinities (K_d ~10^{-11} M) to operators containing the TTAC sequence, a close homolog of the TTAA target. Thus, we have fortuitously isolated mutant sc repressors that show up to a several hundred-fold preference for TTAC over TTGT.

INTRODUCTION

Altering the binding specificities of DNA-binding proteins is a challenging research area with the potential application of developing novel reagents and therapeutic agents of desired specificities. Significant progress has recently been achieved toward this goal by using the Cys2His2 type zinc finger motif, due to its modular nature and the applicability of the filamentous phage display and affinity selection techniques to this type of DNA-binding motif (1–10; for an overview see also 11–15). Earlier (16–19) and more recent examples (20,21) show that the helix–turn–helix (HTH) motif of prokaryotic and phage transcription factors can also accommodate certain changes to obtain altered DNA binding specificities. Most of these proteins, however, bind as homodimers and recognize operators containing symmetrically related half-sites (22,23). This property clearly poses limitations since a specificity change is likely to result in recognition of an altered but again symmetric target sequence. Asymmetric sequences can be recognized by heterodimers containing wild-type and altered recognition specificity mutant monomer units (24–26). Such heterodimers exist in equilibrium mixture together with two homodimers, which is a complicating factor in library-based selections for binding to asymmetric target DNA.

These limitations can, in principle, be circumvented by using covalently linked DNA-binding domains (DBDs) and by performing independent specificity changes in the individual DBDs. It was shown previously that a well-studied member of the HTH family, the 434 repressor, can be engineered to form a single-chain (sc) protein by covalently linking two DBDs through a recombinant peptide linker (27,28). Binding studies, both *in vitro* and *in vivo*, showed that the covalent linker could functionally replace the bulky non-covalent dimerization domain of the natural repressor (28). This class of novel protein was termed sc repressor and the prototype homodimeric member containing two wild-type 434 domains was abbreviated as RR69. By substituting the DNA-contacting amino acids at the –1, 1, 2 and 5 positions of the α3 recognition helix with the corresponding residues of the related P22 repressor, a heterodimeric (RR*69, changes in one domain only) or a new homodimeric (R*R*69, identical changes in both domains) sc repressor was obtained (28). DNA recognition studies showed that the wild-type domain of the sc repressors recognized the consensus ACAA sequence found in the natural 434 operators while the R* domain preferentially recognized the TTAA sequence (28,29). These 4 bp long 'contacted' regions form the core of the operator half-sites and are separated by a 6 bp spacer or 'non-contacted' region in the optimum DNA ligands. The spacer sequence was also shown to strongly influence the DNA binding affinity of the sc repressors (29), indicating that the indirect effects of the spacer sequence on DNA binding are similar to those observed with the natural 434 repressor or with its DBD (30). In other words, only a limited set of spacer sequences can support high affinity DNA binding. These findings

*To whom correspondence should be addressed. Tel: +39 40 3757342; Fax: +39 40 226555; Email: simoncs@icgeb.trieste.it
Present address:
Shenglun Wang, Center for Advanced Research in Biotechnology, 9600 Gudelsky Drive, Rockville, MD 20850, USA

suggest that mutant sc 434 repressors can be built from previously isolated DBDs of known specificities. These molecules could be targeted to relatively long (14 bp) DNA sequences or could be used in gene fusions to develop artificial transcription factors exhibiting novel specificities.

There are only a few examples for changing the binding specificity of the 434 repressor (17,18,31) which could be used in the domains of the mutant sc repressors. A permutational approach using a combinatorial library of the 434 repressor and an *in vivo* selection technique did not result in new binding specificities that were detectable by *in vitro* assays (32). We have previously suggested that combinatorial libraries of the sc 434 repressor could be used to explore new DNA binding specificities in a different manner (28). In such libraries, one domain could be kept unchanged whilst the other one is randomized at certain DNA-contacting positions. The library members are expected to bind to the general ACAA–6 bp–NNNN sequence so that the wild-type 434 domain binds to its cognate ACAA sequence providing an initial, supporting interaction for the mutant domain to approach and recognize the NNNN target located 6 bp away. We constructed such libraries in *Escherichia coli* and used an *in vivo* library screening technique to identify mutants that interact with an asymmetric test operator ACAATAAAACTTAA containing the selection target TTAA (underlined). A number of mutants were expressed, purified and characterized by *in vitro* DNA binding techniques. Several mutants recognize the TTAA and/or TTAC sequence(s) with high affinities (in the range of 10^{-11} M) and also show strong preference for these subsites over the wild-type TTGT and other related sequences. These results show that new specificity mutants of the 434 repressor can be discovered by using the sc framework, and suggest that this approach could be applicable to other DNA-binding proteins.

MATERIALS AND METHODS

Construction of the expression libraries of sc repressor mutants and *in vivo* selection by phenotypic screening

The pRIZ'O_{R*}1RR(KOX)69 and pRIZ'O_{R*}2RR(KOX)69 vectors were used for library construction. These vectors were obtained from the corresponding RR*69 vectors (28) by cloning a 1.1 kb stuffer fragment (KOX) between the *Kpn*I and *Xho*I sites. This cloning step replaced the α3 helix of the second domain and thereby eliminated the potentially positive background in the libraries. The partially randomized oligonucleotide TATTCTCTGGTACCWCTNNSNNSAGTATCNNS**CAGCTCGAGCTG** (AT443, W = A or T; N = A, C, G or T; S = G or C) containing cleavage sites for *Kpn*I and *Xho*I (underlined) and a 12 nucleotide long self-complementary 3'-terminal region (bold) was made double stranded by self-annealing followed by a Klenow polymerase fill-in reaction in the presence of dNTP. The duplex was cleaved with *Kpn*I and *Xho*I and the product containing the randomized region was purified by electrophoresis on non-denaturing 16% polyacrylamide gel. This randomized *Kpn*I–*Xho*I linker

```
      C WCT NNS NNS AGT ATC NNS CAG C
CA TGG WGA NNS NNS TCA TAG NNS GTC GAG CT
```

was then used to replace the KOX stuffer fragment of the pRIZ'O_{R*}1RR(KOX)69 vector. The ligation mixture was electroporated into XL1-Blue cells (Stratagene) and a library of 5.8×10^4 independent transformants was obtained. A library of ~4.4×10^4 members was also constructed by using the corresponding O_{R*}2 vector.

The *in vivo* screening for clones containing sc repressor mutants that reduce the expression of the α-fragment of the β-galactosidase reporter gene was performed briefly as follows. The libraries were plated in aliquots onto LB agar plates containing 75 mg/l ampicillin and 10 mg/l tetracycline to obtain ~1000–2000 colonies per plate. Approximately 20 000–25 000 colonies (15 000–20 000 of the O_{R*}1 library and a few thousand of the O_{R*}2 library) were screened in this work. The agar was covered with a nitrocellulose filter (BA 85 type, Schleicher & Schuell) and the plates were incubated at 37°C for 10–12 h. The nitrocellulose filter was then placed (colonies facing up) onto LB agar plates containing antibiotics as above, 1 mM IPTG and 25 mg/l X-gal. The plates were further incubated for 8–12 h at 37°C. Most of the colonies turned slightly blue, and the differences in color intensity were already visible. For better color discrimination, the plates were usually kept for 2–4 days at 4°C. Colonies which were paler blue than the average were taken for further analysis. Nucleotide sequencing was performed by using a T7 sequencing kit (Pharmacia Biotech) and a 'forward' sequencing primer AGCATGGTTAGAGCTGGATC (AT446) which is located in the linker coding region, and/or by the backward, vector-specific primer GGCAGTTTCCCAGACATTACTC (AT419).

Expression and purification of the mutant sc repressors

The DNA fragments coding for the mutant domains R_{XXXX} were isolated from the pRIZ' selection vectors as *Bam*HI–*Hin*dIII fragments and were cloned into the pSETRR90 vector (28) to obtain pSETRR_{XXXX} vectors. Expression was performed by using the BL21(DE3)pLysS strain (Novagen) as described (28), with slight modifications. Sonication was performed in TE (10 mM Tris–HCl, 2 mM EDTA, pH 8.0) buffer containing 100 mM KCl. After batch absorption onto SP-Sepharose, the enriched fractions were purified by HPLC on a Shodex SP 825 column with a linear gradient of KCl (100–400 mM) in TE. Protein concentrations were determined by using the molar extinction coefficients 12 400 or 18 000 $M^{-1}cm^{-1}$ (mutants with one extra Trp residue) at 280 nm, calculated as described (33).

DNA probes

The DNA probes were generated by PCR assisted labeling of operators cloned in the pRIZ' selection vectors (28). The O_R1-2'A1'A operator was originally isolated by binding site selection for RR*69 as b*10 sequence by Chen *et al.* (29). The probes used in EMSA were 124 (O_R1), 125 (O_{R*}1) and 127 bp (O_R1-2'A1'A) long and were generated by using 5'-^{32}P-labeled AGGCTTTACACTTTATGCTTCCG (AT477) and unlabeled GTTTTCCCAGTCACGACGTT (AT474) primers. DNA probes for DNase I protection assay were 157 (O_R1) and 158 bp (O_{R*}1) long as described (28). The designed O_R1-2'N1'C operators were obtained by cloning the double-stranded linker composed of TACAAGAAAGTTTNCAA and TATTGNAAACTTTCTTG random oligonucleotides into the *Nde*I site of pCP8. This is a derivative of the pCP1(-) vector (34) and contains 1 bp deletion in the Shine–Dalgarno region. These cloned operators were used to generate 95 bp long probes by using the 5'-^{32}P-labeled AT477 and unlabeled CGTTGCTAAAGTATCGAGATG (AT483) PCR primers.

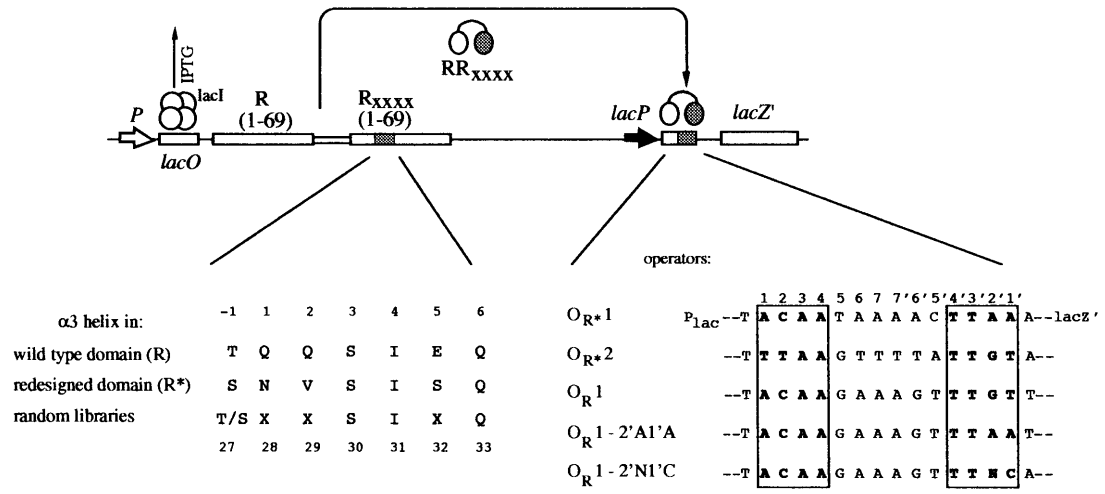

Figure 1. Scheme of the phenotypic screening to detect interaction between mutant DBDs and target DNA. The mutant DBD of the sc repressor and its potential target site are shaded. The amino acid numbering above the residues refers to the positions in the α3 helix and the numbers under the sequences correspond to the positions in the 434 repressor. The operators used in this study are also listed and the boxes indicate the regions that are shown to be in direct contact in the wild-type complex. Only $O_{R*}1$ and $O_{R*}2$ were used as selection targets.

Electrophoretic mobility shift assay (EMSA) and DNase I footprinting

EMSA was performed as previously described (28), but the binding buffer was slightly different and contained 50 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, 2.5 μg/ml sonicated salmon sperm DNA, 0.1 mg/ml bovine serum albumin and 0.02% (w/v) Triton X-100. The concentration of the ^{32}P-labeled DNA probes was generally between 1 and 10 pM and always significantly lower that the equilibrium dissociation constant (K_d) of the analyzed interaction. The protein titration experiments were performed by using 2-fold serial protein dilutions and a minimum of eight concentration points. The binding reactions were performed for 2–3 h at room temperature and analyzed as described (28). Fixed and dried gels were evaluated by using an Instant Imager or a Cyclone Storage Phosphor System (Packard). The K_d was obtained by determining the protein concentration present at half-maximal binding. The fraction of bound DNA (Θ) was plotted against the total protein concentration (P_t) and the data were fit to the equation $\Theta = 1/(1 + K_d/P_t)$ by non-linear least-squares analysis using the KaleidaGraph software (v.3.08, Synergy Software).

DNase I protection assay was performed as described (28) but using the modified binding buffer as above. The binding reactions shown in this work contained 25 nM mutant proteins and 10 times less DNase I, ~1 ng (bovine pancrease, grade II from Boehringer Mannheim) was used in 100 μl reaction volume.

RESULTS AND DISCUSSION

Construction of sc repressor libraries and phenotypic library screening

The previously described one-plasmid system, which showed functional interactions between the intact 434 repressor or its sc derivatives and the corresponding operators, was used for library construction and *in vivo* screening (28). The libraries were constructed in two operator vectors, pRIZ'O_{R*}1 and pRIZ'O_{R*}2 by using XL1-Blue *E.coli* host. We randomized those amino acid residues of the second DBD of the sc repressor which were used in the α-helix redesign experiment to change the DNA binding specificity of the intact 434 repressor to that of the P22 repressor (31), and which were also shown to result in similar specificity change in the sc framework (28). Positions 28, 29 and 32, corresponding to residues 1, 2 and 5 of the α3 helix of the 434 repressor, were fully randomized while position 27 (or –1 residue of α3) was chosen to be either Thr or Ser, the naturally occurring residues in the 434 and P22 repressors respectively. Residue 33 is Gln in both repressors and was kept unchanged in the libraries. Figure 1 shows schematically the principle of the screening as well as the relevant amino acid and nucleotide sequences. It is expected that functional interaction between the mutants and the test operators would lower the β-galactosidase activity causing less intense blue color development in colonies grown on X-Gal indicator plates.

A relatively high proportion of the colonies, at least 1–2 %, appeared paler blue than the average in the library screenings. A number of these clones were sequenced and the amino acid sequences are shown in Table 1. The one-letter amino acid codes are used and the mutants are listed as XXX..X where Xs stand for the selected –1, 1, 2 and 5 residues and the dots for the unchanged 3 and 4 residues of the α3 helix. Certain mutants were isolated more than once and these were often encoded by different codons. Several mutants were isolated for both O_{R*}1 and O_{R*}2; these operators contain identical sequence and differ only in the relative orientation of the two half-sites with respect to the lac promoter. Residue 5 of the α3 helix was most often Ser, Trp or Gly, and the sequences are listed as homology groups mainly according to the identity of this residue. Arg was often found in position 1 when residue 5 was Ser or Trp. Generally, small and/or hydrophobic residues were found in positions 1 and 2 when residue 5 was Gly. No preference for

Table 1. Selection of the sc repressors by *in vivo* library screening and DNA-binding properties of some selected mutants

Sequence in α3 helix	No. of isolates[a]	Affinity (nM) for			Target preference[b]
		O_R1	$O_{R*}1$	O_R1-2'A1'A	
-112 5					
T_{XXSIX}					
S					
TQQ..E R	–	0.017	58.0	14.0	
SNV..S R*	–	4.80	0.0384	0.0124	387
SPS..S	1	1.67	0.420	0.0360	46
TRS..S	2 (1)	1.85	1.90	0.103	18
SRS..S	– (1)				
TRP..S	2	4.14	0.780	0.0525	79
SRP..S	– (1)				
TRE..S	1	1.20	2.12	0.110	11
TRV..S	1	3.20	2.35	0.140	23
TRA..W	2	0.950	6.60	0.970	1
SRA..W	– (1)				
TRE..W	2	2.40	2.74	0.164	15
SRN..W	2	0.260	7.60	1.62	0.16
SRV..W	2 (1)	1.80	1.14	0.0880	20
TRM..W	1	4.10	4.60	0.250	16
SRM..W	– (1)				
TRI..W	1	5.65	5.02	0.220	27
TRT..W	1	0.378	1.43	0.202	1.9
SRQ..W	1	1.57	6.20	1.32	1.2
SRV..A	1 (1)	2.40	10.6	2.00	1.2
TRV..G	1	1.50	18.0	0.970	1.5
SGV..G	1				
SLG..G	1				
TLA..G	1	2.40	32.0	2.00	1.2
TVA..G	1	0.770	10.6	0.830	0.9
TAT..G	1	0.140	6.00	0.400	0.35
SAA..G	1				
SNS..G	1	0.430	2.30	0.250	1.7
TVN..S	1	1.50	7.50	0.170	8.8
SRW..G	5	6.70	28.0	3.40	2.0
SRW..M	2				
SRW..V	1	2.40	34.0	>10.2	<0.23
STW..V	1	5.00	26.0	5.10	1.0
TGE..G	1				
SYG..G	1				
TQA..W	1				
SLL..S	1				
TWA..A	1				
TIR..L	1				
SEH..A	1				
TGR..R	1				
SWR..R	1				
TRR..V	1				

[a]Figures in parentheses are from the $O_{R*}2$ selection experiment.
[b]Affinity ratios calculated as K_d for O_R1/K_d for O_R1-2'A1'A.

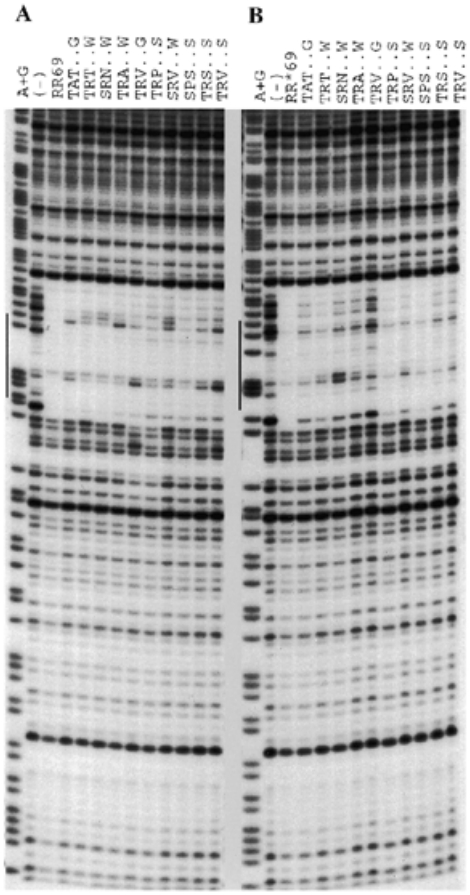

Figure 2. Protection of the operator sites against DNase I cleavage by several sc repressor mutants. (**A**) Protection of O_R1; (**B**) protection of $O_{R*}1$. The operator sites, aligned to the A+G lanes, are marked by vertical lines. RR69 and RR*69 were used as positive controls for O_R1 and $O_{R*}1$, respectively. All proteins were used at a concentration of 25 nM. The mutants are listed on the top of the lanes according to the abbreviations used in Table 1.

either Ser or Thr could be observed in the last position of the turn (or –1 position of the α3 helix) and certain combinations of 1, 2 and 5 residues were obtained with both Ser and Thr in this position. A number of phenotypically selected sc mutant proteins (abbreviated as RR_{XXXX}) were expressed for characterization by *in vitro* binding techniques.

DNA binding properties of the isolated mutant sc repressors: affinity, specificity and context effects

EMSA and purified proteins were used in protein titration experiments to determine the binding affinities for DNA probes (~125 bp long) containing the wild-type O_R1 or the target $O_{R*}1$ operator. The K_d are given as the protein concentrations required for half-maximal binding and are summarized in Table 1. DNase I footprinting experiments showed protection of the operator site of long (~158 bp) DNA probes (Fig. 2), indicating that the binding observed in the EMSA takes place at these sites. Both the EMSA and the footprinting assays were

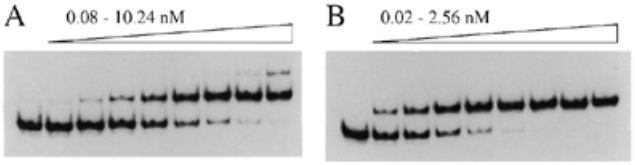

Figure 3. Protein titrations showing that the target site recognition is strongly context dependent. The RR_{TRPS} mutant binds O_R1-2'A1'A operator (**B**) ~15-fold better than $O_{R*}1$ (**A**).

performed in the presence of a large molar excess of non-specific DNA over the probes used (Materials and Methods).

These experiments showed that the *in vivo* selected mutants bound the $O_{R*}1$ site *in vitro* but none of them was as strong at binding as the previously designed RR*69 (RR_{SNVS} using the present abbreviation). At the same time, many of the mutants bound O_R1 better than $O_{R*}1$ and only a few mutants exhibited preference for $O_{R*}1$ over O_R1. Even in these cases, this preference was not more than a modest 5–6-fold (RR_{TRPS}). One could reason, however, that these relative affinities do not necessarily reflect the real, intrinsic preference of the mutant domain for TTAA over TTGT sequence. The generally higher affinities obtained for O_R1 can be explained partly by the observed preference of the wild-type R domain for G (in O_R1) over T (in $O_{R*}1$) at position 5 of the operator (29,35). The two test operators may also differ substantially in terms of DNA flexibility, an important determinant for high affinity binding by the 434 repressor (36,37) and also by the sc repressors RR69 and RR*69 (29). To test this assertion, we used a DNA probe which contains the TTAA narrow selection target in place of the TTGT sequence of the O_R1 operator (O_R1-2'A1'A, see Fig. 1), i.e. we compared the binding affinities for TTAA and TTGT in identical sequence context. It was observed that the mutants generally bound O_R1-2'A1'A significantly better (10–25-fold in most cases) than $O_{R*}1$. The mutant RR_{TRPS} bound O_R1-2'A1'A ~15-fold better than $O_{R*}1$ (Fig. 3); thus, the assumed 5–6-fold target/wild-type preference is in fact ~80-fold. The combinatorial amino acid changes in RR_{TRPS} resulted in an affinity decrease, compared to the wild-type RR69, of 240-fold for O_R1 and an affinity increase of 266-fold for O_R1-2'A1'A. Table 1 shows the affinity data and the target preferences of all expressed mutants for O_R1-2'A1'A. According to these data, >40% of the expressed and tested mutants prefer 10–80-fold the selection target TTAA over TTGT. About half of the mutants showed no preference for either sequence and several mutants preferred slightly (3–6-fold) the wild-type recognition sequence over the selection target.

The *in vitro* binding data thus show that our screening resulted in both new, altered specificity and broadened specificity mutants. The screening reveals mutants that interact with the operator, but it is possible that certain mutants can also bind, even with higher affinities, to other sequences. This is in fact observed with other DNA-binding proteins, for example in zinc finger phage display selections (4,13,14,38) even when stringent selection conditions were used (8,10). Testing the DNA-binding specificities of the selected mutants is therefore important and can also lead to the identification of novel specificities.

Table 2. DNA binding affinities of several mutants for designed operator derivatives containing TTNC subsites

Protein	K_d (pM) for operator with subsite			
	TTAC	TTCC	TTGC	TTTC
RR_{TRES}	9.0	52	1390	2600
RR_{TREW}	25.0	192	2470	2470
RR_{TRPS}	12.6	1180	1380	410
RR_{TRSS}	102.0	1790	740	416
RR_{TRVS}	35.8	1120	1060	171
RR_{TRVG}	83.3	784	420	164
RR_{TRVA}	75.1	744	1320	378
RR_{TRVW}	67.6	1810	535	257

Mutants that recognize the TTAC sequence with high affinities and specificities

Many of the selected mutant DBDs contained Arg in the position 1 of the α3 helix (Table 1). The chemical recognition principles and statistical analysis of protein–DNA complexes suggest that this Arg could favorably interact with a G residue (23,39–41). The corresponding residue in the wild-type protein is Gln which is shown to interact with the A residue of the 1 (or the 1') bp of operator sites in different complexes (35,42,43). Assuming a topologically equivalent interaction between mutant sc repressors and operators, Arg1 of the α3 helix (Arg28) can contact the G residue of the 1' C-G base pair of a general operator O_R1-2'N1'C, where the putative target for the mutant DBD is 4'-TTNC-1' (see Fig. 1 for sequences and numbering schemes). We prepared these O_R1 derivatives with all four possible bases and screened them with the Arg28 containing mutants in EMSA. Figure 4 shows that most of these mutants interact strongly with one or more of these operators at 0.2 nM protein concentration and the highest affinities are generally observed with the TTAC containing probe O_R1-2'A1'C, which is the closest homolog to the original selection target.

Mutants that showed strong binding and preference for the TTAC probe were characterized in more detail by determining their binding affinities for all four TTNC variants (Table 2). Some of these interactions (RR_{TRPS} and RR_{TRES} with the TTAC probe) are at least as strong as those observed between the wild-type RR69 and O_R1 or between the designed RR*69 and its cognate operators: O_{R*}1 (designed) or O_R1-2'A1'A (selected in 29). The specificity of the interactions is also higher in several cases than those obtained with the designed RR*69 mutant, which showed only an ~1.5-fold preference for its optimal TTAA subsite over TTTA (29 and data obtained under the present assay conditions). For example, the high affinity mutant RR_{TRES} bound nearly 6-fold better the TTAC derivative than the next strongest binding site TTCC. The RR_{TRPS} bound TTAC significantly (33–110-fold) better than any other TTNC and 4-fold better than the nearest high affinity target, the TTAA (in the O_R1-2'A1'A context). The slightly lower affinity mutant RR_{TREW} showed at least as high sequence specificity as its close homolog RR_{TRES}. The RR_{TRSS} shows 4–18-fold preference

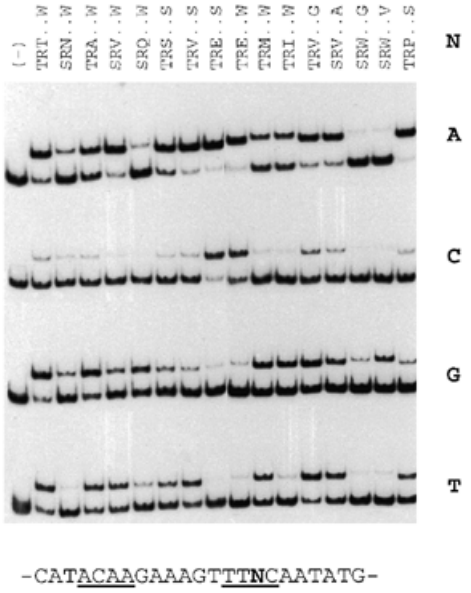

Figure 4. Testing the binding specificities of the Arg28 containing mutants in EMSA by using designed operator derivatives. The operators containing the general TTNC target site (where N = A or C or G or T) were tested individually with 16 mutant proteins at 0.2 nM concentration. Mutant sc repressors are abbreviated as in Table 1.

for TTAC over other TTNC derivatives but this interaction is not particularly strong and is about the same as observed for the TTAA sequence. Mutants with the Arg1Val2 sequence motif (RV motif in α3 helix positions 1 and 2) show different degrees of preferences for TTAC in the TTNC group and also discriminate differentially TTAC and TTAA, depending on the identity of the residue 5. For a better comparison, some of the affinity and specificity data of Tables 1 and 2 are compared as relative affinities and shown in Figure 5.

The examples shown in Figure 5 summarize the main results of the present work. Starting from combinatorial mutant libraries of the sc 434 repressor, it is possible to isolate mutant DBDs of altered binding specificities by using an *in vivo* screening

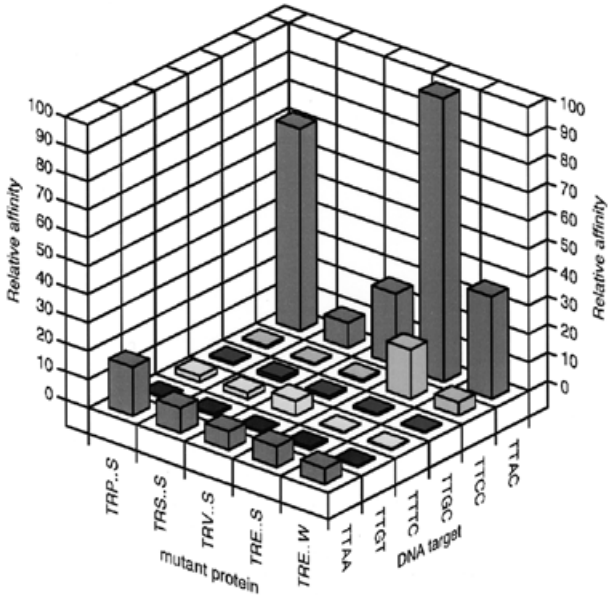

Figure 5. Comparison of the binding affinities and specificities of several mutants. The value 100 was arbitrarily assigned to the highest affinity interaction (with a K_d of 9×10^{-12} M) observed between RR_{TRES} and O_R1-2'A1'C.

method. Mutants showing strong preference for the selection target (TTAA in this work) over the wild-type sequence (TTGT) were isolated. Specificity studies revealed, however, that certain mutants could bind with even higher affinities to a sequence (TTAC) which is a close homolog of the target. Similar situations were also observed in zinc finger selections (6,8,10,38). The binding affinities of several new interactions are in the range of 10^{-11} M and this high affinity is often accompanied by high degree of sequence specificity. We have been using binding site selection techniques to complement the specificity studies of this work and the most specific mutant DBDs are being used as building blocks in the sc arrangement to construct DNA-binding proteins of entirely new specificities.

ACKNOWLEDGEMENTS

Thanks are due to K. Vlahovicek for help with the manuscript preparation and to Dr C. Guarnaccia for protein purifications. S.W. was supported by an ICGEB fellowship. A.S. and M.-L.T. are on leave from the Stockholm University, The Arrhenius Laboratories for Natural Sciences, Stockholm, Sweden.

REFERENCES

1. Greisman,H.A. and Pabo,C.O. (1997) *Science*, **275**, 657–661.
2. Kim,J.-S. and Pabo,C.O. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 2812–2817.
3. Liu,Q., Segal,D.J., Ghiara,J.B. and Barbas,C.F.,III (1997) *Proc. Natl Acad. Sci. USA*, **94**, 5525–5530.
4. Choo,Y., Castellanos,A., Garcia-Hernandez,B., Sanchez-Garcia,I. and Klug,A. (1997) *J. Mol. Biol.*, **273**, 525–532.
5. Kim,J.-S. and Pabo,C.O. (1997) *J. Biol. Chem.*, **272**, 29795–29800.
6. Isalan,M., Klug,A. and Choo,Y. (1998) *Biochemistry*, **37**, 12026–12033.
7. Isalan,M., Choo,Y. and Klug,A. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 5617–5621.
8. Wolfe,S.A., Greisman,H.A., Ramm,E.I. and Pabo,C.O. (1999) *J. Mol. Biol.*, **285**, 1917–1934.
9. Beerli,R.R., Segal,D.J., Dreiler,B. and Barbas,C.F.,III (1998) *Proc. Natl Acad. Sci. USA*, **95**, 14628–14633.
10. Segal,D.J., Dreier,B., Beerli,R.R. and Barbas,C.F.,III (1999) *Proc. Natl Acad. Sci. USA*, **96**, 2758–2763.
11. Berg,J.M. (1997) *Nature Biotechnol.*, **15**, 323.
12. Choo,Y. and Klug,A. (1995) *Curr. Opin. Biotechnol.*, **6**, 431–436.
13. Rebar,E.J., Greisman,H.A. and Pabo,C.O. (1996) *Methods Enzymol.*, **267**, 129–149.
14. Berg,J.M. and Shi,Y. (1996) *Science*, **271**, 1081–1085.
15. Rhodes,D., Schwabe,J.W.R., Chapman,L. and Fairall,L. (1996) *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, **351**, 501–509.
16. Ebright,R.H., Cossart,P., Gicquel-Sanzey,B. and Beckwith,J. (1984) *Nature*, **311**, 232–235.
17. Wharton,R.P., Brown,E.L. and Ptashne,M. (1984) *Cell*, **38**, 361–369.
18. Wharton,R.P. and Ptashne,M. (1987) *Nature*, **326**, 888–891.
19. Lehming,N., Sartorius,J., Niemöller,M., Genenger,G., von Wilcken-Bergmann,B. and Müller-Hill,B. (1987) *EMBO J.*, **6**, 3145–3153.
20. Helbl,V., Tiebel,B. and Hillen,W. (1998) *J. Mol. Biol.*, **276**, 319–324.
21. Helbl,V. and Hillen,W. (1998) *J. Mol. Biol.*, **276**, 313–318.
22. Harrison,S.C. and Aggarwal,A.K. (1990) *Annu. Rev. Biochem.*, **59**, 933–969.
23. Pabo,C.O. and Sauer,R.T. (1992) *Annu. Rev. Biochem.*, **61**, 1053–1095.
24. Hollis,M., Valenzula,D., Pioli,D., Wharton,R. and Ptashne,M. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 5834–5838.
25. Webster,C., Merryweather,A. and Brammar,W. (1992) *Mol. Microbiol.*, **6**, 371–377.
26. Zhou,Y., Busby,S. and Ebright,R.H. (1993) *Cell*, **73**, 375–379.
27. Percipalle,P., Simoncsits,A., Zakhariev,S., Guarnaccia,C., Sanchez,R. and Pongor,S. (1995) *EMBO J.*, **14**, 3200–3205.
28. Simoncsits,A., Chen,J., Percipalle,P., Wang,S., Törö,I. and Pongor,S. (1997) *J. Mol. Biol.*, **267**, 118–131.
29. Chen,J., Pongor,S. and Simoncsits,A. (1997) *Nucleic Acids Res.*, **25**, 2047–2054.
30. Koudelka,G.B., Harrison,S.C. and Ptashne,M. (1987) *Nature*, **326**, 886–888.
31. Wharton,R.P. and Ptashne,M. (1985) *Nature*, **316**, 601–605.
32. Huang,L., Sera,T. and Schultz,P.G. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 3969–3973.
33. Gill,S.C. and von Hippel,P.H. (1989) *Anal. Biochem.*, **182**, 319–326.
34. Gal,J., Szekeres,S., Schnell,R., Pongor,S., Simoncsits,A. and Kalman,M. (1999) *Anal. Biochem.*, **266**, 235–238.
35. Aggarwal,A.K., Rodger,D.W., Drottar,M., Ptashne,M. and Harrison,S.C. (1988) *Science*, **242**, 899–907.
36. Koudelka,G.B., Harbury,P., Harrison,S.C. and Ptashne,M. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 4633–4637.
37. Koudelka,G.B. and Carlson,P. (1992) *Nature*, **355**, 89–91.
38. Choo,Y. and Klug,A. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 11168–11172.
39. Seeman,N.C., Rosenberg,M.J. and Rich,A. (1976) *Proc. Natl Acad. Sci. USA*, **73**, 804–808.
40. Suzuki,M. (1994) *Structure*, **2**, 317–326.
41. Mandel-Gutfreund,Y., Schueler,O. and Margalit,H. (1995) *J. Mol. Biol.*, **253**, 370–382.
42. Shimon,L.J. and Harrison,S.C. (1993) *J. Mol. Biol.*, **232**, 826–838.
43. Rodgers,D.W. and Harrison,S.C. (1993) *Structure*, **1**, 227–240.