

# The DNA-recognition mode shared by archaeal feast/famine-regulatory proteins revealed by the DNA-binding specificities of TvFL3, FL10, FL11 and Ss-LrpB

Katsushi Yokoyama<sup>1,2</sup>, Hideki Nogami<sup>1,2</sup>, Mamiko Kabasawa<sup>1,2</sup>, Sonomi Ebihara<sup>1,2</sup>, Ai Shimowasa<sup>1,2</sup>, Keiko Hashimoto<sup>1,2</sup>, Tsuyoshi Kawashima<sup>1,2</sup>, Sanae A. Ishijima<sup>1,2</sup> and Masashi Suzuki<sup>1,\*</sup>

<sup>1</sup>National Institute of Advanced Industrial Science and Technology, Tsukuba Center 6-10, Higashi 1-1-1, Tsukuba 305-8566 and <sup>2</sup>Japan Science and Technology Agency, Core Research for Evolutional Science and Technology, Sanbancho 5, Chiyoda-ku, Tokyo 102-0075, Japan

Received February 2, 2009; Revised April 26, 2009; Accepted April 27, 2009

## ABSTRACT

The DNA-binding mode of archaeal feast/famine-regulatory proteins (FFRPs), i.e. paralogs of the *Escherichia coli* leucine-responsive regulatory protein (Lrp), was studied. Using the method of systematic evolution of ligands by exponential enrichment (SELEX), optimal DNA duplexes for interacting with TvFL3, FL10, FL11 and Ss-LrpB were identified as **TA CGA[AAT/ATT]TCGTA**, **GTTCGA[AAT/ATT]TCGAAC**, **CCGAAA[AAT/ATT]TTTCGG** and **TTGCAA[AAT/ATT]TTGCAA**, respectively, all fitting into the form abcdeWWWedcba. Here W is A or T, and e.g. a and a are bases complementary to each other. Apparent equilibrium binding constants of the FFRPs and various DNA duplexes were determined, thereby confirming the DNA-binding specificities of the FFRPs. It is likely that these FFRPs recognize DNA in essentially the same way, since their DNA-binding specificities were all explained by the same pattern of relationship between amino-acid positions and base positions to form chemical interactions. As predicted from this relationship, when Gly36 of TvFL3 was replaced by Thr, the b base in the optimal DNA duplex changed from A to T, and, when Thr36 of FL10 was replaced by Ser, the b base changed from T to G/A. DNA-binding characteristics of other archaeal FFRPs, Ptr1, Ptr2, Ss-Lrp and LysM, are also consistent with the relationship.

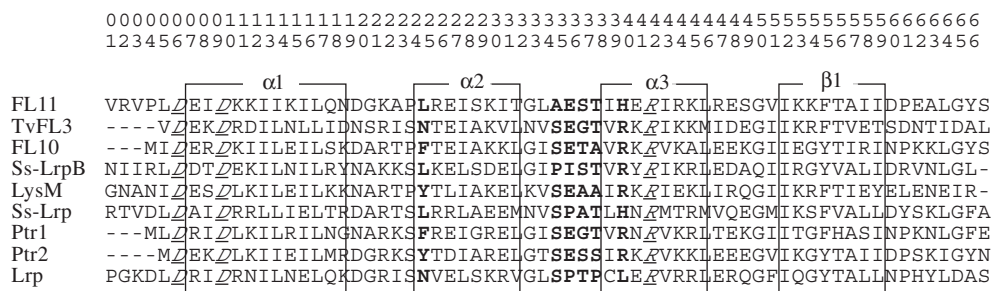
## INTRODUCTION

Homologs of the *Escherichia coli* leucine-responsive regulatory protein (Lrp) are distributed throughout archaea and eubacteria, composing a large family of transcription factors (1–4). Sensing the presence of rich nutrition in many cases by a high concentration of leucine, *E. coli* regulates ~100 transcription units by Lrp, thereby changing its overall metabolism (5–7). In order to summarize this global regulation, Calvo and Matthews (5) used the term feast/famine regulation. For this reason, Lrp homologs can be referred to as feast/famine regulatory proteins (FFRPs) (1,3,4). In fact, sensing the lysine concentration, the archaeon *Pyrococcus* sp. OT3 regulates ~200 transcription units using the FFRP FL11, thereby shifting its metabolism between the feast and famine modes (8). Lrp interacts not only with leucine, but also with alanine, isoleucine and valine (9). Similarly various archaeal FFRPs interact with amino acids, some with multiple types (10,11). It is believed that the last common ancestor of extant organisms first differentiated to archaea and eubacteria. It is likely that this ancestor regulated its overall metabolism using an FFRP, sensing the nutritional condition by the concentration of an amino acid.

We determined the crystal structure of the FL11 dimer in complex with the DNA duplex, **TGAAA[AAT/ATT]TTCA** (8). To our knowledge this is the single FFRP–DNA complex so far determined. Although DNA had been added while crystallizing Lrp, the structure of the DNA was not determined (12). In the FL11–DNA crystal complex, six amino-acid residues of each FL11 monomer bound to 5 bp, TGAAA/TTTCA, at each end of the DNA [summarized in Figure 2 of (8)]. This DNA recognition

\*To whom correspondence should be addressed. Tel: +81 29 861 6583; Fax: +81 29 861 6041; Email: masashi.suzuki@aist.go.jp  
Present address:

Tsuyoshi Kawashima, Yokohama College of Pharmacy, Laboratory of Molecular Biology, Matano 601, Totsuka-ku, Yokohama 245-0066, Japan.



**Figure 1.** Amino-acid sequences of the DNA-binding domains of FFRPs. Secondary structural elements, i.e.  $\alpha$ -helices 1–3 ( $\alpha$ 1–3) and  $\beta$ -strand 1 ( $\beta$ 1), were identified using the crystal structures of FL11 (8), LrpA (24) and Lrp (12). Six residues of FL11 bound to bases in the crystal complex with DNA (8) and residues at the identical positions in the other FFRPs are indicated by bold characters. In the crystal complex Asp6 and Asp9 interacted with Arg41 thereby directing Arg41 to a DNA phosphate (8). The three residues conserved among the FFRPs are italicized and underlined. The sequence identities of the DNA-binding domains of the archaeal FFRPs, FL11-Ptr2, are 31.3–59.6%, and those between the DNA-binding domain of Lrp and the archaeal domains are 30.8–41.9%. The numbering scheme shown here is used to describe amino-acid positions of FFRPs in text.

mode is unusual, since the six residues are not positioned on one side of an  $\alpha$ -helix or a  $\beta$ -sheet, but four of them cluster in a loop between  $\alpha$ -helices (Figure 1). DNA-binding of other archaeal FFRPs was characterized, i.e. LrpA from *Pyrococcus furiosus* (13,14), Ptr1 and Ptr2 from *Methanocaldococcus jannaschii* (15), and LysM (16), Ss-Lrp (17) and Ss-LrpB (18,19) from *Sulfolobus solfataricus*. In these characterizations small numbers of binding sites were analyzed. Many of them were based on the assumption that FFRPs would regulate genes coding themselves, but this can be misleading. Consequently it was not clear whether or not the crystal complex represented a common DNA recognition mode shared by archaeal FFRPs (4,8).

In order to obtain a global insight into DNA recognition by archaeal FFRPs, in this study DNA duplexes optimal for interacting with four archaeal FFRPs and their three variants, where amino-acid residues are replaced, have been identified by selecting ~100 sites tightly interacting with each protein, using the method of systematic evolution of ligands by exponential enrichment (SELEX) (20). By analyzing the nucleotide sequences of these DNA duplexes in the light of the amino-acid sequences of the FFRPs, a common contact pattern relating amino-acid positions with base positions has been deduced.

Archaeal FFRPs used in this study are TvFL3 (genbank ID 13541958) from *Thermoplasma volcanium*, FL10 (14591370) and FL11 (14591302) from *P. OT3*, and Ss-LrpB (15898915) from *S. solfataricus* (Figure 1). These archaea are thermophiles, having optimal growth temperatures of 60°C (*T. volcanium*), 98°C (*P. OT3*) and 80°C (*S. solfataricus*) (21–23). Before this study the DNA-binding specificity of TvFL3 was unknown. Ss-LrpB binds to three sites in the *ss-lrpB* promoter, whose consensus is TTGYAWWWWTRCAA, where Y is T or C, R is A or G and W is A or T (18). Interaction of Ss-LrpB with TTGCAA[AAT/ATT]TTGCAA and its variants, where single or 2 bp each were replaced by other types, was studied (19). FL10 is an ortholog of LrpA from *P. furiosus*, whose crystal structure was determined in the absence of DNA (24). LrpA represses transcription from the *lrpA* promoter (13,14), and consensus of two

LrpA-binding sites in the *lrpA* promoter is GTCGA [AGA/TCT]TCGAC (25). Binding of FL11 to various promoters was characterized, and consensus of these binding sites is TTGAAA[AAT/ATT]TTTCAA (8). Nevertheless it has turned out that not all of these sequences are optimal for interacting with the FFRPs.

## MATERIALS AND METHODS

### Protein purification

The *tvfl3* gene was cloned into the pET15b vector and introduced into the *E. coli* strain BL21(DE3) (Novagen) in order to synthesize TvFL3 by adding a His-tag to its N-terminus. To cells growing at 37°C at an OD<sub>600</sub> of 0.5–0.8, isopropyl-1-thio- $\beta$ -D-galactopyranoside (IPTG) was added to 1 mM, and the culture was continued for additional 4 h. Cells were collected by centrifugation at 8000  $\times$  g for 10 min at 4°C, suspended in 200 ml of 50 mM NaH<sub>2</sub>PO<sub>4</sub> containing 300 mM NaCl and 20 mM imidazole, and sonicated. After centrifugation at 48400  $\times$  g for 30 min at 4°C, the supernatant was incubated at 60°C for 30 min. After another centrifugation at 48400  $\times$  g for 20 min at 4°C, the supernatant was filtered through a membrane of pore size 0.45  $\mu$ m, and subjected to affinity chromatography using Ni-NTA agarose (QIAGEN) and a linear, 0.02–1 M, gradient of imidazole in 50 mM NaH<sub>2</sub>PO<sub>4</sub> containing 300 mM NaCl. Proteins eluting at 40–100 mM imidazole were subjected to gel filtration using Sephacryl S-300 (GE Healthcare) equilibrated with 50 mM NaH<sub>2</sub>PO<sub>4</sub> containing 300 mM NaCl.

The *tvfl3* gene was modified using the method of Higuchi *et al.* (26). A pair of DNA duplexes was amplified by PCR (27), which covered the gene from different ends to the position to be modified. Of two primers used for PCR of each duplex, that corresponding to the position to be modified was designed so that GGA coding Gly36 was replaced by GCA coding Ala or ACA coding Thr in the duplex amplified. In this paper amino-acid positions are described using the positions of corresponding residues in FL11 (Figure 1). The two duplexes were fused by PCR using two of the four primers corresponding to the two

ends of the gene. TvFL3[G36A] and TvFL3[G36T] were synthesized and purified in the same way as the original TvFL3 except that gel filtration using Sephacryl S-300 was not performed.

The *fl10* gene was cloned into the pET21a vector and introduced into the *E. coli* strain Rosetta 2(DE3) (Novagen). After induction using IPTG, cells were suspended in 200 ml of 100 mM Na-phosphate buffer (pH 7.0) containing 150 mM NaCl and 5% (v/v) glycerol, and sonicated. After centrifugation at  $48400 \times g$  for 30 min at 4°C, the supernatant was incubated at 75°C for 30 min. After another centrifugation at  $48400 \times g$  for 20 min at 4°C, the supernatant was dialyzed against 50 mM Na-phosphate buffer (pH 7.0), and subjected to anion-exchange chromatography using Q-sepharose (GE Healthcare) and a linear, 0–1 M, gradient of NaCl in the same buffer. Proteins eluting at 400–650 mM NaCl were subjected to gel filtration using Sephacryl S-300 equilibrated with 50 mM Na-phosphate buffer (pH 7.0) containing 150 mM NaCl. The *fl10* gene was modified using the method of Higuchi *et al.* (26) so that ACA coding Thr36 was replaced by TCA coding Ser. FL10[T36S] was synthesized and purified in the same way as the original FL10.

FL11 was synthesized and purified as has been previously described (11).

The *ss-lrpB* gene was cloned into the pET28a vector and introduced into the *E. coli* strain Rosetta 2(DE3) in order to synthesize Ss-LrpB by adding a His-tag to its N-terminus. After induction using IPTG, cells were suspended to 120 ml of 50 mM K-phosphate buffer (pH 7.0) containing 300 mM NaCl and 20 mM imidazole, and sonicated. After centrifugation at  $11900 \times g$  for 20 min at 4°C, the supernatant was subjected to affinity chromatography using Ni-NTA superflow (QIAGEN) and a linear, 0.2–1 M, gradient of imidazole in 50 mM K-phosphate buffer (pH 7.0) containing 300 mM NaCl. Proteins eluting at 250–550 mM imidazole were concentrated, and applied to gel filtration using Sephacryl S-300 equilibrated with 50 mM K-phosphate buffer (pH 7.0) containing 300 mM NaCl.

### Selection of DNA fragments tightly interacting with FFRPs

A mixture of single-stranded DNA fragments of 79 bases in the form 5'-GAAATTAATACGACTCACTATGGG GAGAGAGANNNNNNNWWNNNNNNNAGAG AGTCGCTAGTTATTGCTCAGCGGTGG-3' was synthesized, where NNNNNNNWWNNNNNNN was inserted, with N being A, T, G or C, and W being A or T. Using PCR (27) with primers, CCACCGCTGAGCAATAACTAGCGACTCTCT and GAAATTAATACGACTCACTATGGGAGAGAGA, DNA duplexes were synthesized and amplified.

The DNA duplexes, 10.0 pmol, were mixed with an FFRP at room temperature in 42 mM Na-phosphate buffer (pH 7.0) containing 125 mM NaCl and 6.7% (w/v) sucrose, 12  $\mu$ l. Altogether  $1.07 \times 10^9$  different sequences fit into the form NNNNNNNWWNNNNNNN, and with 10.0 pmol on average 5626 molecules were expected for each sequence. The quantities of the FFRP dimer used were 10.0 pmol in the first to third cycles, 5.0 pmol

in the fourth cycle, 3.0 pmol in the fifth cycle and 2.0 pmol in the sixth cycle, yielding DNA to FFRP dimer ratios of 1:1, 1:0.5, 1:0.3 and 1:0.2, respectively. After being kept for 10 min at room temperature, the DNA-FFRP solution was submitted to electrophoresis using a 12% polyacrylamide gel and 90 mM Tris-borate buffer (pH 8.3) containing 1 mM ethylenediaminetetraacetic acid (EDTA). The gel was stained with ethidium bromide, and the part containing the band of duplexes bound by the FFRP dimer was cut out. When the band was not visually identified, the part corresponding to markers of 600–800 bp was cut out. By overnight incubation of gel pieces at 37°C, DNA fragments were eluted into 800  $\mu$ l of solution A, i.e. 500 mM ammonium acetate containing 10 mM Mg(CH<sub>3</sub>COO)<sub>2</sub>, 1 mM EDTA and 0.1% sodium dodecyl sulfate (SDS). After proteins were removed using phenol-chloroform, DNA fragments were precipitated by ethanol, and dissolved into water, 20  $\mu$ l.

Using ExTaq polymerase (TaKaRa), the DNA fragments were amplified by two steps of PCR. At the first step, 15 cycles of PCR were carried out using the DNA solution, 10  $\mu$ l, mixed into a reaction mixture, 50  $\mu$ l. After electrophoresis using a 12% polyacrylamide gel and 40 mM Tris-acetate buffer (pH 8.1) containing 1 mM EDTA, DNA fragments were eluted into solution A, 400  $\mu$ l, by overnight incubation at 37°C. After proteins were removed using phenol-chloroform, DNA fragments were precipitated by ethanol, and dissolved into water, 20  $\mu$ l. At the second step, 25 cycles of PCR were carried out using the DNA solution, 8  $\mu$ l, mixed into a reaction mixture, 800  $\mu$ l, and divided into 16 tubes of 50  $\mu$ l each. After electrophoresis using a 12% polyacrylamide gel and 40 mM Tris-acetate buffer (pH 8.1) containing 1 mM EDTA, DNA fragments were eluted into solution A, 800  $\mu$ l, by overnight incubation at 37°C. After proteins were removed using phenol-chloroform, DNA fragments were precipitated by ethanol, and dissolved into 50 mM Na-phosphate buffer (pH 7.0) containing 150 mM NaCl, 20  $\mu$ l. The DNA solution was used in the subsequent cycle of selection. After the sixth cycle, DNA fragments identified as interacting with the FFRP were cloned into the HincII site of the pUC118 vector using Mighty Cloning Kit (TaKaRa), and sequenced using PRISM377 (ABI).

In order to confirm identities of the central 3 bp in optimal DNA duplexes for interacting with the FFRPs, mixtures of single-stranded DNA fragments, where GCT ACGANNNTCGTACG, CGTTCGANNNTCGAAC and CTGCAANNNTTGCAAC, respectively, were inserted between GAAATTAATACGACTCACTATGG GAGAGAGA and TGTGTGTCGCTAGTTATTGCT CAGCGGTGG, were synthesized for selection using TvFL3, FL10 and Ss-LrpB, respectively. Duplexes in each type were synthesized and amplified by PCR, and mixed with the FFRP at room temperature. In experiments using TvFL3 and FL10 the DNA to FFRP dimer ratio was kept to 1:0.05 (first cycle) or 1:0.033 (second to fourth cycles). In another experiment using Ss-LrpB the ratio was kept to 1:0.3 (first) or 1:0.2 (second to fourth). DNA duplexes identified as interacting with each FFRP were amplified by 30 cycles of PCR, with each cycle carried out in the same way as that in the second step of PCR



in the other procedure, and used in the subsequent cycle of selection. After the fourth cycle, DNA fragments interacting with each FFRP were cloned into the pUC118 vector and sequenced.

### Sequence analysis of DNA fragments tightly interacting with FFRPs

The nucleotide sequences of DNA fragments, identified as tightly interacting with each FFRP, were analyzed. Using a sliding window, each fragment was scanned from the last GA of GAGAGAGA through the insertion to the first AG of AGAGAGTC against each reference sequence, thereby searching for the highest number of matches. Reference sequences having the highest average number of matches with insertions were identified from all 13-bp sequences, and also from 1024 sequences in the form **abcdeWWWedcba**. Here, e.g. **a** and **a** are bases complementary to each other, and differences in combining 3 W bases were ignored.

For each insertion, the 13-bp site best matching the consensus **abcdeWWWedcba** sequence was identified as a binding site of the FFRP. Bases corresponding to **edcba** at binding sites were converted to the complementary bases, and jointly analyzed with bases originally corresponding to **abcde**, for calculating the frequencies of A, T, G and C at these positions. Base frequencies at -1 and -2 upstream of **abcde** as well as at +1 downstream of **abcde** were also calculated. The bases whose frequencies were 45% or higher at these positions were identified as optimal for interacting with the FFRP. In combination with the most frequent base combination between **abcde** and **edcba**, an optimal DNA duplex for interacting with the FFRP was identified.

### Determination of apparent equilibrium FFRP-DNA binding constants

Seven DNA duplexes of 79 bp each were synthesized. In each duplex along one of the two strands between GAAA TTAATACGACTCACTATGGGGAGAGAGA at the 5'-end and TCGTAGTTATTGCTCAGCGGTGG at the 3'-end inserted was CGTTCGA[AAT]TCGAACCTGTGTG (referred to as the TTCGA duplex), CGTACG A[AAT]TCGTACCTGTGTG (the TACGA duplex), CGGTTCGA[AAT]TCGACCCTGTGTG (the GTCGA duplex), CGGACGA[AAT]TCGTCCCTGTGTG (the GACGA duplex), CGCCGAAA[ATT]TTTCGGAGAGA G (the CGAAA duplex), CGCTGAAA[ATT]TTTCAG AGAGAG (the TGAAA duplex) or CTTGCAA[AAT]T TGCAACTGTGTG (the TGCAA duplex). One of the duplexes and an FFRP were mixed into 42 mM Na-phosphate buffer (pH 7.0) containing 125 mM NaCl and 6.7% (v/w) sucrose, 12  $\mu$ l. While the concentration of the duplex was fixed at  $2.25 \times 10^{-8}$  M, that of the FFRP dimer was changed in six to seven steps from 0M to a value between  $1.8 \times 10^{-7}$  M and  $14.4 \times 10^{-7}$  M, thereby creating a series. After being kept for 10 min at 22°C, the series was submitted to electrophoresis, using a 12% polyacrylamide gel and 90 mM Tris-borate buffer (pH 8.3) containing 1 mM EDTA. To make a calibration curve, standards, 12  $\mu$ l each, containing the same DNA duplex with

concentrations of  $0.25 \times 10^{-8}$  M,  $0.50 \times 10^{-8}$  M,  $0.75 \times 10^{-8}$  M,  $1.00 \times 10^{-8}$  M,  $1.25 \times 10^{-8}$  M,  $1.50 \times 10^{-8}$  M,  $1.75 \times 10^{-8}$  M,  $2.00 \times 10^{-8}$  M and  $2.25 \times 10^{-8}$  M, were applied to each gel.

After electrophoresis each gel was stained with a DNA-binding fluorescent dye, SYBR<sup>®</sup> Gold (Invitrogen), following a procedure recommended by the manufacturer. Using an image analyzer, PharosFX<sup>™</sup> (BIO-RAD), exciting at 488 nm and filtering at 530 nm, the integrated intensity of the DNA band unbound by the FFRP in each lane was quantified according to the calibration curve. The intensity of the bound band in each lane was not directly measured, but identified as the difference between the unbound bands in the presence and absence of the FFRP due to smearing of the bound band [see (19) for similar identification]. Taking Y as the intensity of the bound band, and X as the total concentration of the FFRP, the curve  $Y = aX^b/(c + X^b)$  was fitted non-linearly, using the KaleidaGraph software version 4.0 (HULINKS). Using the curve, the total concentration of the FFRP at which the bound band intensity became 50% of the intensity of the unbound band in the absence of the FFRP i.e. that of the total DNA, was obtained; here referred to as (FFRP)<sub>50</sub>. The apparent equilibrium binding constant ( $K_B$ ) was calculated, using the equation  $K_B = 1/[(FFRP)_{50} - 0.5(\text{total DNA})]$ .

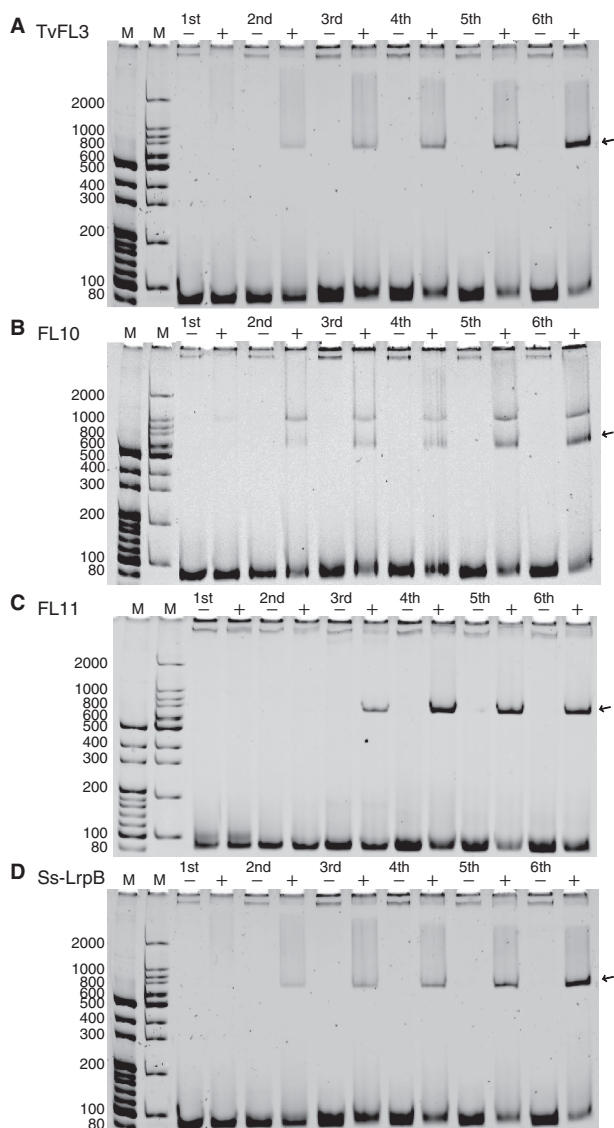
## RESULTS

### Optimal DNA duplexes for interacting with TvFL3, FL10, FL11 and Ss-LrpB

After six cycles of selection from a pool of DNA duplexes, that had insertions in the form NNNNNNNWWWN NNNNNN (Figure 2), 53–84 fragments tightly interacting with each of TvFL3, FL10, FL11 and Ss-LrpB were sequenced (Table 1). Of these, 96.2–100% had 17 bp between GAGAGAGA and AGAGAGTC, where WWW combinations were positioned at the centers. However, the remaining 0–3.8% had insertions of 16 or 18 bp, or 17 bp with combinations of 2 Ws and 1 S, i.e. G or C, at the centers. It is believed that these deviations from the original design were created by incorrect replications while repeating PCR.

From all 13-bp sequences, those best matching sets of insertions were identified as TACGA[AAT/ATT]TCGTA for TvFL3, CGA[AAT]TCGAACC/GGTTTCGA[ATT]TCG for FL10, CGAAA[AAT/ATT]TTTCG for FL11 and TGCAA[AAT/ATT]TTGCA for Ss-LrpB (Table 1). The sequences identified for TvFL3, FL11 and Ss-LrpB fitted into the form **abcdeWWWedcba**, where, e.g. **a** and **a** are bases complementary to each other. That identified for FL10 was closely related with the form.

From 1024 **abcdeWWWedcba** sequences, those best matching insertions were identified as TACGAWWWTCGTA for TvFL3, TTCGAWWWTCGAA for FL10, CGAAWWWWTTTCG for FL11, and TGCAAWWWTTGCA for Ss-LrpB (Table 1). In each insertion the 13 bp best matching the identified **abcdeWWWedcba** sequence constituted a binding site of the FFRP. The frequencies of the bases most common at **a-e** in each set of binding sites were



**Figure 2.** DNA fragments used in six cycles of selection with TvFL3 (A), FL10 (B), FL11 (C) and Ss-LrpB (D). A set of six mixtures of DNA fragments designed to have NNNNNNNWNNNNNNNN insertions, 1.93 pmol each in 12  $\mu$ l of 42 mM Na-phosphate buffer (pH 7.0) containing 125 mM NaCl and 6.7% (w/v) sucrose, was applied to a 12% polyacrylamide gel in the presence (+) and absence (-) of an FFRP, and electrophoresed. After electrophoresis the gel was stained with ethidium bromide. In order to show a general increase in the fraction of fragments interacting with each FFRP following cycles, when an FFRP was present (+), the molar ratio of DNA fragments to the FFRP dimer was kept to 1:2. However, in the real selection, the molar ratio was decreased from 1:1 (first–third cycles), through 1:0.5 (fourth) and 1:0.3 (fifth), to 1:0.2 (sixth) (see ‘Materials and Methods’ section). Also in the real section the DNA concentration was higher, 10 pmol in 12  $\mu$ l. Arrows indicate bands of fragments bound by the FFRP dimers.

52.7–100% (Figure 3). At the  $-1$  position immediately upstream of **abcde** the frequencies of G at FL10-binding sites, C at FL11-binding site, and T at Ss-LrpB-binding sites were higher than 45% (Figure 3).

The most frequent combination between **abcde** and **edcba** was always AAT/ATT (33.3–56.2%), which was followed by AAA/TTT (14.0–27.4%) (Table 2, WWW

columns). At GCTACGANNNTCGTACG sites identified as tightly interacting with TvFL3, CGTTCGANNNTCGAACC sites identified as tightly interacting with FL10, and CTTGCAANNNTTGCAAC sites identified as tightly interacting with Ss-LrpB, the frequencies of AAT/ATT at the centers were even higher, 50.0–88.0% (Table 2, NNN columns).

When selection was carried out using FL11 and DNA fragments having insertions in the NNNNNNNNNNNNNNNNNNN form (Yokoyama, K. *et al.*, unpublished results), the frequencies of C at  $-1$ , C at **a**, G at **b**, A at **c**, A at **d**, and A at **e** at binding sites were 70.0–100%, and most frequent between **abcde** and **edcba** was AAT/ATT (77.8%), which was followed by AAA/TTT (18.9%).

In this way optimal DNA duplexes for interacting with TvFL3, FL10, FL11 and Ss-LrpB were identified as TACGA[AAT/ATT]TCGTA, GTTCGA[AAT/ATT]TCGAAC, CCGAAA[AAT/ATT]TTTCGG and TTGCAA[AAT/ATT]TTGCAA, respectively.

### Interaction between FFRPs and 79-bp DNA duplexes

The apparent equilibrium binding ( $K_B$ ) constant of Ss-LrpB and a 79-bp duplex having TTGCAA[AAT/ATT]TTGCAA, i.e. the TGCAA duplex, was determined to  $3.50 \times 10^7 \text{ M}^{-1}$  (Figure 4A and B, Table 3). This value is similar to  $1.1\text{--}3.0 \times 10^7 \text{ M}^{-1}$ , reported by another group for apparent  $K_B$  of Ss-LrpB and 45–150-bp duplexes having TTGCAA[AAT/ATT]TTGCAA (19).

Among 79-bp duplexes used, FL10 showed the tightest interaction with the TTCGA duplex having GTTCGA[AAT/ATT]TCGAAC (apparent  $K_B$  of  $1.04 \times 10^7 \text{ M}^{-1}$ ), which was followed by the TACGA duplex having GTACGA[AAT/ATT]TCGTAC (apparent  $K_B$  of  $0.66 \times 10^7 \text{ M}^{-1}$ ), the GTCGA duplex having GGTCGA[AAT/ATT]TCGACC (apparent  $K_B$  of  $0.33 \times 10^7 \text{ M}^{-1}$ ), and the GACGA duplex having GGACGA[AAT/ATT]TCGTCC (apparent  $K_B$  of  $0.18 \times 10^7 \text{ M}^{-1}$ ), in this order (Figure 4C, Table 3). This observation is consistent with the binding specificity of FL10, identified using SELEX procedures. Also it suggests that the presence of T at position **a** is more important than that of T at position **b** for interaction with FL10, which is consistent with a higher frequency of T at **a** in the statistics (Figure 3D).

TvFL3 showed the tightest interaction with the TACG A duplex (apparent  $K_B$  of  $1.28 \times 10^7 \text{ M}^{-1}$ ) (Figure 4D, Table 3). Binding to the GACGA duplex (apparent  $K_B$  of  $0.90 \times 10^7 \text{ M}^{-1}$ ) or the TTCGA duplex (apparent  $K_B$  of  $0.82 \times 10^7 \text{ M}^{-1}$ ) was weaker, and that to the GTCGA duplex was the weakest (apparent  $K_B$  of  $0.73 \times 10^7 \text{ M}^{-1}$ ). Although apparent  $K_B$  of FL10 and the four duplexes varied between 100% and 17% of the highest value, those of TvFL3 varied between 100% and 57% of the highest value (Table 3), showing smaller differences.

Apparent  $K_B$  of FL11 and the CGAAA duplex having CCGAAA[AAT/ATT]TTTCGG was  $3.45 \times 10^7 \text{ M}^{-1}$  (Figure 4E and F, Table 3). That of FL11 and the TGA AA duplex having CTGAAA[AAT/ATT]TTTCAG was  $1.54 \times 10^7 \text{ M}^{-1}$ , i.e. 45% of the value obtained with the CGAAA duplex (Figure 4F, Table 3).

**Table 1.** Sequences best matching insertions in fragments tightly interacting with FFRPs, identified from all 13 bp sequences or 1024 sequences in the **abcdeWWWedcba** form, where W is A or T, and, e.g. **a** and **a** are bases complementary to each other

FFRP	No. fragments	Best 13 bp sequence	Score <sup>a</sup>	Best abcdeWWWedcba	Score <sup>a</sup>
TvFL3	84	TACGA[AAT/ATT]TCGTA	10.65	TACGAWWWTCGTA	11.05
TvFL3[G36A]	73	<u>GTTCGA</u> [AAT]TCGT/ ACGA[ATT]TCGAAC	10.70	<u>TACGAWWWTCGTA</u>	10.70
TvFL3[G36T]	80	<u>GTTCGA</u> [AAT]TCGT/ ACGA[ATT]TCGAAC	10.36	<u>TTCGAWWWTCGAA</u>	10.10
FL10	79	<u>CGA</u> [AAT]TCGAACC/ GGTTCGA[ATT]TCG	10.65	<u>TTCGAWWWTCGAA</u>	10.98
FL10[T36S]	126	<u>GTGCGA</u> [ATA]TCGA/ TGCA[TAT]TCGCAC	10.55	<u>TGCGAWWWTCGCA</u>	10.71
FL11	73	<u>CGAAA</u> [AAT/ATT]TTTCG	11.48	<u>CGAAAWWTTTCG</u>	11.73
Ss-LrpB	53	<u>TGCAA</u> [AAT/ATT]TTGCA	9.25	<u>TGCAAWWWTTGCA</u>	9.72

<sup>a</sup>An average of the highest number of matches found between the sequence and DNA fragments in regions from the last GA of GAGAGAGA through insertions to the first AG of AGAGAGTC.

### Optimal DNA duplexes for interacting with TvFL3[G36A], TvFL3[G36T] and FL10[T36S]

SELEX experiments were carried out using variants of TvFL3, where Gly36 was replaced by Ala, i.e. TvFL3[G36A], and Thr, i.e. TvFL3[G36T], respectively, and a variant of FL10, i.e. FL10[T36S]. In this paper amino-acid positions are described using the positions of corresponding residues in FL11 (Figure 1). After six cycles of selection from a pool of DNA duplexes, that had insertions in the NNNNNNNWWNNNNNNN form, 73–126 fragments tightly interacting with each variant were sequenced (Table 1).

For TvFL3[G36A] the **abcdeWWWedcba** sequence best matching insertions was identified as TACGAWWWTCGTA, which was followed by the second best, GACGAWWWTCGTC and the third best, TTCGAWWWTCGAA. When statistics were made using the sites best matching TACGAWWWTCGTA, the frequency of A at **b**, 56.9%, was smaller than that obtained with the original TvFL3, 70.2% (Figure 3B). The frequency of T at **b** increased to 25.3% from 3.6%, obtained with the original TvFL3. At **a** the frequency of T decreased to 56.2% from 67.3%, and that of G increased to 30.1% from 24.4%. At **-1** the frequency of G increased to 57.5% from 41.7%.

For TvFL3[G36T] the **abcdeWWWedcba** sequence best matching insertions was identified as TTCGAWWWTCGAA, which was followed by the second best, GTCGAWWWTCGAC, and the third best, TACGAWWWTCGTA. When statistics were made using the sites best matching TTCGAWWWTCGAA, at **b**, the frequency of T, 45.0%, was higher than that of A, 30.6% (Figure 3C). At **a** the frequency of T further decreased to 45.6%, and that of G further increased to 41.3%. At **-1** the frequency of G further increased to 61.3%.

For FL10[T36S] the **abcdeWWWedcba** sequence best matching insertions was identified as TGCGAWWWTCGCA, which was followed by the second best, TACGAWWWTCGTA, and the third best, TTCGAWWWTCGAA. When statistics were made using the sites best matching TGCGAWWWTCGCA, at **b** the frequency of T decreased to 17.4% from 60.8%, obtained with the original FL10, and that of G increased to 42.6% from 20.3%

(Figure 3E). Also the frequency of A at **b** increased to 31.2% from 10.1%. The frequency of T at **a** decreased to 69.3% from 83.5%, and that of G at **d** decreased to 83.5% from 98.7%.

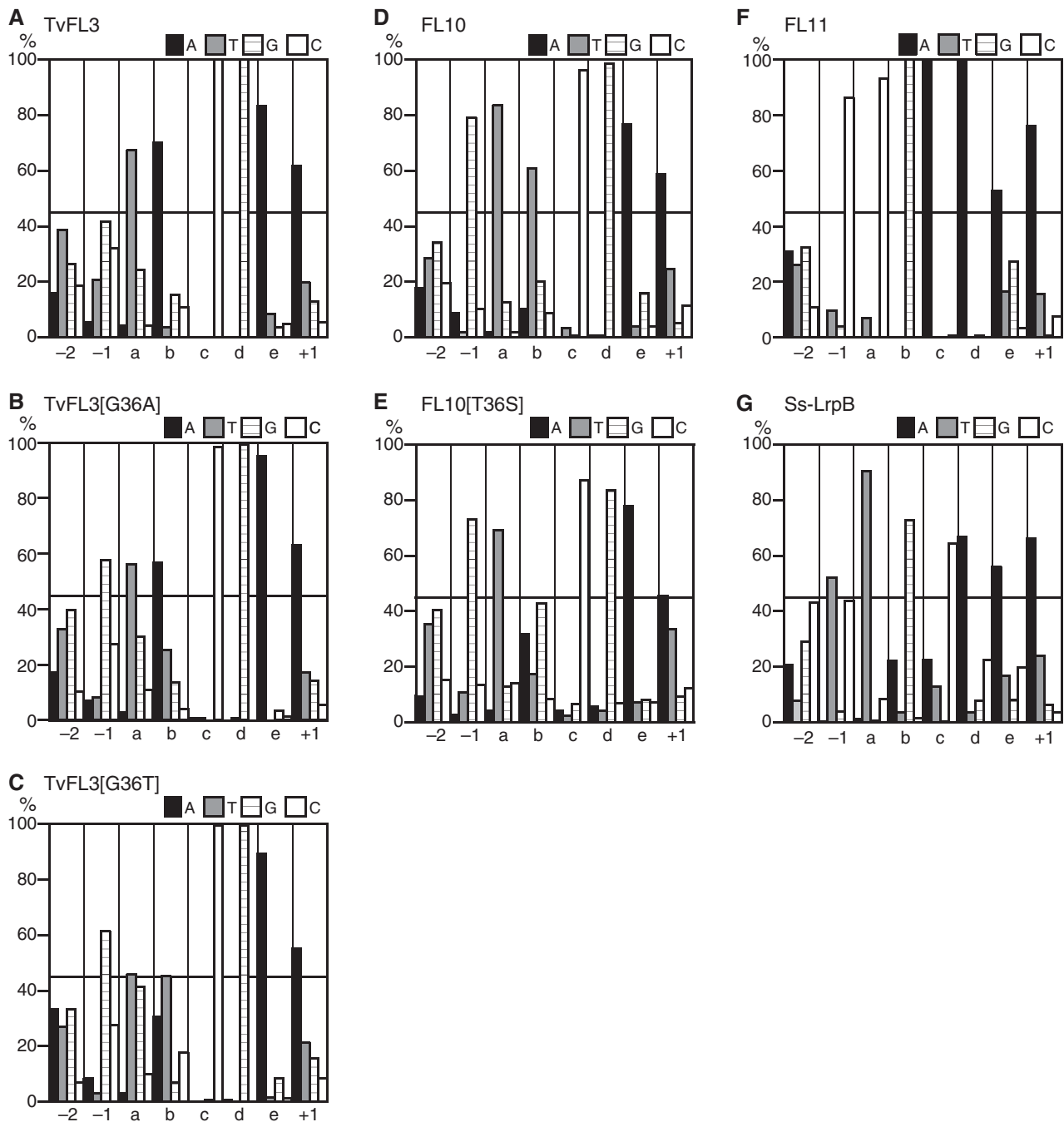
## DISCUSSION

### Recognition of CCGAAA[AAT/ATT]TTTCGG by FL11

Using the SELEX method an optimal DNA duplex for interacting with FL11 was identified as CCGAAA[AAT/ATT]TTTCGG. While, the duplex co-crystallized with FL11 was TGTGAAA[AAT]TTTCACT/AGTGAAA[AAT]TTTCACA, where central 13 bp correspond to consensus of four tight binding sites of FL11 in the *fl11* and lysine synthesis promoters (8). Compared with the optimal duplex, C is not positioned at **-1**, and at position **a** C is replaced by T.

In the crystal complex Thr37 and Ala34 of FL11 formed hydrophobic interactions with the methyl group of T at **a**, i.e. T(**a**), on its major groove side (Figure 5A). Similar hydrophobic interactions seem to be possible also with C<sub>5</sub>H and C<sub>6</sub>H of C at **a** (Figure 5B). It is difficult to predict which of C and T will be the optimal partner for a hydrophobic residue, since it depends on small differences in the binding geometry (28,29). In fact, as will be discussed in the next subsection, it is likely that Thr37 of Ss-LrpB and Thr37 of TvFL3 both bind to T(**a**). Binding of FL11 at the four sites in the *fl11* and lysine synthesis promoters is essential for transcriptional regulation of the units (8,11). At **a** of the four sites, T is most frequent (50.0%), and only 1 C is found (12.5%). With the A/T content of the *P. OT3* genome being 58.1% (30), in this genome the number of TG steps, 180702, is larger than that of CG steps, 92316, and the CCGAAAWWTTTCGG sequence is not present. The consensus sequence of biologically functioning FL11-binding sites can be different from the tightest binding sequence in order to produce binding constants appropriate for regulations.

In *P. OT3* intracellular concentrations of K<sup>+</sup> and Mg<sup>++</sup> are as high as 431 mM and 63 mM, respectively (Kawashima-Ohya, Y. *et al.*, in preparation). Depending on the concentrations of K<sup>+</sup> and Mg<sup>++</sup>, *K<sub>B</sub>* of the



**Figure 3.** Frequencies of the four bases at positions of sites identified as tightly interacting with FFRPs. Statistics of 84 sites best matching TACGAWWWTCGTA identified using DNA fragments having NNNNNNNNWWNNNNNNNN insertions (A), those of 73 sites best matching TACGAWWWTCGTA (B), those of 80 sites best matching TTCGAWWWTCGAA (C), those of 79 sites best matching TTCGAWWWTCGAA (D), those of 126 sites best matching TGCGAWWWTCGCA (E), those of 73 sites best matching CGAAAWWWTTTCG (F), and those of 53 sites best matching TGCAAWWWTTGCA (G). The frequency 45% is indicated by horizontal lines.

TATA-binding protein (TBP) and the TATA-box DNA changes considerably, although it generally increases with increasing temperature (Kawashima-Ohya, Y. *et al.*, in preparation). At high temperature with high concentrations of  $K^+$  and  $Mg^{++}$ ,  $K_B$  of FL11 and the TGAAA duplex relative to that of FL11 and the CGAAA duplex might be different from that observed at 22°C with 125 mM NaCl (Table 3). During electrophoresis, which

is essential for the methods used in this study, it is difficult to maintain high metal concentrations. A different method is needed in order to characterize effects of high metal concentrations on DNA recognition by archaeal FFRPs.

In the optimal duplex for interacting with FL11, position -1 is fixed to C. Also positions -1 of the optimal duplexes for interacting with FL10 and Ss-LrpB are fixed to G and T, respectively. In the crystal complex bases at



**Table 2.** Frequencies of base combinations between **abcde** and **edcba** at binding sites

	TvFL3		FL10		FL11	Ss-LrpB	
	WWW <sup>a</sup>	NNN <sup>b</sup>	WWW <sup>a</sup>	NNN <sup>b</sup>	WWW <sup>a</sup>	WWW <sup>a</sup>	NNN <sup>b</sup>
No. of binding sites	84	50	79	44	73	53	36
AAT/ATT	33.3	88.0	34.2	68.2	56.2	36.8	50.0
AAA/TTT	17.9	6.0	13.9	15.9	27.4	22.6	27.8
ATA/TAT	9.5	0	13.9	2.3	0	13.2	8.3
TAA/TTA	2.4	0	5.1	6.8	0	1.9	2.8
2W1S <sup>c</sup>	36.9	6.0	32.9	6.8	16.4	23.6	8.3
1W2S <sup>d</sup>	0	0	0	0	0	1.9	2.8
3S <sup>e</sup>	0	0	0	0	0	0	0

Values are given in percentages unless otherwise mentioned.

<sup>a</sup>Statistics of binding sites identified using DNA fragments having insertions in the NNNNNNNWWWNNNNNNNN form.

<sup>b</sup>Statistics of binding sites in the form, GCTACGANNNTCGTACG (TvFL3), CGTTCGANNNTCGAAC (FL10) or CTGCAANNNTTGCAAC (Ss-LrpB).

<sup>c</sup>The combinations where 2 Ws and 1 S are combined along each strand.

<sup>d</sup>The combinations where 1 W and 2 Ss are combined along each strand.

<sup>e</sup>The combinations where 3 Ws are combined along each strand.

–1 or the complementary bases did not interact with any amino-acid residue of FL11 (8), and so mechanisms of fixing –1 to these bases remain unknown.

#### A common contact pattern relating amino-acid and base positions

Optimal DNA duplexes for interacting with TvFL3, FL10, FL11 and Ss-LrpB were identified as TACGA [AAT/ATT]TCGTA, GTTCGA[AAT/ATT]TCGAAC, CCGAAA[AAT/ATT]TTTCGG and TTGCAA[AAT/ATT]TTGCAA, respectively. Although the AAT/ATT combination is conserved at the centers, it is likely that this combination does not directly interact with amino-acid residues in the FFRPs. In the crystal complex 3 A:T basepairs at the center of the duplex did not interact with FL11 (8). Any combinations of A:T basepairs were tolerated there, but when all three were replaced by G:C, interaction with FL11 was severely weakened. An A:T basepair has two hydrogen bonds only, and it is less planar than a G:C basepair, which has three hydrogen bonds. The central A:T basepairs are important for bending the duplex around the FL11 dimer by propeller twisting (8).

It is expected that outside AAT/ATT similarities of, and differences between the nucleotide sequences of the optimal duplexes reflect similarities of, and differences between amino-acid residues forming chemical interactions with DNA bases. Some amino-acid side-chains form specific interactions with DNA bases (29). For example, Arg can donate two hydrogen bonds to the G base on the major groove side (Figure 5D), and Glu can accept a hydrogen bond from the C or A base on the major groove side (Figure 5E).

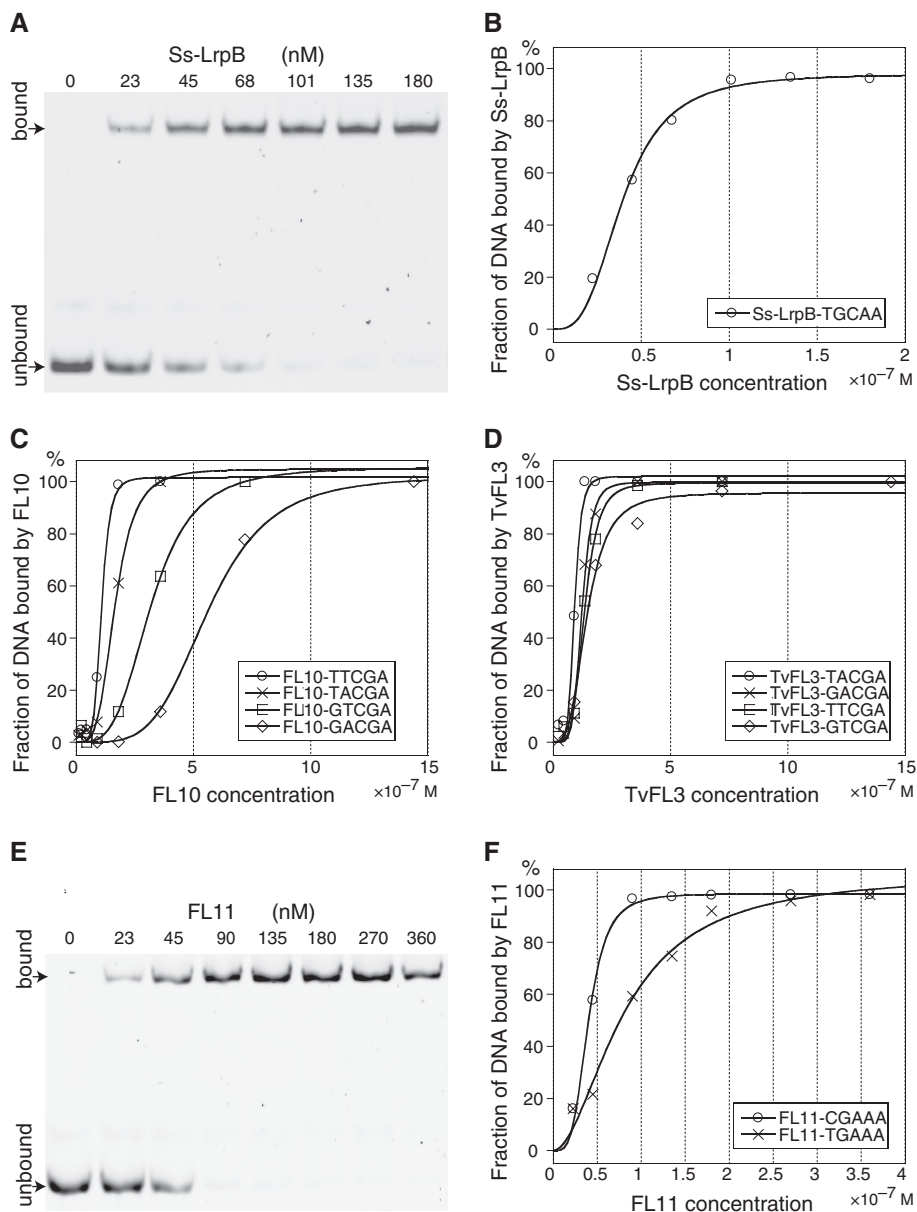
In the crystal complex (8) six amino-acid residues of each FL11 monomer, i.e. Leu24 in  $\alpha$ -helix 2, Ala34-Thr37 in the loop connecting  $\alpha$ -helices 2 and 3, and His39 in  $\alpha$ -helix 3 (Figure 1), bound to 5 bp, TGAA A/TTTCA, at each end of the duplex in its major groove (Figure 6). Of the six residues, Leu24, Ser36 and Thr37 are also present in Ss-LrpB (Figure 1). In the crystal

complex, Thr37 formed a hydrophobic interaction with T(**a**) (Figure 5A), and Leu24 formed those with T(**d**) and T(**e**). Ser36 donated a bifurcated hydrogen bond to G(**b**) (Figure 5C). All four bases are retained in the optimal duplex for interacting with Ss-LrpB (Figure 6). In the crystal complex T(**c**) was bound by His39. While T(**c**) is replaced by G in the optimal duplex for interacting with Ss-LrpB, in Ss-LrpB His39 is replaced by Arg, a specific partner of G. Glu35 of FL11 is changed to Ile in Ss-LrpB. In the crystal complex, Glu35 of FL11 formed hydrogen bonds with A(**c**), A(**d**), T(**d**) and T(**e**), indirectly through water molecules (8). Of the corresponding positions in the optimal duplex, **d** and **e** are occupied by T, the specific partner of Ile (Figure 6).

In the optimal duplexes for interacting with Ss-LrpB, TvFL3 and FL10, G bases are present at **c**, and Arg residues are found at positions 39 in the FFRPs (Figure 6). While T bases are present at **a**, positions 37 of the FFRPs are occupied by Thr and Ala (Figure 6). It is likely that these Thr and Ala form hydrophobic interactions with the T bases. TvFL3 and FL10 have Glu35, and in the optimal duplexes C bases are positioned at **d** and **c** (Figure 6). So among the four base positions indirectly bound by Glu35 of FL11, the **d** positions in the optimal duplexes for interacting with TvFL3, FL10 and also Ss-LrpB are all occupied by partner bases of residues 35 of the FFRPs (Figure 6). Glu35 of FL11 was unable to bind to a C base, because **d** is occupied by T in order to interact with Leu24 and His39. In this way the DNA-binding specificities of the archaeal FFRPs can be well explained by the same pattern of relationship between amino-acid positions and base positions.

In the crystal complex Ala34 bound to T(**a**) together with Thr37 (8). So it is likely that Pro34 of Ss-LrpB forms a hydrophobic interaction with T(**a**), and Ser34 of TvFL3 and FL10 donate hydrogen bonds to T(**a**). It is likely that C(**d**) and T(**e**) in the optimal duplex for interacting with FL10 are determined by Glu35 and Phe24 of FL10, but C(**d**) and T(**e**) in that for interacting with TvFL3 are determined by Glu35 and Asn24 of TvFL3





**Figure 4.** Interaction between FFRPs and 79-bp DNA duplexes. (A) A gel stained with a DNA-binding fluorescent dye, SYBR<sup>®</sup> Gold, after electrophoresis of the TGCAA duplex having TTGCAA[AAT/ATT]TTGCAA in the presence and absence of Ss-LrpB. On top the concentrations of Ss-LrpB are indicated. Bands of the duplex bound or unbound by Ss-LrpB are pointed to with arrows. (B) The binding profile of Ss-LrpB to the TGCAA duplex, obtained by analyzing the gel in A. (C) The binding profiles of FL10 to the TTCGA duplex having GTTCGA[AAT/ATT]TCGAAC, the TACGA duplex having GTACGA[AAT/ATT]TCGTAC, the GTCGA duplex having GGTCGA[AAT/ATT]TCGACC and the GACGA duplex having GGACGA[AAT/ATT]TCGTCC. (D) The binding profiles of TvFL3 to the TACGA, GACGA, TTCGA and GTCGA duplexes. (E) A gel stained with SYBR<sup>®</sup> Gold, after electrophoresis of the CGAAA duplex having CCGAAA[AAT/ATT]TTTCGG in the presence and absence of FL11. On top the concentrations of FL11 are indicated. Bands of the duplex bound or unbound by FL11 are pointed to with arrows. (F) The binding profile of FL11 to the CGAAA duplex, obtained by analyzing the gel in E, and that of FL11 to the TGAAA duplex having CTGAAA[AAT/ATT]TTTCAG, obtained by analyzing another gel.

(Figure 6). A hydrophobic residue such as Phe can bind to two T/C bases neighboring along the same strand (29). Asn has a hydrogen bond donor as well as an acceptor, and can simultaneously bind to T and C (29). Alternatively, it can form two hydrogen bonds with A (31), which is positioned at **e** in the optimal duplex for interacting with TvFL3 (Figure 5F).

When Gly36 of TvFL3 was replaced by Thr, A(**b**) in the optimal duplex changed to T(**b**) (Figure 6). When Thr36 of

FL10 was replaced by Ser, T(**b**) in the optimal duplex changed to G(**b**) or A(**b**) (Figure 6). It is likely that Ser36 of an FFRP forms a hydrogen bond with G or A at **b**, as indicated by the binding specificities of FL11, Ss-LrpB and FL10[T36S] (Figure 6). While, Thr36 will form a hydrophobic interaction with T(**b**), as indicated by the binding specificities of FL10 and TvFL3[G36T].

A(**b**) in the optimal duplex for interacting with TvFL3 will not interact with the side-chain proton of Gly36

**Table 3.** Apparent and relative  $K_B$  constants of FFRPs and 79-bp DNA duplexes

FFRP	DNA	Apparent $K_B$ ( $M^{-1}$ )	Relative $K_B^a$
TvFL3	TACGA <sup>b</sup>	$1.28 \times 10^7$	1
TvFL3	GACGA <sup>b</sup>	$0.90 \times 10^7$	0.70
TvFL3	TTCGA <sup>b</sup>	$0.82 \times 10^7$	0.64
TvFL3	GTCGA <sup>b</sup>	$0.73 \times 10^7$	0.57
FL10	TTCGA <sup>b</sup>	$1.04 \times 10^7$	1
FL10	TACGA <sup>b</sup>	$0.66 \times 10^7$	0.63
FL10	GTCGA <sup>b</sup>	$0.33 \times 10^7$	0.32
FL10	GACGA <sup>b</sup>	$0.18 \times 10^7$	0.17
FL11	CGAAA <sup>b</sup>	$3.45 \times 10^7$	1
FL11	TGAAA <sup>b</sup>	$1.54 \times 10^7$	0.45
Ss-LrpB	TGCAA <sup>b</sup>	$3.50 \times 10^7$	1

<sup>a</sup>The value relative to the highest apparent  $K_B$  obtained with each FFRP.

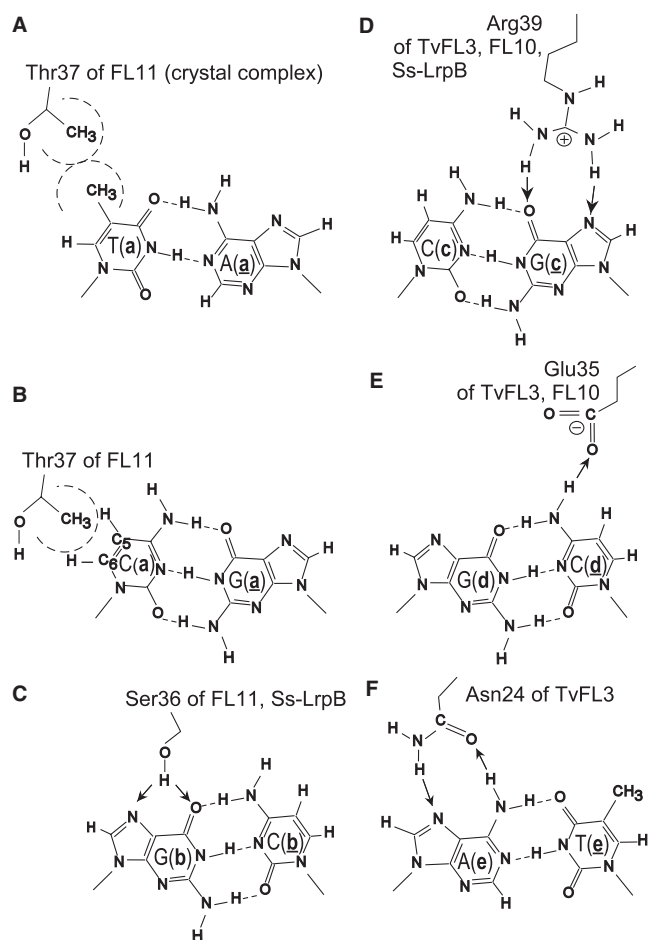
<sup>b</sup>The 79-bp duplex that has GTACGA[AAT/ATT]TCGTAC (TACGA), GGACGA[AAT/ATT]TCGTCC (GACGA), GTTCGA[AAT/ATT]TCGAAC (TTCGA), GGTCGA[AAT/ATT]TCGACC (GTCGA), CCGAAA[AAT/ATT]TTTCGG (CGAAA), CTGAAA[AAT/ATT]TTTCAG (TGAAA) or TTGCAA[AAT/ATT]TTGCAA (TGCAA).

of TvFL3. To explain the presence of A(b) in the duplex, essential physical characteristics such as high bendability, where a pyrimidine–purine step, T(a)A(b), would be required, might better be considered (32). Alternatively, the OH group of Thr37 or Ser34 might form a hydrogen bond with A(b). In the crystal complex Thr37 of FL11 donated a hydrogen bond from the side-chain OH to the DNA–phosphate group between T(a) and G(–1) (8). When another OH of Ser34 is positioned nearby, the two OH might interact with each other so that one of them will form a hydrogen bond with A(b). If this happens, although residues 34 and 37 primarily interact with the a base, and residue 37 does so with the b base, more precisely speaking, the a and b bases are recognized by residues 34, 36 and 37 in combination.

By the amino-acid replacements at positions 36 of TvFL3 and FL10 the b bases in the optimal duplexes changed as predicted from the contact pattern (Figure 3). However, the frequencies of these bases at binding sites are not so high as one might expect. The most frequent bases at a did not change, and this fact is consistent with the contact pattern. However, their frequencies decreased by the replacements (Figure 3). These facts appear to be consistent with the idea of combinatorial recognition of the a and b bases by residues 34, 36 and 37.

#### DNA-binding characteristics of other archaeal FFRPs consistent with the contact pattern

Consensus of a LysM-binding site in the *lysWXJK* promoter in the genome of *S. solfataricus* and identical sites in three related genomes was reported as GGTWYKA AWWWSGWACC, where Y is C or T, K is G or T, and S is C or G (16). However, by fitting into a self-complementary form, the consensus can be described as GGTTTCGA[AAT/ATT]TCGAACC. Consensus of 16 Ptr1-binding sites is TACGC[AAT/ATT]GCGTA, and

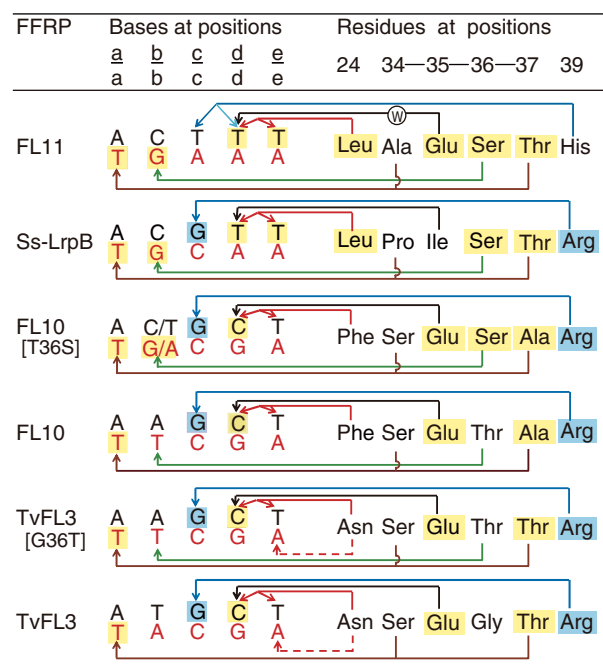


**Figure 5.** Chemical interactions formed (A and C) or predicted to be formed (B and D–F) between amino-acid side-chains of FFRPs and DNA bases on their major groove sides. A and B represent hydrophobic interactions. In C–F arrows indicate the donor to acceptor directions of hydrogen bonds. In the crystal complex (8) Thr37 of FL11 bound to T(a) (A), and the residue is predicted to bind to C(a) in the optimal duplex for interacting with FL11 (B). The pair of NH<sub>2</sub> groups or an NH<sub>2</sub> and the NH of Arg39 can be used for donating two hydrogen bonds to G(c), and D shows an example.

that of 13 Ptr2-binding sites is GGACGA[AAA/TTT]TCGTCC (15). The abcde bases in these sequences can be explained using the contact pattern discussed in the preceding subsection and residues 24, 34–37 and 39 of the FFRPs (Figure 1).

The abcde sequence of the consensus LysM-binding site in the self-complementary form is the same as those of the optimal duplexes for interacting with FL10 and TvFL3[G36T] (Table 4). The presence of C(c), G(d) and A(e) can be explained by binding of Tyr24, Glu35 and Arg39 of LysM to bases in the same way as that of Phe24, Glu35 and Arg39 of FL10 predicted (Figure 6). A LysM ortholog from *Sulfolobus acidocaldarius* has Phe24 instead of Tyr24, and so the two residues are expected to have the same functions. The presence of T(a) and T(b) can be explained by hydrophobic interactions of Ala37 of LysM and T(a), and Ala36 and T(b) (Table 4).

Ptr2 has the same residues as LysM at positions 24, 34, 35 and 39 (Table 4). Positions 36 and 37 are occupied by



**Figure 6.** Chemical contacts predicted to be formed from residues of FFRPs to bases in optimal DNA duplexes for interacting with the FFRPs in comparison with those formed in the FL11–DNA crystal complex. In the crystal complex Glu35 of FL11 formed hydrogen bonds with A(c), A(d), T(d) and T(e), indirectly through water molecules. Here the bond with T(d) only is shown, with the water molecule indicated by W. Among bases possibly bound by residues 35 of other FFRPs, bases at **d** only are indicated, since only the **d** positions in the optimal duplexes are always occupied by specific partners of residues 35 in the FFRPs. Amino-acid residues the same as or similar to those of FL11, and bases predicted to be bound by these residues are highlighted by yellow backgrounds. Arg39 of the FFRPs, and G bases at **c** in the optimal duplexes are highlighted by light-blue backgrounds. Contacts relating residues at the same positions with bases at the same positions are indicated by lines in the same colors. Chemical interactions predicted to be formed from residues 24 to the **d** and **e** bases, and those predicted to be formed alternatively to the **e** bases are differentiated by line types.

Ala in LysM but Ser in Ptr2. As has been discussed, Ser36 will form a hydrogen bond with G or A at **b**. At the 13 Ptr2-binding sites the most and next frequent bases at **b** are A and G, respectively (15). Among proteins listed in Table 4 only Ptr2 has Ser37, and the consensus Ptr2-binding site only has G(a). It is likely that Ser37 of Ptr2 donates a hydrogen bond to G(a).

Ptr1 has the same residues as TvFL3 at five of the six positions (Table 4). Instead of Asn present at position 24 of TvFL3, Ptr1 has Phe. As has been discussed for FL10 and TvFL3, this difference will not differentiate the binding specificity of the FFRPs. The consensus Ptr1-binding site has T(a)A(b)C(c)G(d)C(e) (15), which is almost the same as T(a)A(b)C(c)G(d)A(e) in the optimal duplex for interacting with TvFL3 (Table 4).

Binding of Ss-Lrp to the *ss-lrp* promoter was characterized by foot-printing and contact-probing experiments (17). Ss-Lrp has Leu24, Ser34, Pro35, Ala36, Thr37 and His39 (Figure 1). Using the contact pattern, an optimal duplex for interacting with Ss-Lrp is predicted to be TTA AAWWTTTAA, since Thr37 and Ser34 will determine **a** to T in the same way as those of TvFL3, and Ala36 will determine **b** to T in the same way as that of LysM, and Leu24 and His39 will determine **cde** to AAA in the same way as those of FL11 (Table 4). Inside the region tightly protected by Ss-Lrp from DNase I cleavage around a cluster of bases identified as tightly contacted by amino-acid residues (17), TGAAA[ATT]TTTTA/TAAAA[AAT]TTTCA, and, shifted by 1 bp, GAAAA[TTT]TTTAA/TTAAA[AAA]TTTTTC are present, which are very close to the site predicted to be optimal. While it is likely that the downstream half of the region was protected by a dimer binding around these sequences, Ss-Lrp forms a tetramer (17), and so the upstream half might have been protected by the other dimer in the same tetramer.

### Factors modifying the contact pattern

Summarizing the discussion, it is likely that various archaeal FFRPs recognize DNA in essentially the same

**Table 4.** Bases in consensus FFRP-binding sites and amino-acid residues of the FFRPs

FFRP	Bases at positions					Residues at positions					
	<b>a</b>	<b>b</b>	<b>c</b>	<b>d</b>	<b>e</b>	24	34	35	36	37	39
LysM	<u>T</u>	T	C	G	A	<b>Tyr</b>	<u>Ser</u>	<b>Glu</b>	<u>Ala</u>	<u>Ala</u>	<b>Arg</b>
FL10	<u>T</u>	T	C	G	A	<b>Phe</b>	<u>Ser</u>	<b>Glu</b>	<u>Thr</u>	<u>Ala</u>	<b>Arg</b>
TvFL3[G36T]	<u>T</u>	T	C	G	A	<b>Asn</b>	<u>Ser</u>	<b>Glu</b>	<u>Thr</u>	<u>Thr</u>	<b>Arg</b>
Ptr2	<u>G</u>	<u>A</u> <sup>a</sup>	C	G	A	<b>Tyr</b>	<u>Ser</u>	<b>Glu</b>	<u>Ser</u>	<u>Ser</u>	<b>Arg</b>
FL11	C	<u>G</u>	A	A	A	<b>Leu</b>	<u>Ala</u>	<b>Glu</b>	<u>Ser</u>	<u>Thr</u>	<b>His</b>
Ss-LrpB	<u>T</u>	<u>G</u>	C	A	A	<b>Leu</b>	<u>Pro</u>	<b>Ile</b>	<u>Ser</u>	<u>Thr</u>	<b>Arg</b>
FL10[T36S]	<u>T</u>	<u>G</u> <sup>b</sup>	C	G	A	<b>Phe</b>	<u>Ser</u>	<b>Glu</b>	<u>Ser</u>	<u>Ala</u>	<b>Arg</b>
Ptr1	<u>T</u>	A	C	G	C	<b>Phe</b>	<u>Ser</u>	<b>Glu</b>	Gly	<u>Thr</u>	<b>Arg</b>
TvFL3	<u>T</u>	A	C	G	A	<b>Asn</b>	<u>Ser</u>	<b>Glu</b>	Gly	<u>Thr</u>	<b>Arg</b>

Bases at **a** and residues 34 and 37 are italicized and underlined. Bases at **b** and residues 36, except for Gly, are underlined. Bases at **c**, **d** and **e**, and residues 24, 35 and 39 are shown in bold. It is likely that these residues interact with the bases shown in the same expressions, i.e. underlined, italicized and underlined, or bold, or the complementary bases.

<sup>a</sup>G is frequent next to A at **b** of Ptr2-binding sites (15).

<sup>b</sup>The frequency of G at **a** of FL10[T36S]-binding sites does not reach 45%, and A is frequent next to G.



way as FL11 in the crystal complex. However, it is difficult to explain the DNA-binding specificity of *E. coli* Lrp by exactly the same contact pattern. Using another SELEX procedure, Cui *et al.* (33) identified an optimal duplex for interacting with Lrp as YAGHAW[AAT/ATT]WTD CTR, where H is a non-G base, and D is a non-C base. We have obtained preliminary results suggesting that an optimal duplex for interacting with the Lrp dimer is CAG CAT[AAT/ATT]ATGCTG (Yokoyama, K. *et al.*, unpublished results). While Lrp has Asn24, Ser34, Pro35, Thr36, Pro37 and Leu39 (Figure 1), A(a) will not interact with the hydrophobic side-chain of Pro37, and G(c) will not interact with that of Leu39. Pro35 and Pro37 positioned nearby might affect the flexibility of the loop containing residues 34–37, thereby modifying the contact pattern.

Ser36 of FL11 bound to G(b) in the crystal complex (Figure 5C), and it is likely that Arg39 of an FFRP binds to G(c) (Figure 5D). However, it will be difficult to replace Arg39 by Ser, since the side-chain of Ser will be too short to reach G(c) from position 39. Similarly, Leu24 or Ile35 of Ss-LrpB, or Phe24 of FL10 will not be replaced by Ala. So it is important to identify sizes of amino-acid side-chains able to relate amino-acid and base positions as in the contact pattern. According to our analysis of the amino-acid sequences of FFRPs (Suzuki, M. *et al.*, unpublished results), positions 34, 36 and 37 are generally occupied by small residues (see sequences in Figure 1). The average volumes of residues at the three respective positions are all close to the volumes of Ser and Thr. While, the average volumes of residues at positions 24 and 39, respectively, are larger and close to the volumes of Glu-Lys. That of residues at position 35 are intermediate and close to the volumes of Asn-Glu. If all six residues are replaced by residues of fundamentally different sizes, in many cases the FFRP will become unable to bind to DNA. However, some FFRPs do have such combinations. Another contact pattern might be needed for characterizing their DNA recognition mode.

## ACKNOWLEDGEMENTS

The authors thank Dr L. Clowney and Dr J. Finch for their critical reading of the manuscript.

## FUNDING

Core Research for Evolutional Science and Technology program of the Japan Science and Technology Agency. Funding for open access charge: Core Research for Evolutional Science and Technology program of the Japan Science and Technology Agency.

*Conflict of interest statement.* None declared.

## REFERENCES

- Suzuki, M. (2003) Structure and function of the feast/famine regulatory proteins, FFRPs. *Proc. Jpn Acad.*, **B79**, 274–289.
- Brinkman, A.B., Ettema, T.J.G., de Vos, W.M. and van der Oost, J. (2003) The Lrp family of transcriptional regulators. *Mol. Microbiol.*, **48**, 287–294.
- Yokoyama, K., Ishijima, S.A., Clowney, L., Koike, H., Aramaki, H., Tanaka, C., Makino, K. and Suzuki, M. (2006) Feast/famine regulatory proteins (FFRPs): *Escherichia coli* Lrp, AsnC and related archeal transcription factors. *FEMS Microbiol. Rev.*, **30**, 89–108.
- Kawashima, T., Aramaki, H., Oyamada, T., Makino, K., Yamada, M., Okamura, H., Yokoyama, K., Ishijima, S.A. and Suzuki, M. (2008) Transcription regulation by feast/famine regulatory proteins, FFRPs, in archaea and eubacteria. *Biol. Pharm. Bull.*, **31**, 173–186.
- Calvo, J.M. and Matthews, R.G. (1994) The leucine-responsive regulatory protein: a global regulator of metabolism in *Escherichia coli*. *Microbiol. Rev.*, **58**, 466–490.
- Newman, E.B. and Lin, R. (1995) Leucine-responsive regulatory protein: a global regulator of gene expression in *E. coli*. *Ann. Rev. Microbiol.*, **49**, 747–775.
- Cho, B.-K., Barrett, C.L., Knight, E.M., Park, Y.S. and Palsson, B.Ø. (2008) Genome-scale reconstruction of the Lrp regulatory network in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **105**, 19462–19467.
- Yokoyama, K., Ishijima, S.A., Koike, H., Kurihara, C., Shimowasa, A., Kabasawa, M., Kawashima, T. and Suzuki, M. (2007) Feast/famine regulation by transcription factor FL11 for the survival of the hyperthermophilic archaeon *Pyrococcus* OT3. *Structure*, **15**, 1542–1554.
- Lahooti, M., Roesch, P.L. and Blomfield, I.C. (2005) Modulation of the sensitivity of FimB recombination to branched-chain amino acids and alanine in *Escherichia coli* K-12. *J. Bacteriol.*, **187**, 6273–6280.
- Okamura, H., Yokoyama, K., Koike, H., Yamada, M., Shimowasa, A., Kabasawa, M., Kawashima, T. and Suzuki, M. (2007) A structural code for discriminating between transcription signals revealed by the feast/famine regulatory protein DM1 in complex with ligands. *Structure*, **15**, 1325–1338.
- Yamada, M., Ishijima, S.A. and Suzuki, M. (2009) Interactions between the archaeal transcription repressor FL11 and its coregulators lysine and arginine. *Proteins*, **74**, 520–525.
- de los Rios, S. and Perona, J.J. (2007) Structure of the *Escherichia coli* leucine-responsive regulatory protein Lrp reveals a novel octameric assembly. *J. Mol. Biol.*, **366**, 1589–1602.
- Brinkman, A.B., Dahlke, I., Tuininga, J.E., Lammers, T., Dumay, V., de Heus, E., Lebbink, J.H.G., Thomm, M., de Vos, W.M. and van der Oost, J. (2000) An Lrp-like transcription regulator from the archaeon *Pyrococcus furiosus* is negatively autoregulated. *J. Biol. Chem.*, **275**, 38160–38169.
- Dahlke, I. and Thomm, M. (2002) A *Pyrococcus* homolog of the leucine-responsive regulatory protein, LrpA, inhibits transcription by abrogating RNA polymerase recruitment. *Nucleic Acids Res.*, **30**, 701–710.
- Ouhammouch, M. and Geiduschek, E.P. (2001) A thermostable platform for transcription regulation: the DNA-binding properties of two Lrp homologs from the hyperthermophilic archaeon *Methanococcus jannaschii*. *EMBO J.*, **20**, 146–156.
- Brinkman, A.B., Bell, S.D., Lebbink, R.J., de Vos, W.M. and van der Oost, J. (2002) The *Sulfolobus solfataricus* Lrp-like protein LysM regulates lysine biosynthesis in response to lysine availability. *J. Biol. Chem.*, **277**, 29537–29549.
- Enoru-Eta, J., Gigot, D., Glansdorff, N. and Charlier, D. (2002) High resolution contact probing of the Lrp-like DNA-binding protein Ss-Lrp from the hyperthermoacidophilic crenarchaeote *Sulfolobus solfataricus* P2. *Mol. Microbiol.*, **45**, 1541–1555.
- Peeters, E., Thia-Toong, T.L., Gigot, D., Maes, D. and Charlier, D. (2004) Ss-LrpB, a novel Lrp-like regulator of *Sulfolobus solfataricus* P2, binds cooperatively to three conserved targets in its own control region. *Mol. Microbiol.*, **54**, 321–336.
- Peeters, E., Wartel, C., Maes, D. and Charlier, D. (2007) Analysis of the DNA-binding sequence specificity of the archaeal transcriptional regulator Ss-LrpB from *Sulfolobus solfataricus* by systematic mutagenesis and high resolution contact probing. *Nucleic Acids Res.*, **35**, 623–633.
- Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.

21. Segerer, A., Langworthy, T.A. and Stetter, K.O. (1988) *Thermoplasma acidophilum* and *Thermoplasma volcanium* sp. nov. from solfatara fields. *Syst. Appl. Microbiol.*, **10**, 161–171.
22. González, J.M., Masuchi, Y., Robb, F.T., Ammerman, J.W., Maeder, D.L., Yanagibayashi, M., Tamaoka, J. and Kato, C. (1998) *Pyrococcus horikoshii* sp. nov., a hyperthermophilic archaeon isolated from a hydrothermal vent at the Okinawa trough. *Extremophiles*, **2**, 123–130.
23. Zillig, W., Arnold, H.P., Holz, I., Prangishvili, D., Schweier, A., Stedman, K., She, Q., Phan, H., Garrett, R. and Kristjansson, J.K. (1998) Genetic elements in the extremely thermophilic archaeon *Sulfolobus*. *Extremophiles*, **2**, 131–140.
24. Leonard, P.M., Smits, S.H.J., Sedelnikova, S.E., Brinkman, A.B., de Vos, W.M., van der Oost, J., Rice, D.W. and Rafferty, J.B. (2001) Crystal structure of the Lrp-like transcription regulator from the archaeon *Pyrococcus furiosus*. *EMBO J.*, **20**, 990–997.
25. Suzuki, M. (2005) Comparison of foot-prints of FFRPs on promoters designed for autoregulation: FL10 (pot0377090, LrpA), FL11 (pot0434017), and Ss/Sa-Lrp. *Proc. Jpn Acad.*, **B81**, 403–409.
26. Higuchi, R., Krummel, B. and Saiki, R.K. (1988) A general method of *in vitro* preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic Acids Res.*, **16**, 7351–7367.
27. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487–491.
28. König, P. and Richmond, T.J. (1993) The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility. *J. Mol. Biol.*, **233**, 139–154.
29. Suzuki, M. (1994) A framework for the DNA-protein recognition code of the probe helix in transcription factors: the chemical and stereochemical rules. *Structure*, **2**, 317–326.
30. Kawarabayashi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A. *et al.* (1998) Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3. *DNA Res.*, **5**, 55–76.
31. Seeman, N.C., Rosenberg, J.M. and Rich, A. (1976) Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl Acad. Sci. USA*, **73**, 804–808.
32. Suzuki, M. and Yagi, N. (1995) Stereochemical basis of DNA bending by transcription factors. *Nucleic Acids Res.*, **23**, 2083–2091.
33. Cui, Y., Wang, Q., Stormo, G.D. and Calvo, J.M. (1995) A consensus sequence for binding of Lrp to DNA. *J. Bacteriol.*, **177**, 4872–4880.