**Note**

**Binding of the Protein from *Thermus aquaticus* ISLtaqI to Its Inverted Repeat in Vitro**

Hiroyuki Tanabe, Yoshihiko Kakuta, Sachiko Yoshioka, and Ryutaro Utsumi

Department of Agricultural Chemistry, Kinki University, Naka-machi, Nara 631, Japan

Received September 1, 1997

We have isolated from *Thermus aquaticus* an insertion-sequence-like genetic element (ISLtaqI) that induces thermostolerance and has a high sequence similarity to IS150 belonging to the IS3 family. An open reading frame on ISLtaqI, termed ORF1, encodes the ORF1 protein, which carries a DNA-binding motif. In this study, we found an imperfect inverted repeat in ISLtaqI. We next overproduced and purified a His-tagged ORF1 protein. Gel retardation analysis demonstrated that this protein specifically bound to an DNA fragment containing the inverted repeat in ISLtaqI. These results suggest that ISLtaqI and the ORF1 protein are an insertion sequence and part of the transposase encoded by ISLtaqI, respectively.

**Key words:** ISLtaqI; inverted repeat; insertion sequence; transposon; *Thermus aquaticus*

Regulation of expression of insertion sequence (IS) functions is exercised at several levels.1,2 Generally, endogenous promoters are at least partially located within one (or both) imperfect terminal inverted repeats (IRs) which delimit most ISs. Since most of the specialized recombinases of DNA elements, transposases, specifically recognize the IRs of their cognate transposons before the cleavage and strand-transfer steps necessary for their displacement, expression may be autoregulated at the transcriptional level by transposase binding. Control can also be done at the level of translation initiation. Transcription entering some elements such as IS10 and IS50 from the outside generates an mRNA capable of forming a stem-loop structure, which sequesters the transposase translation initiation signals. Another form of translational regulation, translational frameshifting, has been observed in several ISs. Although many ISs encode their transposases in a single open reading frame (ORF) spanning almost the entire element, others have two consecutive overlapping ORFs placed in different reading phases. Slippage of the elongating ribosome between the ORFs can lead to synthesis of a transframe protein composed of part of the upstream ORF fused to the second, downstream, domain. The sequence that authorizes the backward shifting (−1 frameshifting) of the ribosome, so called 'slippery' codons, is a pentanucleotide of the type XXXXYYZ (X XX YYZ in the 0 reading phase and XXX YYY Z in the −1 reading phase). Another motif that stands out in many ISs is a triad of acidic amino acids called the DDE motif.4,5 The DDE triad has been proposed to be involved directly in catalysis.4,5

Recently, we have cloned from *Thermus aquaticus* a genetic element that is responsible for induction of thermostolerance in *Escherichia coli*.6 This element, ISLtaqI, contains two ORFs termed ORF1 and ORF2 which are significantly similar to ORFA and ORFB, respectively, of IS150,7 a member of the IS3 family. Members of the IS3 family show limited conservation in the nucleotide sequence but share a similar organization; a small upstream reading frame (ORFA) in phase 0 and a longer downstream frame (ORFB) in phase −1.8 In IS150, IS3, and IS911 of the IS3 family, translational frameshifting has been demonstrated and these ISs carry a slippery codon of AAAAAAG where ORFA and ORFB are fused.11,12 The importance of the downstream regions capable of generating a stem-loop structure in stimulating frameshifting has also been demonstrated.13,14 The portion of the fusion protein encoded by ORFA displays a motif predicted to assume a helix-turn-helix configuration, reflecting a capacity for sequence-specific binding to the ends of the cognate element. The region encoded by ORFB carries the DDE motif. The AAAAAAG frameshifting motif, the stem-loop structure, and the DDE triad were observed also in ORF2 of ISLtaqI, and the predicted amino acid sequence of the ORF1 protein had motifs for a helix-turn-helix and a leucine zipper.8

On the basis of this information, we hypothesized that the ORF1 protein has a DNA-binding ability and recognizes unidentified IRs. Inspection of the nucleotide sequence of ISLtaqI (see Fig. 3 in ref. 8 for the complete sequence) found the presence of an imperfect IR, in which its one half of 19 bp was located 5’ to ORF1 and the other half was present within ORF2, and they were separated by a 785-bp DNA domain (Fig. 1A). There is no prominent sequence similarity between the half-site of this IR and that of known IRs. We next constructed a gene expression system for high level production of the ORF1 protein. The entire ORF1 was engineered by polymerase chain reaction to bear a BamHI and a HindIII site at the 5’- and 3’- ends, respectively, and then subcloned in-frame into the BamHI/HindIII site of pQE-30 (Qiagen). The resultant plasmid was transformed into E. coli SG13009 (pREP4) and the subcloned frag-

---

1 Corresponding author.

**Abbreviations:** IR, inverted repeat; IS, insertion sequence; ORF, open reading frame, SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis.
Fig. 1. Functional DNA Element in ISLtaq1.
A) the nucleotide sequence of and around the IR in ISLtaq1. Nucleotide positions are given relative to the translational initiation point (+1) of ORF1. Half-sites of the IR is indicated by arrows facing each other. B) probe DNA fragments used in gel retardation assay. The numbering of the nucleotide position follows that in A. Arrows represent half-sites of the IR in ISLtaq1.

Fig. 2. SDS-PAGE of the ORF1 Protein.
SDS-PAGE on a 15% gel and the following silver-staining were done by the methods of Laemmli and Oakley et al., respectively. Lanes: 1, molecular weight markers; 2, cell extract prepared from E. coli SGI3099 (pREP4) overproducing a His-tagged ORF1 protein; 3, purified His-tagged ORF1 protein. The molecular weights of the marker proteins in lane 1 are indicated to the left of the lane.

The present study suggested that both of the half-sites, while fragment U2 or D2 lacks the downstream or the upstream half-site, respectively. Interaction between the ORF1 protein and the IR was investigated by means of gel retardation assay using the purified His-tagged ORF1 protein and these DNA fragments as probes. The binding reaction was done in a total volume of 10 μl of a binding buffer consisted of 10 mM HEPES-NaOH, pH 7.5, 50 mM KCl, 10% glycerol, and 0.5 mM dithiothreitol. Each reaction mixture contained 1 μg of the ORF1 protein or bovine serum albumin as a negative control as well as 1 μg of probe DNA. After incubated at 37°C for 30 min, the mixtures were put onto a 4% NuSieve 3:1 (FMC BioProducts) agarose gel containing 1 × TAE buffer (40 mM Tris, 5 mM sodium acetate, 2 mM EDTA, pH 8.0) and run at 8.5 volts/cm in 0.5 × TAE at room temperature. The result is shown in Fig. 3. Both U1 and D1 presented a retarded migration pattern when tested for affinity with the ORF1 protein, but these probes did not react with bovine serum albumin. On the other hand, U2 and D2 remained unsheared even after being incubated with the ORF1 protein as in the negative control running. Thus, we tentatively concluded that the IR in ISLtaq1 is a specific binding domain for the ORF1 protein and the half-sites of the IR are essential for recognition by the protein. U2 and U4, however, lack not only the half-site but also its flanking regions. Therefore, further investigation is necessary for the final conclusion.

By analogy with previous studies concerning IS, there should exist another IR to which the ORF1 protein responds, probably, downstream from ISLtaq1. However, because the further downstream region of ISLtaq1 has not yet been cloned, nothing is known about the putative counterpart IR. The remote position of the recognition sites for the ORF1 protein raises the possibility that DNA flexibility, such as loop formation or bending, could play an important role in the interaction. In
addition, it is tempting to speculate that the ORF1 protein may form higher-order complexes to bridge the gap that separates the two half-sites. Sequence-specific DNA binding has been reported for the amino-terminal region of the IS1 transposase\(^\text{10}\) as well as IS30\(^\text{19}\) and IS50\(^\text{20}\) transposases, and this seems to be a general characteristic of this type of transposable element. IRs are the only genetic element required in cis for transposition and they are recognized and cleaved by transposase before being inserted into the target DNA. Therefore, our data suggest that the ORF1 protein is the amino-terminal portion of a transposase protein and ISLtaq1 is its cognate transposable element.

Acknowledgments
This work was supported in part by grants from the Environmental Science Research Institute, Kinki University, and the Ministry of Education, Science, and Culture of Japan.

References