JOURNAL OF BACTERIOLOGY, Oct. 1997, p. 6238–6243 Vol. 179, No. 20
0021-9193/97/$04.00+0

Functional Domains of the InsA Protein of IS*2*

GUANG-SHENG LEI AND SHIAU-TING HU*

Department of Microbiology and Graduate Institute of Microbiology and Immunology, National Yang-Ming University, Taipei, Taiwan, Republic of China

Received 5 February 1997/Accepted 20 July 1997

The InsA protein is a transcriptional regulator. It binds to the promoter region of *insA* and *insAB′*. To understand the molecular mechanism for the interaction between InsA and its binding sequence, the functional domains of InsA were identified. The glutaraldehyde cross-linking method and the two-hybrid expression system were used to study the protein-protein interaction of InsA. The results of these experiments showed that InsA forms homodimers. Deletion of the last 44 amino acid residues at its C terminus, but not the first 12 or 57 residues at the N terminus, abolished the ability of InsA to form homodimers, indicating that the protein-protein interaction domain of InsA is located at its C terminus. Gel retardation assays revealed that deletion of the last 29 amino acid residues at its C terminus had no effect on the DNA binding ability of InsA. In contrast, deletion of the first N-terminal 12 residues abolished the DNA binding capability of InsA. These results indicate that the DNA binding domain of InsA is located at its N terminus.

IS*2* is a transposable element commonly found in *Escherichia coli* K-12 and B as well as *Shigella* spp. (10, 23, 28). It is a cause of polar effects (5, 8) and is responsible for the transposition of Tn*4521*, which carries the enterotoxin STII gene of enterotoxigenic *E. coli* (11, 18). The entire IS*2* is 1,331 bp, with a pair of 42-bp imperfect terminal repeats.

IS*2* contains five open reading frames (ORF1 to ORF5) of greater than 50 amino acids. Only two IS*2*-encoded proteins (14 and 46 kDa) have been detected. The 14-kDa protein (121 amino acids), designated InsA, is encoded by ORF1 (12), and the 46-kDa protein is encoded by ORF1 and ORF2 (Fig. 1) via a −1 translational frameshift at the frameshift signal AAAA AAG located near the stop codon of *insA* (13). This 46-kDa protein is referred to as InsAB′. It has typical transposase motifs, including WxxD, N3, and C1 (15, 20, 26, 27), and a helix-turn-helix DNA binding domain (25). It is believed to be the transposase of IS*2* (13).

The 14-kDa InsA protein has been shown to negatively regulate the major IS*2* promoter located at the left terminal repeat (13). This promoter is responsible for the expression of both InsA and InsAB′. When InsA is overexpressed, a significant decrease in the frequency of IS*2* transposition is observed (12). InsA is a DNA binding protein and binds to the sequence 5′-TAAATAA-3′, which is part of the 9-bp repeats (5′-TATC ACTTAAATAAGTGATA-3′) located at the end of the left terminal repeat and its adjoining region (Fig. 1). The binding sequence of InsA is located 6 bp upstream from the −10 sequence of the major IS*2* promoter. The binding of InsA to this sequence may hinder the binding of RNA polymerase to the promoter to initiate transcription. Since InsA is a helix-turn-helix DNA binding protein (12), InsA may bind DNA as a dimer. In this study, we have identified the domains responsible for InsA dimerization and DNA binding.

MATERIALS AND METHODS

Construction of plasmids. To detect dimerization of intact or truncated InsA, a series of plasmids were constructed with pGBT9 and pGAD424 (Clontech), which were designed by Bartel et al. (1) for expression of hybrid (fusion) proteins. These two plasmids are shuttle vectors capable of replicating autonomously in both *E. coli* and *Saccharomyces cerevisiae*. pGBT9 allows the generation of hybrid proteins containing the GAL4 DNA binding domain (amino acids 1 to 147), whereas pGAD424 enables the production of hybrid proteins with the GAL4 activation domain (amino acids 768 to 881). The DNA fragments containing various portions of InsA were generated by PCR with the primers listed in Table 1. The regions amplified, vectors used, and plasmids constructed are summarized in Table 2, and the structures of these plasmids are shown in Fig. 2.

Two-hybrid assay. The two-hybrid assay was performed by methods recommended by the commercial supplier (Clontech). One colony of *S. cerevisiae* SFY526 (1) was inoculated into 20 ml of YPD medium (20 g of Difco peptone per liter, 10 g of yeast extract per liter), and the culture was incubated at 30°C overnight with shaking. This 20-ml overnight culture was transferred into 300 ml of fresh YPD medium. The new culture was incubated at 30°C with shaking until an absorbance at 600 nm of 0.2 was reached. The cells were then pelleted, washed with 50 ml of H_2O, and resuspended in 1.5 ml of sterile 1× TE/LiAc (10 mM Tris-HCl, 1 mM EDTA, 100 mM lithium acetate [pH 7.5]). To introduce plasmids into yeast cells, 0.1 μg of plasmid DNA and 100 μg of salmon sperm carrier DNA were mixed with 100 μl of the competent yeast cells (Clontech). After addition of 0.6 ml of PEG/LiAc solution (40% [wt/vol] polyethylene glycol 4000, 1× TE buffer, 100 mM lithium acetate [pH 7.5]), the DNA-cell mixture was incubated at 30°C for 30 min with shaking (200 rpm). Dimethyl sulfoxide was then added to a final concentration of 10% (vol/vol). The mixture was heat shock treated for 15 min at 42°C and then chilled on ice. The cells were then pelleted, resuspended in 0.5 ml of TE buffer, plated on agar containing appropriate synthetic selection medium, and incubated at 30°C for 2 to 3 days until colonies appeared.

To investigate whether the hybrid proteins interacted with each other and brought the GAL4 binding and activation domains together to bind and activate the GAL4 promoter in the transformants, the β-galactosidase activity of each transformant was determined as described by Miller (22).

TABLE 1. Primers used in this study

Primer	IS*2* coordinates	Sequence
InsA-N	1240–1219	CCGGAATTCATGATTGATGTCTTAGG GCCGG
InsA-C	871–893	CAGGTCGACAAGCTTACTCCCCATCC CGGGC
InsAdN12	1200–1180	CCGGAATTCATGACCACACAGGAAAA GATCGCA
InsA77	1010–1028	CTACATGGCGGCAGCAAGTTCA
InsA92	965–983	TTACATCGTTTTCTTGCCGAGC
InsA105	911–942	TATCCACTTTTTTTTGCCCGTCCATAT TCAAC
InsAdN51	1088–1068	TATGCAATACCAGGAAGGAAGTCT
InsAdN57	1086–1062	ATACCAGGAAGGAATTCTTACTGCT
GAD-C′		CGATTCATAGATCTCTGCAG

* Corresponding author. Mailing address: Department of Microbiology, National Yang-Ming University, Taipei, Taiwan, Republic of China. Phone: 011-886-2-826-7107. Fax: 011-886-2-821-2880.

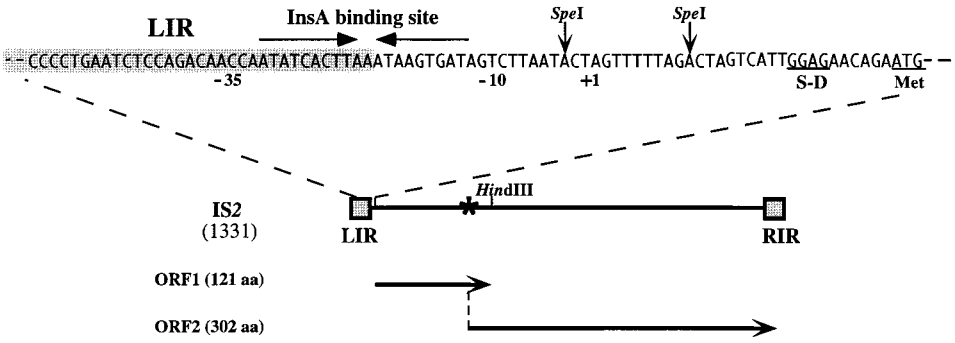

FIG. 1. Promoter region and major ORFs of IS*2*. IS*2* is 1,331 bp in length. The terminal repeats (LIR and right inverted repeat [RIR]) are represented by shaded squares. The location and the number of amino acids (aa) encoded by each ORF are indicated. The asterisk indicates the location of the bacterial frameshift signal A_6G. The nucleotide sequence shown contains the IS*2* promoter region (−10 and −35) and part of the LIR (boxed region). The 9-bp inverted repeats (5′-TATC ACTTA) are indicated by thin arrows.

Glutaraldehyde cross-linking assay. A glutaraldehyde cross-linking assay was also performed to determine whether InsA has the ability to form dimers. The cell lysates used for the glutaraldehyde cross-linking assay were prepared as follows. *E. coli* JM109/DE3 cells (31) containing pT7-7 (32), pT7insA, pT7InsA13–121, pT7InsA1–105, pT7InsA1–92, pT7InsA1–77, pT7InsA58–121, or pT7InsA58–105 (Fig. 2C) were grown to an absorbance at 600 nm of 0.4 in 200 ml of M9 medium (21) containing 50 μg of ampicillin per ml. The cells were harvested, washed with M9 buffer (21), and resuspended in 1 ml of a methionine assay medium (methionine free) (Difco Laboratories, Detroit, Mich.). After a 100-min incubation at 37°C, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to the culture to a final concentration of 1 mM to induce the production of the T7 RNA polymerase. After an additional 40-min incubation, rifampin and [^{35}S]methionine were added to 200 μg/ml and 10 μCi/ml, respectively. The culture was further incubated at 37°C for 5 min. The cells were then pelleted, resuspended in 200 ml of buffer I (20 mM HEPES [pH 8.0], 100 mM NaCl, 2 mM EDTA [pH 8.0], 7 mM β-mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride, 10% [vol/vol] glycerol), and lysed by repeated freezing and thawing followed by sonication. The cell lysates were then clarified by centrifugation in a microcentrifuge (Kubota, KM15200) at 14,000 rpm for 15 min. The cell lysates were mixed with glutaraldehyde at a final concentration of 0.005% (vol/vol) and incubated at room temperature for 10 min. The untreated or glutaraldehyde-treated cell lysates were then mixed with electrophoresis sample buffer (50 mM Tris-HCl [pH 6.8], 2% [wt/vol] sodium dodecyl sulfate [SDS], 7 mM β-mercaptoethanol, 10% [vol/vol] glycerol, 0.1% [wt/vol] bromophenol blue), boiled for 5 min, and electrophoresed on an SDS–12% (wt/vol) polyacrylamide gel (16). An autoradiography of the gel was performed to visualize radioactive bands.

Gel retardation assay. Cell lysates of *E. coli* JM109(DE3) containing pT7-7 (32), pT7insA, pT7InsA13–121, pT7InsA1–105, pT7InsA1–92, pT7InsA1–77, pT7InsA58–121, or pT7InsA58–105 (Fig. 2C) were used as the source of proteins in the gel retardation assays. The preparation of cell lysates and performance of the gel retardation assay were as described previously (6, 12).

RESULTS

Protein-protein interaction of intact or truncated InsA. The two-hybrid assay was performed to investigate whether the intact or truncated InsA can interact with each other. The host cell used in this assay was *S. cerevisiae* SFY526, which contains the *lacZ* gene driven by the GAL4 promoter. When both the GAL4 binding (bd) and activation (ad) domains bind to the GAL4 promoter, the *lacZ* gene is expressed to produce β-galactosidase. Therefore, the level of β-galactosidase activity would reflect the degree of binding of both the bd and ad domains to the GAL4 promoter. In this study, the two domains were brought together by the intact or truncated InsAs, which

TABLE 2. Various portions of InsA amplified by PCR and then cloned for two-hybrid assays or for expression

Plasmid	Region[a]	Primer Pair	Template	Vector	Cloning sites
pGTInsA1–121	1–121	InsA-N/InsA-C	pT7insA	pGBT9	*Eco*RI and *Sal*I
pADInsA1–121	1–121	InsA-N/InsA-C	pT7insA	pGAD424	*Eco*RI and *Sal*I
pGTInsA1–105	1–105	InsA-N/InsA105	pT7insA	pGBT9	*Eco*RI and *Sma*I
pADInsA1–105	1–105	InsA-N/InsA105	pT7insA	pGAD424	*Eco*RI and *Sma*I
pGTInsA1–92	1–92	InsA-N/InsA92	pT7insA	pGBT9	*Eco*RI and *Sma*I
pADInsA1–92	1–92	InsA-N/InsA92	pT7insA	pGAD424	*Eco*RI and *Sma*I
pGTInsA1–77	1–77	InsA-N/InsA77	pT7insA	pGBT9	*Eco*RI and *Sma*I
pADInsA1–77	1–77	InsA-N/InsA77	pT7insA	pGAD424	*Eco*RI and *Sma*I
pGTInsA13–121	13–121	GAD-C′/InsAdN12	pADInsA1–121	pGBT9	*Eco*RI and *Pst*I
pADInsA13–121	13–121	GAD-C′/InsAdN12	pADInsA1–121	pGAD424	*Eco*RI and *Pst*I
pGTInsA58–121	58–121	GAD-C′/InsAdN57	pADInsA1–121	pGBT9	*Eco*RI and *Pst*I
pADInsA58–121	58–121	GAD-C′/InsAdN57	pADInsA1–121	pGAD424	*Eco*RI and *Pst*I
pGTInsA58–105[b]	58–105	InsAdN57/InsA105	pGTInsA	pGBT9	*Eco*RI and *Sma*I
pADInsA58–105[b]	58–105	InsAdN57/InsA105	pGTInsA	pGAD424	*Eco*RI and *Sma*I
pT7InsA1–105	1–105	T7/InsA105	pT7insA	pT7insA	*Xba*I and *Sma*I
pT7InsA1–92	1–92	T7/InsA92	pT7insA	pT7insA	*Xba*I and *Sma*I
pT7InsA1–77	1–77	T7/InsA77	pT7insA	pT7insA	*Xba*I and *Sma*I
pT7InsA13–121[c]	13–121	GAD-C′/InsAdN12	pADInsA1–121	pT7-7	*Eco*RI and *Pst*I
pT7InsA58–121[d]	58–121	GAD-C′/InsAdN51	pADInsA1–121	pT7-7	*Nde*I and *Pst*I
pT7InsA58–105[b]	58–105	InsAdN57/InsA105	pADInsA1–121	pT7-7	*Eco*RI and *Sma*I

[a] Amino acid positions of InsA.
[b] The *Eco*RI site on these plasmids was modified by filling in the 5′ overhang ends to make the region in frame.
[c] To reduce the distance between the initiation codon and the ribosome binding site from 21 to 7 bp, the region between *Nde*I and *Eco*RI was removed and the ends were digested with mung bean nuclease and then religated.
[d] The *Nde*I site was blunted by digesting the 5′ overhang ends with the mung bean nuclease.

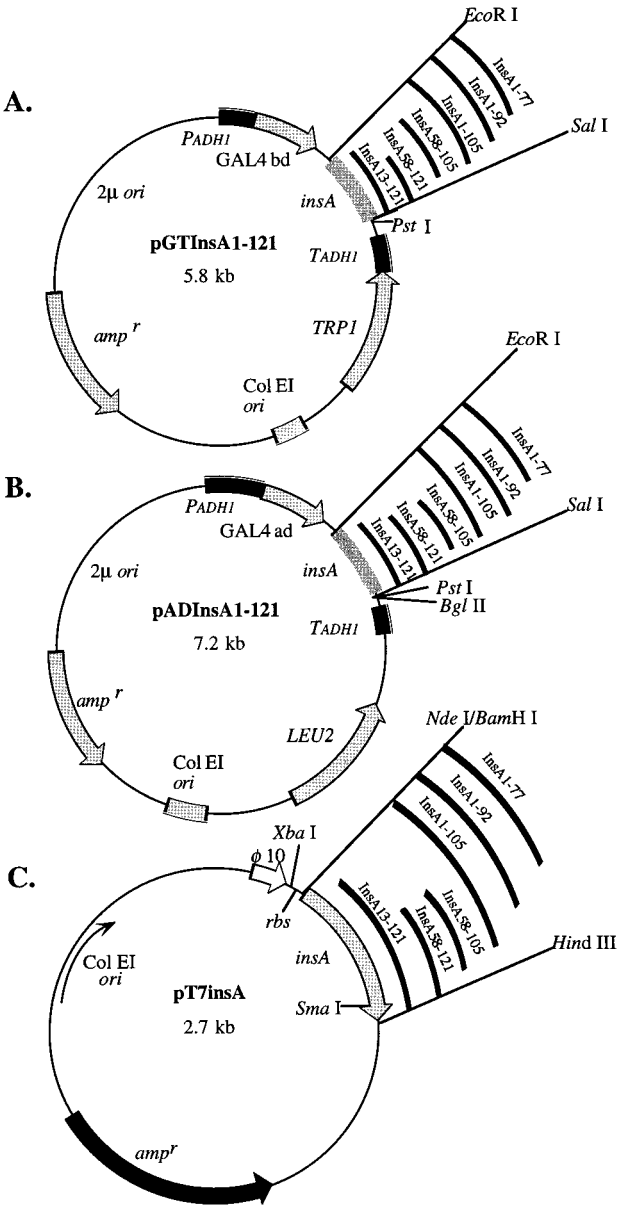

FIG. 2. Plasmids used in this study. The thick bars located between the *Eco*RI and *Sal*I sites on pGTInsA and PADInsA and the *Nde*I-*Bam*HI and *Hin*dIII sites on pT7insA represent various portions of *insA* cloned. Abbreviations: P_{ADH1}, *ADH1* promoter; rbs, ribosome binding sequence; T_{ADH1}, *ADH1* transcription termination sequence; ϕ10, ϕ10 gene promoter of bacteriophage T7.

were expressed as hybrids of bd-InsA or ad-InsA by the same DNA fragment cloned into two different plasmids (pGBT9 and pGAD424). If the InsA portions of the hybrids interact with each other, the bd and ad domains of GAL4 will be brought together to bind the GAL4 promoter to drive the expression of the *lacZ* gene.

S. cerevisiae SFY526 cells containing both the cloning vectors pGBT9 and pGAD424 were used as negative controls. The culture of these cells produced no β-galactosidase activity (Table 3), indicating that the bd and ad domains expressed from pGBT9 and pGAD424 do not interact with each other. SFY526 cells containing pTD1 (3, 19) and pVA3 (14), which contain functional bd and ad domains of GAL4, respectively, were used as the positive control. This cell culture produced β-galactosidase activity at the level of 44.6 ± 0.6 U per assay (Table 3).

Cells containing both pGTInsA and pADInsA, which contain the entire *insA* gene fused with the GAL4 bd and ad domains, respectively, produced 15.9 ± 0.2 U of β-galactosidase, indicating that InsA proteins have the ability to bind to each other. The plasmid pairs pGTInsA1–105 plus pADInsA1–105 and pGTInsA58–105 plus pADInsA58–105, which express amino acids 1 to 105 and 58 to 105, respectively, conferred upon *S. cerevisiae* SFY526 cells the ability to produce almost the same level of β-galactosidase activity (14.2 ± 0.1 and 13.4 ± 0.1 U) as did those containing pGTInsA and pADInsA. Plasmid pairs pGTInsA1–92 plus pADInsA1–92 and pGTInsA1–77 plus pADInsA1–77, which express amino acids 1 to 92 and 1 to 77, respectively, did not produce any β-galactosidase activity in SFY526 cells. However, plasmid pairs pGTInsA13–121 plus pADInsA13–121 and pGTInsA58–121 plus pADInsA58–121 enabled SFY526 cells to produce 6.3 ± 0.1 and 9.5 ± 0.3 U of β-galactosidase, respectively.

Dimerization of intact and truncated InsA. To determine whether the interaction between InsA molecules is due to InsA-InsA binding or is mediated by host proteins, a glutaraldehyde cross-linking experiment was performed. If InsA formed a homodimer, a molecule twice the molecular weight of InsA would be observed. The intact or truncated InsAs were expressed with the T7 expression system as described above. Cell lysates containing [^{35}S]methionine-labeled proteins of *E. coli* JM109(DE3) harboring pT7-7 (32), pT7insA, pT7InsA1–105, pT7InsA1–92, pT7InsA1–77, pT7InsA13–121, pT7InsA58–121, or pT7InsA58–105 (Fig. 2C) were used to study the dimerization of InsA. Cell lysates were treated or not treated with glutaraldehyde and then electrophoresed on a polyacrylamide gel.

The results of these experiments are shown in Fig. 3. No bands were seen in lanes 1 and 2, which contained cell lysate of *E. coli* JM109(DE3) harboring pT7-7 (32), indicating that no host proteins are visible in this system. Lane 3 shows a 14-kDa band of the intact InsA expressed by pT7insA. When the same cell lysate was treated with glutaraldehyde, a band of 28 kDa in addition to the 14-kDa band was observed (lane 4). The 28-kDa band represents the cross-linked dimer of InsA, and the 14-kDa band represents residual InsA proteins that were not cross-linked by glutaraldehyde. This result clearly indicates that InsA has the ability to form homodimers which can be cross-linked by glutaraldehyde.

The truncated InsA which contains the N-terminal 105 amino acids (with a 16-amino-acid deletion at the C-terminal end)

TABLE 3. β-Galactosidase activity in *S. cerevisiae* SFY526 expressing various combinations of hybrid proteins from plasmids

Plasmids	β-Galactosidase activity[a]
pGBT9 + pGAD424	0
pTD1 + pVA3	44.4 ± 0.6
pGTInsA1–121 + pADInsA1–121	15.9 ± 0.2
pGTInsA1–105 + pADInsA1–105	14.2 ± 0.1
pGTInsA1–92 + pADInsA1–92	0
pGTInsA1–77 + pADInsA1–77	0
pGTInsA13–121 + pADInsA13–121	6.3 ± 0.1
pGTInsA58–121 + pADInsA58–121	9.5 ± 0.3
pGTInsA58–105 + pADInsA58–105	13.4 ± 0.1

[a] Determined as described in the text after transformation into yeast strain SFY526. Each value represents the mean of at least three independent assays.

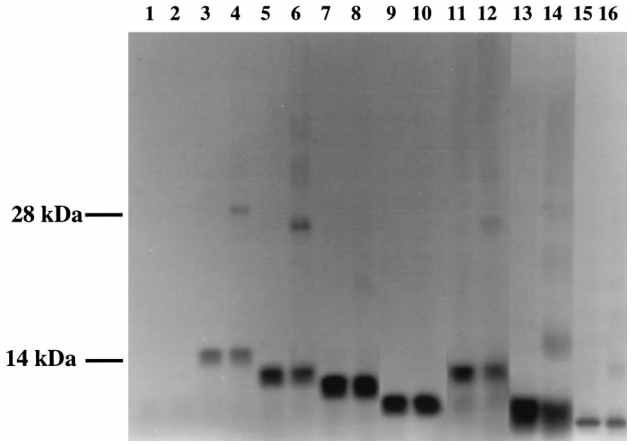

FIG. 3. Dimerization of InsA or truncated InsA. The ^{35}S-labeled InsA (lanes 3 and 4) or truncated InsA proteins including InsA1–105 (lanes 5 and 6), InsA1–92 (lanes 7 and 8), InsA1–77 (lanes 9 and 10), InsA13–121 (lanes 11 and 12), InsA58–121 (lanes 13 and 14), and InsA58–105 (lanes 15 and 16) were treated with glutaraldehyde (lanes 2, 4, 6, 8, 10, 12, 14, and 16) and then electrophoresed on an SDS–12% (wt/vol) polyacrylamide gel. The non-glutaraldehyde-treated InsA proteins (lanes 1, 3, 5, 7, 9, 11, 13, and 15) were run as controls. Lanes 1 and 2 contained cell lysates without InsA.

of InsA expressed by pT7InsA1–105 showed a similar pattern. The non-glutaraldehyde-treated cell lysate had a band of approximately 11.5 kDa (lane 5), and the glutaraldehyde-treated cell lysate showed a 23-kDa band as well as the 11.5-kDa band (lane 6). A much weaker dimerization ability was seen in the truncated InsA expressed by pT7InsA1–92. This truncated InsA1–92 contains amino acids 1 to 92 (missing the C-terminal 29 amino acids). The non-cross-linked InsA1–92 produced a 10-kDa band (lane 7), and the cross-linked InsA1–92 showed a 20-kDa band (lane 8) in addition to the 10-kDa band. The intensity of the dimer (20 kDa) band is much weaker than those of InsA1–121 and InsA1–105. A further deletion (44 amino acids) made from the C terminus, such as InsA1–77, abolished the dimerization ability of InsA. This result is shown in lanes 9 and 10; no band representing the dimeric form of InsA1–77 is observed after treatment with glutaraldehyde.

Truncated InsA which lacks the first 12 N-terminal amino acids also had the ability to form dimers. The non-cross-linked InsA13–121 expressed by pT7InsA13–121 showed a band of approximately 12.0 kDa (lane 11), and the cross-linked one had a band of 24 kDa (lane 12). Similar results were obtained with InsA58–121 and InsA58–105. Non-cross-linked InsA58–121 and InsA58–105 samples showed a 7-kDa and a 5.5-kDa protein (lanes 13 and 15), respectively, and the glutaraldehyde-treated sample had a weak 14-kDa and an 11-kDa band (lanes 14 and 16) in addition to the 7- and 5.5-kDa bands, respectively.

DNA binding domain of InsA. InsA has been shown to bind the left inverted repeat (LIR) of IS*2* (13). To identify the DNA binding domain of InsA, the intact and various truncated InsA proteins expressed from different plasmids were assayed for their DNA binding activity. The LIR of IS*2* was isolated from pKS$^+$IS2 (13) as a 59-bp *Spe*I-*Hin*fI fragment. This DNA fragment was labeled by filling in the ends with [α-^{32}P]dATP and then incubated with 10 μg of cell lysates containing the intact InsA1–121, InsA1–105, InsA1–92, InsA1–77, InsA13–121, InsA58–121, or InsA58–105. Binding of these proteins to the LIR fragment was detected by a mobility shift of the fragment on a native polyacrylamide gel.

As shown previously (12), the intact InsA1–121 bound the LIR fragment and shifted the mobility of the fragment (Fig. 4, lane 3). InsA1–105 also bound the LIR fragment and produced a mobility-shifted band slightly smaller than that produced by the intact InsA (lane 4). This is conceivable since InsA1–105 is 16 amino acids smaller than the intact InsA. InsA1–92 also had the ability to bind the LIR fragment (lane 5). InsA1–77, which lacks the C-terminal 44 amino acids of InsA, did not bind the LIR fragment (lane 6). InsA13–121, InsA58–121, and InsA58–105, which lack the first N-terminal 12, the first N-terminal 57, or both N-terminal 57 and C-terminal 16 amino acids, respectively, also did not bind the LIR fragment (lanes 7 to 9). These results suggest that the last 29 amino acids at the C-terminal end of InsA are not essential for its DNA binding activity. Deletion of any other part of InsA would probably abolish its ability to bind the LIR.

DISCUSSION

In this study, the two-hybrid expression system and the glutaraldehyde cross-linking method were used to detect the protein-protein interaction activity of InsA and gel retardation assays were performed to investigate the dimerization of InsA. In the two-hybrid expression system, InsA1–105 and InsA58–105 generated almost the same result as the intact InsA (Table 3), suggesting that the protein-protein interaction domain of InsA is located within amino acids 58 to 105.

In the glutaraldehyde cross-linking experiments, InsA58–105 was able to form dimers. This result is consistent with that of the two-hybrid expression system, i.e., that the first 57 N-terminal and the last 16 C-terminal amino acids are not essential for protein-protein interaction of InsA (Fig. 5). The observation that InsA1–77 with a 44-amino-acid deletion at the C terminus lost the dimerization ability suggests that the protein-protein interaction domain of InsA is located in the C-terminal portion of the molecule. The observation that InsA1–92 can form dimers suggests that the protein-protein interaction domain of InsA is located around amino acids 58 to 92. Although InsA92 had the ability to form dimers, it generated a negative result in the two-hybrid expression system. A possible explanation is that the binding activity of InsA92 is weakened when it is fused to the yeast GAL4 binding domain.

In the gel retardation assay, InsA13–121 had no DNA binding activity, indicating that the integrity of the N terminus of InsA is critical for its DNA binding activity. The presence of a helix-turn-helix DNA binding motif at amino acids 31 to 50

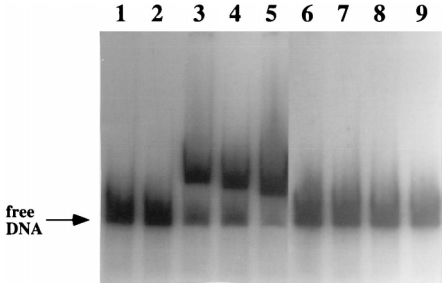

FIG. 4. Binding of InsA or truncated InsA to the left terminal repeat of IS*2*. Cell lysates containing InsA1–121 (lane 3), InsA1–105 (lane 4), InsA1–92 (lane 5), InsA1–77 (lane 6), InsA13–121 (lane 7), InsA58–121 (lane 8), and InsA58–105 (lane 9) were reacted with the ^{32}P-labeled 59-bp *Spe*I-*Hin*fI fragment containing the LIR. The protein-DNA complexes were assayed for mobility shift on an 8% (wt/vol) native polyacrylamide gel. The free DNA fragments are as indicated. The retarded protein-DNA complex bands are located at various positions above the free DNA fragment band. The cell lysate containing no InsA was used as a negative control (lane 2). Lane 1 contained the labeled LIR-containing DNA not reacted with cell lysate.

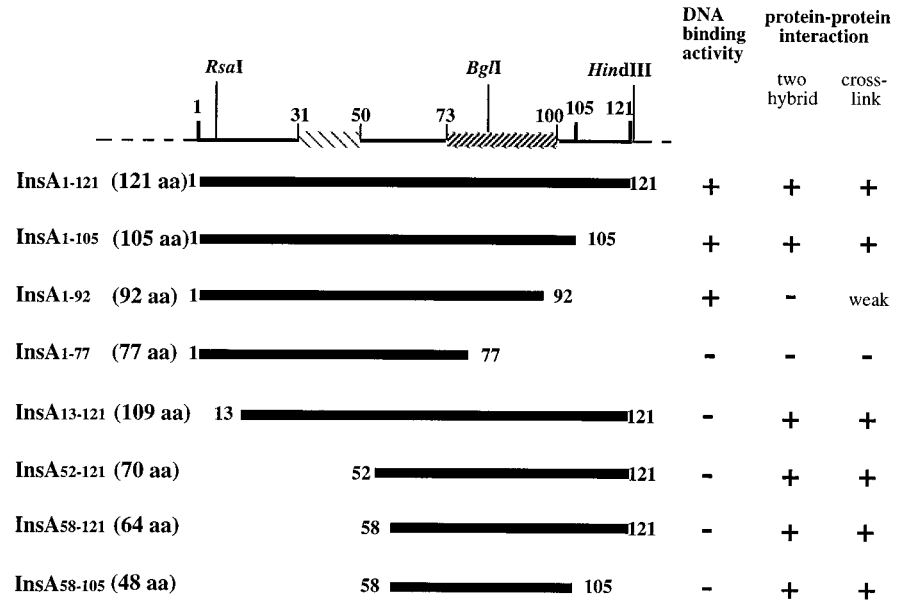

FIG. 5. Essential regions of InsA for dimerization and DNA binding. The full length of InsA or various portions of InsA assayed for DNA binding, two-hybrid formation, and cross-linking are shown as bold solid lines. The numbers of amino acid (aa) residues of various InsA derivatives are as indicated. The helix-turn-helix DNA binding motif is located within amino acids 31 to 50 (diagonal lightly shaded box), and the coiled-coil protein-protein interaction motif is located within amino acids 73 to 101 (diagonal heavily shaded box). The results derived from each assay are indicated as + or −.

(13) supports this finding. Although the deletion made on InsA13–121 was not within the DNA binding motif, this 12-amino-acid deletion may have severely affected its DNA binding ability. The observation that InsA1–92 bound the LIR as efficiently as did the intact InsA and InsA1–105 indicates that deletion of 29 amino acid residues at its C terminus had no effect on its DNA binding ability. However, a further deletion from the C terminus, such as that made on InsA1–77, destroyed its DNA binding ability. Since this 44-amino-acid deletion removed the protein-protein interaction domain of InsA, its inability to bind the LIR may be due to its inability to form a dimer, which is probably the DNA binding entity of InsA.

A pairwise residue correlation (2) analysis of InsA revealed the sequence LAAAMKQIKELQRLLGKKTMENELLKEAV located within amino acids 73 to 101. The underlined residues are either hydrophobic or neutral amino acids located at positions 73, 80, 87, 94, and 101 (separated by 6 residues). This sequence resembles the heptad repeat sequence found in many proteins that function as coiled-coil dimers. Examples of these proteins are α-keratin (4, 7), muscle proteins, lamins, intermediate filaments, dynein (9, 29, 30), and leucine zipper DNA binding proteins (17, 24). This heptad repeat usually spans 30 to 40 amino acid residues with four to six repeats. A typical heptad repeat has hydrophobic amino acid residues spaced every four and then every three residues apart; the first and third residues of the last three residues are usually charged amino acids. For leucine zippers, the heptad repeats of leucines are responsible for the formation of coiled-coil intertwining of α-helices.

Although the segment of InsA from amino acids 73 to 101 does not have a typical heptad repeat, it was identified by the PAIRCOIL″ computer program (2) as having the potential of forming coiled-coil dimers with a likelihood value of 0.51 on a scale of 0 to 1.0. In this sequence, residue 94 is asparagine, which is a neutral instead of a hydrophobic amino acid, and the first and third residues of the last three residues of the heptad repeat are not all charged amino acids. The finding that the protein-protein interaction domain of InsA is located within amino acids 58 to 105 is consistent with that identified by the PAIRCOIL″ computer program, i.e., amino acids 73 to 101.

ACKNOWLEDGMENTS

We thank C.-H. Lee for critically reading and editing the manuscript, S.-C. Lo and S.-T. Liu for helpful comments on this work, H.-S. Yuan for coiled-coil sequence analyses, and S. Tarbor for kindly providing plasmids and host cells.

This work was supported by grants NSC84-2331-B-010-058 and NSC85-2331-B-010-009 from the National Science Council, R.O.C., to S.-T. Hu.

REFERENCES

1. **Bartel, P. L., C.-T. Chien, R. Sterng, and S. Fields.** 1993. Using the two-hybrid system to detect protein-protein interactions, p. 153–179. *In* D. A. Hartley (ed.), Cellular interactions in development: a practical approach. Oxford University Press, Oxford, United Kingdom.
2. **Berger, B., D. B. Wilson, E. Wolf, T. Tonchev, M. Milla, and P. S. Kim.** 1995. Predicting coiled coils by use of pairwise residue correlations. Proc. Natl. Acad. Sci. USA **92:**8259–8263.
3. **Chien, C. T., P. L. Bartel, R. Sternglanz, and S. Fields.** 1991. The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. Proc. Natl. Acad. Sci. USA **88:**9578–9582.
4. **Crewther, W. G., A. S. Inglis, and N. M. McKern.** 1978. Amino acid sequences of alpha-helical segments from S-carboxymethylkerateine-A. Complete sequence of a type-II segment. Biochem. J. **173:**365–371.
5. **Fiandt, M., W. Szybalski, and M. H. Malamy.** 1972. Polar mutations in *lac*, *gal*, and phage lambda consist of a few IS-DNA sequences inserted with either orientation. Mol. Gen. Genet. **119:**223–231.
6. **Garner, M. M., and A. Revzin.** 1981. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. Nucleic Acids Res. **9:**3047–3060.
7. **Gough, K. H., A. S. Inglis, and W. G. Crewther.** 1978. Amino acid sequences of alpha-helical segments from S-carbosymethylkerateine-A. Complete sequence of a type-I segment. Biochem. J. **173:**373–385.
8. **Hirsch, H. J., P. Starlinger, and P. Brachet.** 1972. Two kinds of insertions in bacterial genes. Mol. Gen. Genet. **119:**191–206.
9. **Hodges, R. S., J. Sodek, and L. B. Smillie.** 1972. The cyanogen bromide fragments of rabbit skeletal-muscle tropomyosin. Biochem. J. **128:**102.
10. **Hu, M., and R. C. Deonier.** 1981. Comparison of IS*1*, IS2, and IS*3* copy

number in *Escherichia coli* strains K12, B and C. Gene **16:**161–170.
11. **Hu, S. T., and C. H. Lee.** 1988. Characterization of the transposon carrying the STII gene of enterotoxigenic *Escherichia coli*. Mol. Gen. Genet. **214:**490–495.
12. **Hu, S. T., J. H. Hwang, L. C. Lee, C. H. Lee, P. L. Li, and Y. C. Hsieh.** 1994. Functional analysis of the 14 kDa protein of insertion sequence 2. J. Mol. Biol. **236:**503–513.
13. **Hu, S. T., L. C. Lee, and G. S. Lei.** 1996. Detection of an IS*2*-encoded 46-kilodalton protein capable of binding terminal repeats of IS*2*. J. Bacteriol. **178:**5652–5659.
14. **Iwabuchi, K., B. Li, P. Bartel, and S. Fields.** 1993. Use of the two-hybrid system to identify the domain of p53 involved in oligomerization. Oncogene **8:**1693–1696.
15. **Kulkosy, J., K. S. Jones, R. A. Katz, J. P. G. Mack, and A. M. Skalka.** 1992. Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. Mol. Cell. Biol. **12:**2331–2338.
16. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
17. **Landschulz, W. H., P. F. Johnson, and S. L. McKnight.** 1988. The leucine-zipper: a hypothetical structure common to a new class of DNA binding proteins. Science **240:**1759–1764.
18. **Lee, C. H., S. T. Hu, P. J. Swiatek, S. L. Moseley, S. D. Allen, and M. So.** 1985. Isolation of a novel transposon which carries the *Escherichia coli* enterotoxin STII gene. J. Bacteriol. **162:**615–620.
19. **Li, B., and S. Fields.** 1993. Identification of mutations in p53 that affect its binding to SV40 T antigen by using the yeast two-hybrid system. FASEB J. **7:**957–963.
20. **Mahillon, J., J. Seurinck, L. Van Rompuy, J. Delcour, and M. Zabeau.** 1985. Nucleotide sequence and structural organization of an insertion sequence element (IS*231*) from *Bacillus thuringiensis* strain verliner 1715. EMBO J. **4:**3895–3899.
21. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. **Nisen, P., M. Purucker, and L. Shapiro.** 1979. Deoxyribonucleic acid sequence homologies among bacterial insertion sequence elements and genomes of various organisms. J. Bacteriol. **140:**588–596.
24. **O'Shea, E. K., R. Rutkowski, and P. S. Kim.** 1989. Evidence that the leucine zipper is a coiled coil. Science **243:**538–542.
25. **Pabo, C. O., and R. T. Sauer.** 1984. Protein-DNA recognition. Annu. Rev. Biochem. **53:**293–321.
26. **Rezsohazy, R., B. Hallet, J. Delcour, and J. Mahillon.** 1993. The IS*4* family of insertion sequences: evidence for a conserved transposase motif. Mol. Microbiol. **9:**1283–1295.
27. **Rezsohazy, R., B. Hallet, J. Mahillon, and J. Delcour.** 1993. IS*231* V and W from *Bacillus thuringiensis* subsp. *israelensis*, two distant members of the IS*231* family of insertion sequences. Plasmid **30:**141–149.
28. **Saedler, H., and B. Heiss.** 1973. Multicopies of the insertion-DNA sequences IS*1* and IS*2* in the chromosome of *E. coli* K12. Mol. Gen. Genet. **122:**267–277.
29. **Sodek, J., R. S. Hodges, L. B. Smillie, and L. Jurasek.** 1972. Amino-acid sequence of rabbit skeletal tropomyosin and its coiled-coil structure. Proc. Natl. Acad. Sci. USA **69:**3800–3804.
30. **Stone, D., and L. B. Smillie.** 1978. The amino acid sequence of rabbit skeletal alpha-tropomyosin. The NH2-terminal half and complete sequence. J. Biol. Chem. **253:**1137–1148.
31. **Studier, F. W., and B. A. Moffatt.** 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-Level expression of cloned gene. J. Biol. Chem. **189:**113–130.
32. **Tabor, S.** Personal communication.