Zinc and DNA Binding Properties of a Novel LIM Homeodomain Protein Isl-2†

Zhiyuan Gong* and Choy L. Hew

Research Institute, Hospital for Sick Children, Toronto, and Departments of Clinical Biochemistry and Biochemistry, University of Toronto, Toronto, Ontario, Canada M5G 1L5

Received July 6, 1994; Revised Manuscript Received September 27, 1994®

ABSTRACT: LIM homeodomain proteins are a family of recently characterized proteins which contain, in addition to a homeodomain, two tandem repeats of conserved Cys-His motifs termed as LIM domains. We have recently isolated several clones from a chinook salmon pituitary cDNA library that encode two novel LIM homeodomain proteins, Isl-2 and Isl-3, which are structurally related to rat Isl-1. In the present study, we used the salmon Isl-2 to determine the role of LIM domains in DNA binding. Several glutathione *S*-transferase (GST) fusion proteins containing either full length Isl-2 or various portions of this protein were expressed in bacteria. Zinc blot analysis reveals that the LIM domains produced in bacteria are capable of binding zinc. Gel shift analysis indicates that all homeodomain-containing fusion proteins are able to bind to a TAAT target sequence while the fusion proteins containing only the LIM domain are not. In contrast to a previous observation that the LIM domains of rat Isl-1 have an inhibitory role in DNA binding, full length salmon Isl-2 containing both the LIM domains and a homeodomain can bind to a TAAT target sequence. To further examine the role of LIM domains in DNA binding, several GST fusion proteins were used to select specific target DNA sequences from a pool of randomly incorporated oligonucleotides. Specific target DNAs were selected by fusion proteins containing the homeodomain or the full length Isl-2, but not by LIM domain only fusion proteins, indicating that the LIM domain alone is not involved in DNA binding. The selected target DNAs were cloned and sequenced. They revealed two classes of consensus, $^{C}/_{T}$TAAT$^{G}/_{T}{}^{G}/_{A}$ and $^{C}/_{T}$TAAGTG, for both the homeodomain and full length Isl-2. The two classes of consensus competed with each other for binding to the homeodomain. The equilibrium dissociation constants for DNA binding, estimated by Scatchard analysis, were similar for the homeodomain and full length Isl-2. These observations suggest that the LIM domains of Isl-2 have no involvement in DNA binding and they do not affect the specificity and affinity of the homeodomain for the target DNA.

Homeodomain proteins are a class of transcription factors with important roles in determining cell lineage and pattern formation during development. All homeodomain proteins contain a conserved 60 amino acid domain responsible for specific DNA binding. On the basis of their amino acid sequence homology, these proteins can be classified into at least 20 distinct families (Gehring et al., 1994). LIM homeodomain proteins are a recent addition to this class of transcription factors. The acronym LIM is derived from three homeobox genes encoding structurally similar proteins, *lin-11*, *isl-1*, and *mec-3* (Freyd et al., 1990; Karlsson et al., 1990). These three proteins contain, in addition to a homeodomain, two copies of conserved finger-like Cys-His motif termed as LIM domains. Recently, several new genes coding for LIM homeodomain proteins have been described, including *Xlim-1, -2,* and *-3* from *Xenopus* (Taira et al., 1992, 1993), *apterous* from *Drosophila* (Cohen et al., 1992; Bourgouin et al., 1992), *lmx-1* from hamster (German et al., 1992), and LH-2 from the rat (Xu et al., 1993). In addition, the LIM domains with 1−3 copies of the Cys-His motif have been identified in many proteins without the homeodomain, such as the rat cysteine-rich intestinal protein (CRIP) (Birkenmeiser & Gordon, 1986), human T-cell oncogene rhombotin-1 and -2 (McGuire et al., 1989; Boehm et al., 1991), human cysteine-rich protein (CRP) (Wang et al., 1992), cytoskeleton protein zyxin (Sadler et al., 1992), a plant pollen-specific protein SF3 (Baltz et al., 1992), and a LIM domain kinase LIMK (Mizuno et al., 1994).

The structure common to the LIM homeodomain proteins is reminiscent of two other families of homeodomain proteins, paired homeodomain proteins (Bopp et al., 1986) and POU homeodomain proteins (Herr et al., 1988). These two families of proteins also contain a second conserved DNA binding domain, paired domain, and POU-specific domain, respectively, in addition to a homeodomain. The paired domain and the homeodomain display different binding specificities to target DNA sequences (Treisman et al., 1989) while the POU-specific domain and the homeodomain act in concert to specify binding to one target sequence (Ingraham et al., 1990). The function of LIM domains is largely unknown although it has been proposed that it may be involved in DNA binding or in protein−protein interaction during transcriptional regulation (Freyd et al., 1990; Karlsson et al., 1990). Recently, it has been found that the LIM domains of rat Isl-1 inhibit DNA binding of the homeodomain (Sanchez-Garcia et al., 1993). The primary structure of the LIM domain resembles that of a zinc finger DNA binding domain (Berg, 1990). Zinc has been found in the purified LIM domain polypeptides for

† C.L.H. acknowledges support by a grant from the National Science and Engineering Research Council of Canada.

* Please address correspondence to this author at the Department of Clinical Biochemistry, Banting Institute, 100 College St., University of Toronto, Toronto, Ontario, Canada M5G 1L5. Tel: (416) 978-8622; Fax: (416) 978-8802; E-mail: gongzy@sickkids.on.ca.

® Abstract published in *Advance ACS Abstracts,* November 15, 1994.

0006-2960/94/0433-15149$04.50/0

several LIM proteins (Li et al., 1991; Michelsen et al., 1993; Archer et al., 1994; Kosa et al., 1994). However, behavior typical of zinc fingers involving direct binding to DNA remains to be demonstrated. Since the presumptive binding sequence for LIM domains is unknown, one approach to test the DNA binding capacity of the LIM domain is to use LIM domain polypeptides to select specific DNA sequences from a pool of randomly incorporated oligonucleotides. Repeated binding selection coupled to PCR amplification will enrich or purify specific DNA sequences which can be cloned and sequenced to deduce a binding consensus (Thiesen & Bach, 1990; Chittenden et al., 1991).

Recently, we have isolated several novel LIM domain homeobox cDNA clones from a salmon pituitary cDNA library. These cDNA clones encode a family of related but distinct proteins which are structurally similar to Isl-1 and are named Isl-2 and Isl-3. The expression patterns of Isl-2 and Isl-3 are similar to that of Isl-1, suggesting that they have similar roles during development and in maintenance of certain cell lineage (Gong et al., 1994a,b). In the present study, we have used Isl-2 as a model to study the potential DNA binding role of the LIM domains and their effect on DNA binding with a homeodomain. Our data indicated that the LIM domain binds to zinc, but has no specific or nonspecific DNA binding activity. The homeodomain of Isl-2 is able to bind to two similar classes of DNA consensus, $^{C}/_{T}$TAAT$^{G}/_{T}{}^{G}/_{A}$ and $^{C}/_{T}$TAAGTG. The LIM domains, when associated with the homeodomain, affect neither the specificity nor the affinity of DNA binding.

EXPERIMENTAL PROCEDURES

Plasmid Construction. Two Isl-2 cDNA clones (a and b, EMBL access numbers X64885 and X64884, respectively) were isolated from the chinook salmon, *Oncorhynchus tschawytscha,* by screening a pituitary cDNA library using an EST clone which contains a partial homeobox sequence from the winter flounder, *Pleuronectes americanus* (Gong et al., 1994a,b). The Isl-2a clone (S11) contains an unspliced intron, and the Isl-2b clone (S9) is incomplete, lacking the 5′ sequence for the first 18 amino acid residues. To construct a full length Isl-2 cDNA clone without the intron sequence, the first 96 bp (*Eco*RI/*Sph*I fragment) of the S9 clone was replaced by the first 208 bp (*Eco*RI/*Sph*I fragment) from the S11 clone. The resulting plasmid, S11/9, contains a 1.7 kb insert including the complete coding sequence for the 358 amino acids of Isl-2. All GST fusion protein constructs were made by in frame ligation of Isl-2 DNA fragments to the GST fusion protein vector pGEX-2T (Pharmacia). The detailed construction of these constructs is as follows: (1) pGST-ISL2, the 1.7 kb *Eco*RI/*Xho*I fragment of S11/9 was ligated to the *Eco*RI site of pGEX-2T. (2) pGST-ΔN18, the 1.6 kb *Eco*RI/*Xho*I fragment of S9, was blunt end ligated to the *Eco*RI site of pGEX-2T. The blunt ends were generated for both the insert and the vector using Klenow fragment in order to produce a correct reading frame. (3) pGST-ΔL1 (ΔN59), the 260 bp of *Sma*I/*Bsm*I fragment, was deleted from pGST-ISL-2. (4) pGST-ΔL1.5 (ΔN106), the 380 bp of *Eco*RI/*Sst*I fragment, was deleted from pGST-ISL2. (5) pGST-HD (ΔN160), the 1 kb Ssp I fragment of S9, was ligated to the *Eco*RI site of pGEX-2T. (6) pGST-L2 (19−159), the 0.55 kb *Eco*RI/Ssp I fragment of S9, was blunt end ligated to the *Eco*RI site of pGEX-2T as for pGST-ΔN18. (7) pGST-L1 (N83), the 0.6 kb *Eco*RI fragment of S11, was ligated to the *Eco*RI site of pGEX-2T.

The schematic representation of the fusion proteins is shown in Figure 1B, and the sequences included in these fusion proteins are indicated in Figure 1A.

Bacterial Expression and Purification of GST Fusion Proteins. The GST recombinant plasmids were transformed into *Escherichia coli* DH5α. Expression of the fusion proteins was induced by the addition of isopropyl 1-thio-β-D-galactoside (IPTG) to a final concentration of 0.1 mM in a bacterial culture with an OD_{600} at approximately 0.5. Bacteria were collected 3 h after the addition of IPTG and then suspended in 15% sucrose/50 mM Tris, pH 8.3/2 mM EDTA supplemented with 10 μM $ZnSO_4$, 5 μg/mL aprotinin, 1 mM phenylmethanesulfonyl fluoride (PMSF), and 1 mM dithiothreitol (DTT). Lysozyme was added to the suspension to a final concentration of 5 mg/mL. The bacteria were then incubated for 5 min at room temperature and lysed with Triton-X100 to a final concentration of 1%. After lysis, $MgCl_2$ and DNase I were added to final concentrations of 5 mM and 10 μg/mL, respectively, to digest viscous DNA. The lysate was then cleared by centrifugation. GST-HD was expressed in the soluble fraction without the presence of notable inclusion bodies. All other fusion proteins occurred predominantly as insoluble inclusion bodies, but were also present in the soluble fractions. To purify the GST fusion proteins, the glutathione Sepharose 4B beads (Pharmacia) were washed three times in NTN (100 mM NaCl; 20 mM Tris, pH 8.0; 0.5% NP-40; 1mM DTT) and then incubated with the soluble fractions of bacterial lysate for 20 min at 4 °C and washed twice with NTN. The fusion proteins were eluted with 15 mM reduced glutathione. Thrombin digestion of purified GST fusion proteins was carried out as previously described (Smith & Johnson, 1988).

Refolding Inclusion Body Proteins. The insoluble inclusion bodies were washed with 4 M urea in 50 mM Tris, pH 8.5, and solubilized in 8 M urea/50 mM Tris, pH 8.5/10 mM DTT. The urea was removed by gel filtration with PD-10 columns (Pharmacia) pre-equilibrated with 50 mM Tris, pH 8.5/1 mM DTT. The desalted proteins were then supplemented with 10 μM $ZnSO_4$ and renatured overnight at 4 °C. A fraction of the renatured GST fusion proteins was capable of binding to the glutathione Sepharose 4B beads and was used for binding site selection as described below.

SDS−Polyacrylamide Gel Electrophoresis and Zinc Blot Analysis. SDS−polyacrylamide gel electrophoresis was carried out by the method of Laemmli (1970). Zinc blotting was performed essentially as described by Schiff et al. (1988). Briefly, following electrophoresis, proteins were electroblotted onto a nitrocellulose filter. The proteins on the filter were renatured in metal binding buffer (100 mM Tris, pH 6.8; 40 mM NaCl; 2 mM DTT) for at least 3 h and probed for 40 min with 1 μCi/mL $^{65}ZnCl_2$ (1.85 mCi/mg, Du Pont) in the binding buffer, followed by three washes (5 min each) in the same buffer and autoradiography. After autoradiography, the filter was stained with 0.1% amido black in 25% 2-propanol/10% acetic acid and destained with water.

Gel Shift Assay. The TAAT target DNA was synthesized based on the P1 element in the promoter region of the rat insulin I gene (Ohlsson et al., 1991). The upper strand of oligonucleotide sequence is 5′-TCGACGCCCTTAATGGGC-

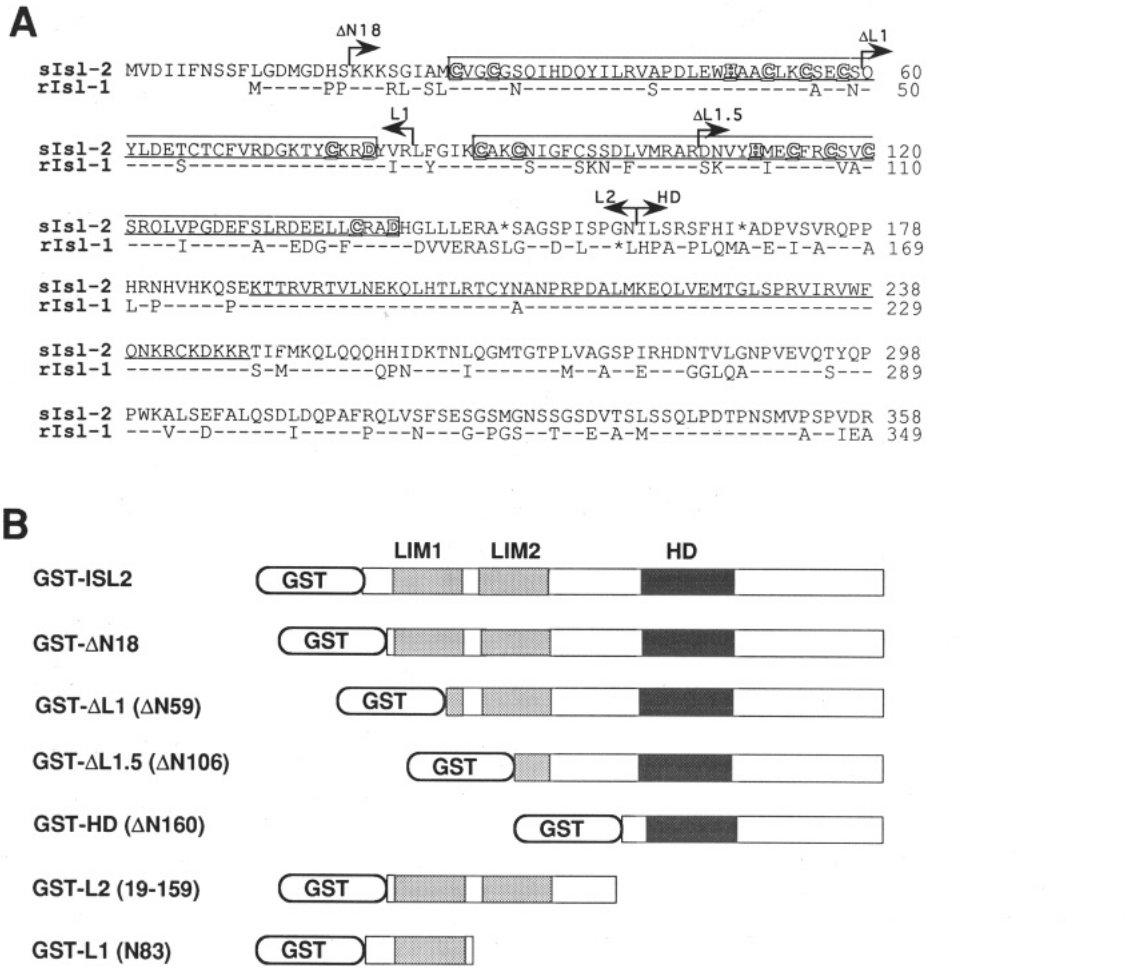

FIGURE 1: Amino acid sequence of salmon Isl-2 aligned with that of rat Isl-1 (panel A) and diagram of GST fusion proteins (panel B). In panel A, the two LIM domains are boxed, and the homeodomain is underlined. Residues likely to bind zinc are represented by double-lined letters. Breaking points for deletions in GST fusion proteins are indicated by arrows. In panel B, the two LIM domains (LIM1 and LIM2) are represented by light-shaded bars and the homeodomain (HD) by dark-shaded bars. Detailed description of these fusion proteins are given in the text.

CAAACGGCAG-3′, and the lower strand of oligonucleotide is 5′-TCGACTGCCGTTTGGCCCATTAAGGGCG-3′. The two oligonucleotides were annealed in the buffer (150 mM Tris-HCl, pH 7.6; 15 mM $MgCl_2$) by heating at 65 °C for 10 min and cooling at 37 °C for 20 min. Double strand DNA was purified from an 8% polyacrylamide gel after electrophoresis. The annealed P1 oligonucleotides had 5′ protruding ends and were radioactively labeled by Klenow filling-in. For gel shift assays, glutathione Sepharose 4B bead purified fusion proteins were preincubated on ice with 2 μg of poly(dI-dC) and 1 μg of acetylated bovine serum albumin in 100 mM NaCl, 20 mM Tris (pH 8), 10 μM $ZnSO_4$, and 2.5 mM DTT. After 30 min, radioactive target DNA (20 000 cpm) was added to continue the incubation for another 30 min, prior to loading onto a 6% polyacrylamide gel. After electrophoresis in 0.5× TBE (45 mM Tris–borate, pH 8; 1 mM EDTA) at 100 V, the gel was exposed to an X-ray film for autoradiography.

Binding Site Selection. Selection of target DNA sequences was performed essentially as described by Chittenden et al. (1991). The oligonucleotides were synthesized by HSC/Pharmacia Service Center, Toronto. The 68mer oligonucleotides (16N) contain 16 randomly incorporated nucleotides flanked by two defined PCR primer sequences: 5′-GACAGTTGTTGTCTTCCTACTAGCTGCAG-NNNNNNNNNNNNNNNNCTCGAGTGGGCGTAGTACTTCAT-3′. The underlined sequences are two PCR primers used for amplification. Two restriction sites, *Pst*I and *Xba*I, are between the PCR primers and the 16N core sequence. Double strand oligonucleotides were complemented by Klenow fragment using the right PCR primer. For each round of binding site selection, 50 μL of glutathione Sepharose 4B beads was washed three times each with 1 mL of NTN and suspended in 0.5 mL of NTN. GST fusion proteins were added to the suspension either using the soluble fraction of a bacterial lysate or using the refolded proteins from solubilized inclusion bodies, and rocked at 4 °C for 20 min. The Sepharose beads were then washed with 3 × 1 mL of NTN and suspended in 0.5 mL of NTN containing 4 μg of poly(dI-dC), followed by the addition of 40 ng of gel purified oligonucleotides. After 30 min of incubation, the beads were washed with four times each with 1 mL of ice-cold NTN. DNA which remained bound to the beads was eluted by incubation for 3 h at 50 °C in 400 μL of proteinase K digestion buffer (500 mM Tris, pH 8.8; 20 mM EDTA; 10 mM NaCl; 1% SDS; 200 μg/mL proteinase K), followed by phenol extraction and ethanol precipitation. The DNA

was then resuspended in 20 μL of water, and 10 μL of this material was used for PCR. PCR was performed with 30 cycles of the following profile: 94 °C, 1 min; 50 °C, 2 min; 60 °C, 2min. Following the 30th cycle of amplification, fresh primers and Taq DNA polymerase were added to amplify one additional cycle to ensure that the majority of the amplified 68 bp oligonucleotides were perfectly complemented double strand molecules rather than heteroduplexes containing mismatched bases at the 16 bp core segment. The PCR-amplified 68 bp DNA was purified by polyacrylamide gel electrophoresis and subjected to the next round of DNA binding. To test the affinity of the fusion protein for selected oligonucleotides, gel purified 68 bp DNA fragments were labeled with ^{32}P by nick-translation and applied to the fusion protein Sepharose beads as above. The radioactive DNA bound to the beads was counted, without prior elution, by scintillation counting.

DNA Cloning and Sequencing. PCR-amplified DNA fragments were directly cloned into pT7Blue-T vector using a kit from Novagen based on the manufacturer's instruction. White colonies were picked up for plasmid preparation. Double strand DNA sequencing was performed by dideoxynucleotide chain termination using the T7 Sequencing Kit from Pharmacia.

Competitive DNA Binding Analysis. DNA was labeled by 8 cycles of PCR using [α-^{32}P]dCTP (~1000 Ci/mmol). A fixed and limiting amount of fusion protein was incubated with equal amounts of radiolabeled DNA and increasing amounts of unlabeled competitor DNA which had been synthesized by PCR from the plasmids containing selected target sequences. Each incubation was carried out in a total volume of 0.4 mL containing 4 μg of poly(dI-dC) as performed for binding site selection except that DNA bound to the beads after two washes were counted using a scintillation counter. In order to estimate the amount of free unbound DNA for Scatchard plot, the maximal binding capacity of labeled DNA in a fixed cpm input was determined by increasing the amounts of fusion protein until a plateau of radioactive binding was achieved. The amount of free DNA was determined from the difference between the maximal binding capacity and the bound radioactivity.

RESULTS

Bacterial Expression of GST-Isl-2 Fusion Proteins and Zinc Blot Analysis. We have previously isolated several cDNA clones from a salmon pituitary cDNA library, and these clones encode a family of structurally related LIM homeodomain proteins which include Isl-1 and two related proteins named Isl-2 and Isl-3 (Gong et al., 1994b). The alignment of salmon Isl-2 and rat Isl-1 is shown in Figure 1A, and they share 75% amino acid sequence identity. The 60 amino acid homeodomains differ only at one amino acid. The two LIM domains and the Gln-rich C-terminal are also conserved between these two proteins, but the region between the second LIM domain and the homeodomain in the two proteins is highly divergent. In order to study the zinc and DNA binding properties of the salmon Isl-2 and to define the regions of zinc and DNA binding, we have expressed in bacteria several GST fusion proteins containing the following: (i) full length Isl-2 (GST-ISL2); (ii) deletion of the first

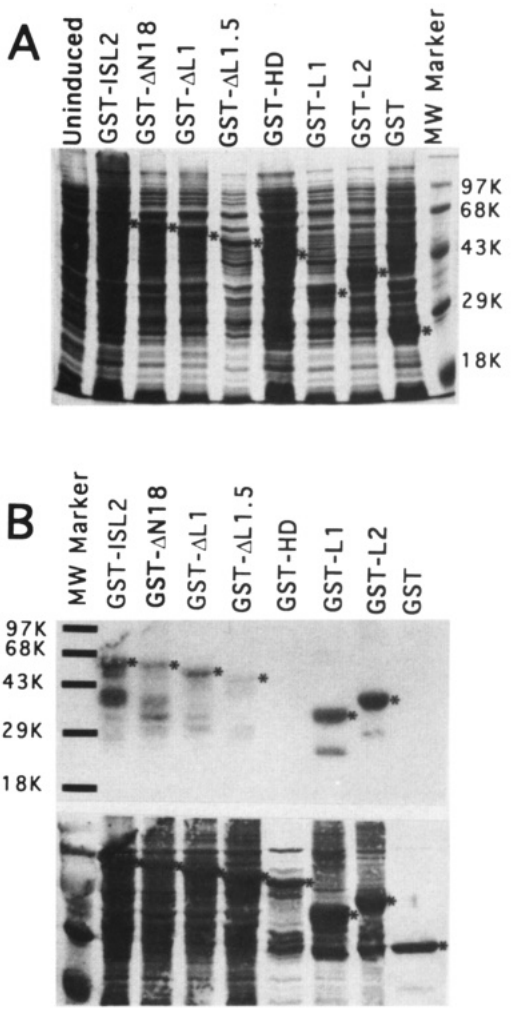

FIGURE 2: Bacterial expression of GST fusion proteins and zinc blot analysis. (A) SDS−polyacrylamide gel electrophoresis of total lysates of the bacteria expressing GST fusion proteins. Names of GST fusion proteins are indicated at the top of each lane, and their positions are indicated by asterisks on the right of each lane. Uninduced, bacterial lysate for GST-ISL2 without IPTG induction. MW marker, molecular weight markers in kilodaltons (K) as shown on the right of the gel. The gel was stained with Coomassie Blue. (B) Zinc binding of the fusion proteins. The proteins were electroblotted onto a nitrocellulose filter, and zinc blot analysis was performed as described in the Experimental Procedures. GST-HD and GST were partially purified by glutathione Sepharose 4B beads, and all others were from crude inclusion bodies. Molecular weights are indicated on the left. Upper panel, autoradiogram of a zinc blot; lower panel, the same blot was stained with amido black after autoradiography.

18 amino acids prior to the first LIM motif (GST-N18); (iii) deletion of the first LIM motif to amino acid 59 (GST-ΔL1); (iv) deletion of one and a half LIM domains to amino acid 106 (GST-ΔL1.5); (v) homeodomain and C-terminal after complete deletion of both LIM domains to amino acid 160 (GST-HD); (vi) both the first and the second LIM domains from amino acid 19 to 159 (GST-L2); and (vii) the first Cys-His motif from amino acid 1 to 83 (GST-L1), respectively (Figure 1B). The inclusions of the Isl-2 sequence in these fusion proteins are indicated in Figure 1A. In addition, the glutathione *S*-transferase (GST) without any fusion sequence was produced using the vector plasmid pGEX-2T (purchased from Pharmacia). As shown in Figure 2A, all of the GST fusion proteins were successfully expressed in bacteria. In each case, the fusion protein was one of the most predomi-

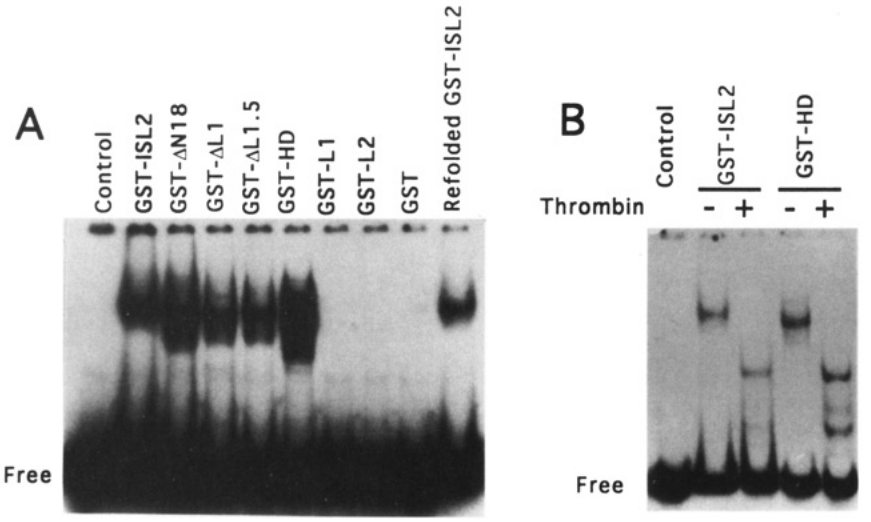

FIGURE 3: DNA binding analysis by gel shift. (A) Target DNA binding of Isl-2 and truncated Isl-2's as recombinant GST fusion proteins. Approximately equal amounts of partially purified GST fusion proteins were incubated with 20 000 cpm of ^{32}P labeled P1 element and electrophoresed onto a low salt, 6% polyacrylamide gel as described in the Experimental Procedures. Recombinant proteins, as indicated at the top of each lane, were prepared from soluble fractions, except for "Refolded GST-ISL2" which was prepared from solubilized inclusion bodies. Control lane, no recombinant protein is added. Free unbound oligonucleotides are indicated on the left. (B) Target DNA binding of full length and the homeodomain of Isl-2 after thrombin digestion. Partially purified GST-ISL2 and GST-HD were digested by thrombin and assayed as above. +, thrombin digested; −, undigested.

nant proteins in total bacterial extracts. To test whether these recombinant LIM proteins were capable of binding zinc, zinc blot analysis was carried out. All fusion proteins containing the Cys-His motif of the LIM domains were capable of binding zinc, while no zinc binding was detected for GST-HD and GST proteins (Figure 2B). This indicates that zinc was bound specifically to the LIM domain. It is interesting to note that GST-ΔL1.5 contains only half of the second LIM domain but retains the ability to bind zinc. This result supports a previous proposal that each LIM motif consists of two zinc fingers (Wang et al., 1992) and is consistent with the estimation that each LIM motif contains two zinc ions in the purified recombinant LIM proteins (Michelsen et al., 1993; Kosa et al., 1994). The 16 residues presumed to bind zinc in the two LIM domains of Isl-2 are shown as double-lined letters in Figure 1A. GST-ΔL1.5 contained the last 6 zinc binding residues in the second LIM domain. These observations confirm that the LIM domains of salmon Isl-2 are able to bind zinc.

DNA Binding Properties of Recombinant Isl-2 Proteins. It has been shown that rat Isl-1 is able to bind to P1 element of the rat insulin gene promoter which contains a TAAT core sequence (Ohlsson et al., 1991). Since the homeodomain of salmon Isl-2 is almost identical to that of rat Isl-1 (Figure 1A), it may also bind to the P1 element. To demonstrate this, the P1 element was synthesized and used for binding analysis. Preliminary experiments using total soluble bacterial lysates were unsuccessful because the GST fusion proteins, except for GST-HD, were predominantly expressed as inclusion bodies with very little occurring in the soluble fractions. Addition of more crude soluble GST-ISL2 in the binding assay often resulted in multiple bands including one with an identical mobility to that in the presence of purified GST-ISL2 (data not shown). Therefore, in order to show a specific band shift, the fusion proteins from the soluble fraction were enriched using the glutathione Sepharose 4B beads. As shown in Figure 3A, all fusion proteins containing the homeodomain were capable of binding the P1 element. In contrast, no DNA binding was detected for the fusion proteins containing only the LIM domain (GST-L1 and GST-L2) and GST itself. The binding was specific as evidenced by competition assays with homologous or nonhomologous DNA (data not shown). It was unexpected that the full length Isl-2 fusion protein (GST-ISL2) bound the target DNA since a previous report indicated that the LIM domains of rat Isl-1 had an inhibitory role in DNA binding when associated with a homeodomain (Sanchez-Garcia et al., 1993). To examine whether the discrepancy was caused by GST in the recombinant Isl-2 proteins, the GST portions of GST-ISL2 and GST-HD were removed by digestion with thrombin and the resulting GST-free proteins were used for DNA binding assay. As shown in Figure 3B, the GST-free proteins retained a strong binding activity. Thus, the GST portion had no influence on DNA binding properties of the fusion proteins. The GST-ISL2 refolded from solubilized inclusion bodies produced a gel shift band identical to that by purified GST-ISL2 from the soluble fraction, indicating that the GST-ISL2 from both fractions have similar properties in DNA binding (Figure 3A). Together, these observations suggest that the homeodomain of Isl-2 binds to a TAAT target DNA and the LIM domains of Isl-2 have no inhibitory role in DNA binding.

Binding Site Selection by the LIM Domain, the Homeodomain, and Full Length Isl-2. Although the LIM domain fusion proteins did not bind to a TAAT core sequence, it may interact with a different target DNA motif, like the paired domain and POU-specific domain in other homeodomain proteins (Treisman et al., 1989; Verrijzer et al., 1992). To test this hypothesis, it is necessary to identify the potential target DNA sequences of Isl-2. Therefore, the GST fusion proteins, including GST-L1, GST-L2, GST-HD, and GST-ISL2, were used here to select DNA target sequences from a pool of randomly incorporated oligonucleotides (16N) which were flanked by two defined primer sequences for PCR amplification (Chittenden et al., 1991). The purpose of this selection included the following: (i)

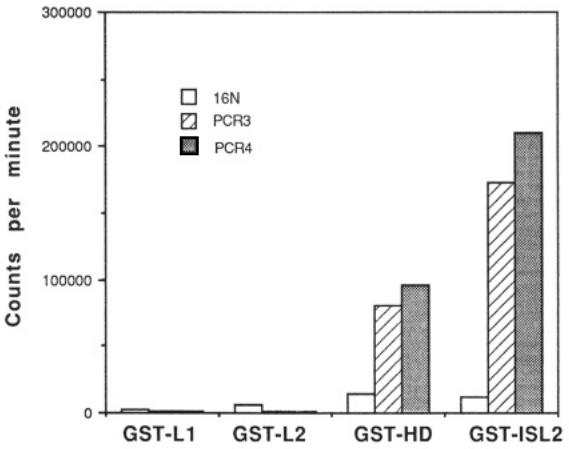

FIGURE 4: Quantitative DNA binding assay using enriched binding oligonucleotides. The enriched populations of binding oligonucleotides by GST-L1, GST-L2, GST-HD, and GST-ISL2 after the 3rd round (PCR3) and the 4th round (PCR4) of binding site selection were labeled with ^{32}P by nick-translation, and 5×10^5 cpm of the labeled oligonucleotides were incubated respectively with the glutathione Sepharose beads of GST-L1, GST-L2, GST-HD, and GST-ISL2, as described the Experimental Procedures. The starting oligonucleotides (16N) were also included as controls.

identification of potential target DNA sequences for LIM domains; (ii) characterization of optimal sequences for binding by the homeodomain; and (iii) examination of the effect of LIM domains on the specificity of the homeodomain to target DNA.

The binding site selection was carried out by repeated bindings of the 16N random oligonucleotides to glutathione Sepharose 4B beads which has been preincubated with GST fusion proteins. After 4 rounds of selection, the selected oligonucleotides from the 3rd and the 4th rounds were radioactively labeled and applied to the glutathione Sepharose−GST fusion protein beads to test their binding activity. As shown in Figure 4, significant increases in binding activity were observed for GST-HD and GST-ISL2 while there was no corresponding increase of binding activity for GST-L1 and GST-L2, indicating that specific target DNAs were successfully selected by the homeodomain but not by the LIM domain. To test whether the LIM domains are capable of binding DNA nonspecifically, the DNA binding assay was also carried out in the absence of nonspecific competitor poly (dI-dC), and again there was no detectable of DNA binding activity (data not shown). Numerous attempts using refolded GST-L1 and GST-L2 proteins also failed to detect DNA binding activity, indicating the lack of potential to convert the LIM domains into an active form for DNA binding. Therefore, despite the similarity of LIM domains and zinc finger motifs, there was no evidence to suggest a role for LIM domain in either specific or nonspecific DNA binding.

Two Identical Classes of DNA Consensus Sequences Were Selected by the Homeodomain and the Full Length Isl-2. In order to deduce the DNA binding consensus of Isl-2, the oligonucleotides selected by GST-HD after the 5th and the 6th rounds were cloned into pT7Blue-T vector. Twelve random clones from the 5th round and 36 clones from the 6th round were sequenced (H5 and H6 series accordingly). The sequences were aligned, and two classes of consensus could be easily deduced (Figure 5A). The first class of consensus is $^{C}/_{T}$TAAT$^{G}/_{T}{}^{G}/_{A}$ (Class I), where TAAT core nucleotides are invariant and some preference of the flanking nucleotides is noted. The second class, or Class II, of consensus is $^{C}/_{T}$TAAGTG, which has not been reported previously for any homeodomain protein. The DNA binding sites after five rounds of selection by GST-ISL2 (F5 series) were also cloned and sequenced (Figure 5B). To examine whether the homeodomain plus LIM domains can recognize a longer consensus which may need more rounds of selection, we carried out the selection to eight rounds. The selected oligonucleotides were then cloned and sequenced (F8 series). The apparent sequence specificity does not increase beyond that obtained with five rounds of selection. The target DNA sequences identified from the 5th and 8th rounds of selection were aligned and are shown in Figure 5B. The two classes of consensus sequences were the same as that selected by homeodomain alone. Class I contains the invariant TAAT core motif, and class II has at least 5 of the 6 bases matching to the invariant bases TAAGTG. These observations indicated that the LIM domains have no effect on the DNA binding specificity of the homeodomain.

The Two Classes of DNA Consensus Sequences Interact with the Same DNA Binding Domain. To test whether the two classes of target DNA sequences interact with the same region of the homeodomain, competition experiments were carried out. Using the two flanked PCR primers, several target DNA sequences were amplified from selected clones. In a preliminary experiment, the H5.3 PCR fragment (68 bp) was labeled and mixed with the GST-HD Sepharose beads in the presence of increasing amounts of unlabeled H5.3 PCR fragments. Under our experimental conditions, the competition could be carried out effectively with between 160 and 320 ng of unlabeled H5.3 fragment (Figure 6A). To test whether the binding of H5.3 fragment can be competed by other purified target DNAs, unlabeled PCR fragments (68 bp) of H5.1, H5.3, F8.3, F8.8, and F8.32 were included for binding competition. As shown in Figure 6B, effective competition was observed for both Class I binding sites (H5.3, F8.8) and Class II binding sites (H5.1 and F8.3) irrespective of their selection origins by GST-HD or by GST-ISL2. It is interesting to note that both pools of binding sites selected by GST-HD and GST-ISL2 contain a few sequences which do not fall into either classes of consensus (unclassified). These sequences might represent impurities resulting from the selection process rather than authentic target sequences. Consistent with this, the unclassified F8.32 did not compete significantly against H5.3. Similarly, no significant competition was observed for an unrelated DNA fragment (~130 bp), OP-93, which was purified from a promoter region of a ocean pout antifreeze protein gene (Gong et al., 1991), indicating that the competition is specific. The fact that the Class II binding sites (H5.1 and F8.3) and the Class I binding site (H5.3 and F8.8) compete equally well suggests that these two classes interact with the same region of the homeodomain.

The LIM Domains Have No Effect on Binding Affinity. To test whether the LIM domain affects the DNA binding affinity, Scatchard analysis was performed to determine the equilibrium dissociation constants (K_D) for GST-HD and GST-ISL2 in DNA binding. The H5.3 PCR fragment was labeled and incubated with limiting amounts of GST-HD or GST-ISL2 in the presence of increasing amounts of unlabeled H5.3 DNA as performed for competition analysis. After quantitation, the ratio of bound DNA to free DNA was plotted against the amount of bound DNA (Figure 7). The

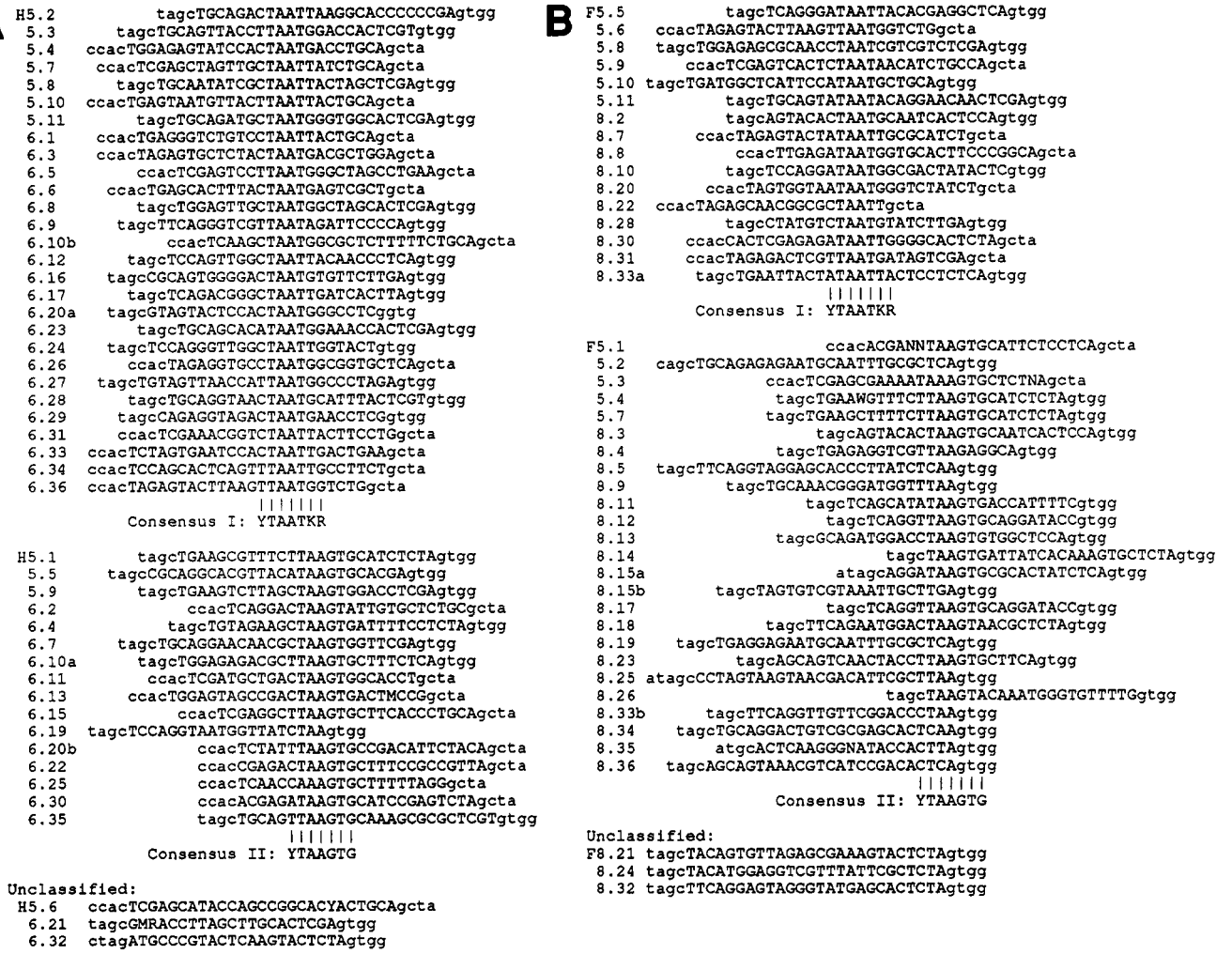

FIGURE 5: Sequence alignment of oligonucleotides isolated after several rounds of binding site selection by GST-HD (panel A) and GST-ISL2 (panel B). The sequences selected by GST-HD and GST-ISL2 are labeled by H and F, respectively. H5.1 through H5.11 were isolates from round 5 of the selection, and H6.1 through H6.36 were isolated from round 6. Similarly, F5.1 through F5.11 were isolated from round 5 and F8.1 through F8.36 from round 8. A few clones contain double inserts, whose sequences are denoted such as H6.10a and H6.10b. Capital letters denote the nucleotide portion corresponding to the region of the original 16 random bases and two flanking restriction sites. Lower case letters indicate ends of PCR primers, and only 4 nucleotides of each primer are shown. K = G, T; M = A, C; R = A, G; Y = C, T; W = A, T.

K_D in a bimolecular reaction can be obtained from the slope of this plot, which is equal to $-1/K_D$. As estimated from Figure 7, the K_D's are very similar for GST-HD (1.7×10^{-9} M) and GST-ISL2 (2.2×10^{-9} M). Similar results were also observed for H5.1 DNA, a Class II consensus (data not shown). The strong affinity of DNA binding can be achieved by the homeodomain alone, implying that no other protein domain is required. In contrast, the POU homeodomain protein pit-1 has a K_D of $\sim 1 \times 10^{-9}$ M to its target DNA while its homeodomain alone (without the POU-specific domain) has a K_D of $\sim 1 \times 10^{-6}$ M, a decrease of approximately 1000-fold in DNA binding (Ingraham et al., 1990).

DISCUSSION

The purpose of the present study is to examine the presumptive DNA binding role of the LIM domains of a newly characterized homeodomain protein Isl-2. We have expressed in bacteria several truncated salmon Isl-2 as GST fusion proteins and used these recombinant proteins to study their zinc and DNA binding properties. We found that the LIM domains were capable of binding zinc but had no role in DNA binding. Although zinc ions have been identified for several purified LIM proteins, this is the first report to demonstrate using zinc blots that the renatured LIM proteins have the ability to bind zinc. It is interesting to note that half of a LIM domain retains zinc-binding activity, thus supporting an earlier proposal that each LIM domain binds two zinc ions (Wang et al., 1992). The potential role of LIM domains in DNA binding was further examined by binding site selection. While the LIM domains alone failed to select any specific target DNA sequence, both the homeodomain and the full length Isl-2 selected the same two classes of DNA consensus, $^{C}/_{T}TAAT^{G}/_{T}{}^{G}/_{A}$ and $^{C}/_{T}TAAGTG$. Thus, in Isl-2, homeodomain is likely the sole DNA binding domain while the LIM domain neither recognizes a specific DNA target nor affects the DNA binding specificity. Further DNA binding analysis indicated that both the homeodomain and the full length Isl-2 have similar binding affinity to the selected target DNAs. Thus, the DNA binding affinity of the homeodomain was not affected by the LIM domains.

Our observation that the LIM domains of the salmon Isl-2 do not affect DNA binding by the homeodomain is in contrast to the previous observation by Sanchez-Garcia et

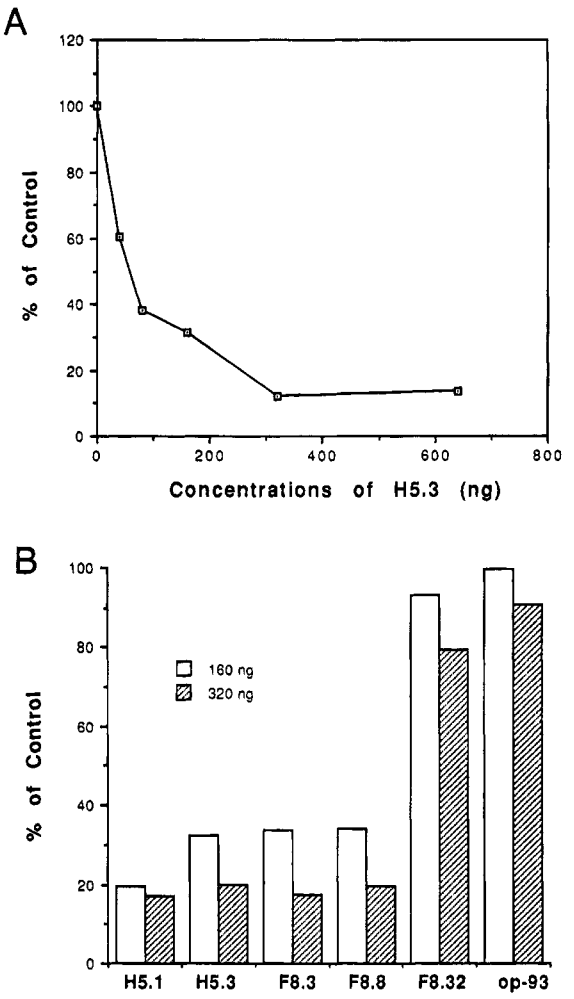

FIGURE 6: Competitive DNA binding assay. H5.3, a binding site selected by GST-HD, was radioactively labeled by PCR, and equal amounts (1.3×10^5cpm per reaction) of radioactive probes were incubated with GST-HD Sepharose beads in the presence of competitor DNA. (A) Percentage of binding activity of H5.3 to GST-HD in the presence of increasing amounts of unlabeled H5.3: 0, 40, 80, 160, 320, and 640 ng. The bound radioactivity in the absence of unlabeled H5.3 (0 ng) is arbitrarily set to 100%. (B) Competitive binding assay in the presence of unlabeled competitor DNA: H5.1, H5.3, F8.3, F8.8, F8.32, and an unrelated DNA fragment, OP-93. 160 and 320 ng of these competitor DNA were included, respectively. The values are percentage of bound radioactivity in the absence of any DNA competitor.

al. (1993) that the LIM domains of rat Isl-1 inhibit DNA binding. In our experiments, the full length GST-ISL2 was able to select specific DNA binding sites, while in their experiments the GST fusion protein containing the rat full length Isl-1 failed to select specific DNA targets because of the alleged inhibition from its LIM domains (Sanchez-Garcia et al., 1993). The discrepancy in the two sets of experiments is surprising since the two proteins have highly conserved primary structures (Figure 1A). Our observation of the lack of an inhibitory role of LIM domains is consistent with several *in vivo* transfection studies. For example, full length Isl-1 has been found to activate the somatostatin gene promoter (Vallejo et al., 1992), and the full length LIM homeodomain protein LH-2 was also shown to stimulate a target gene expression (Roberson et al., 1994). Furthermore, a synergistic stimulation of an insulin gene promoter activity by an LIM homeodomain protein *lmx-1* and a basic helix–loop–helix protein requires two complete LIM domains

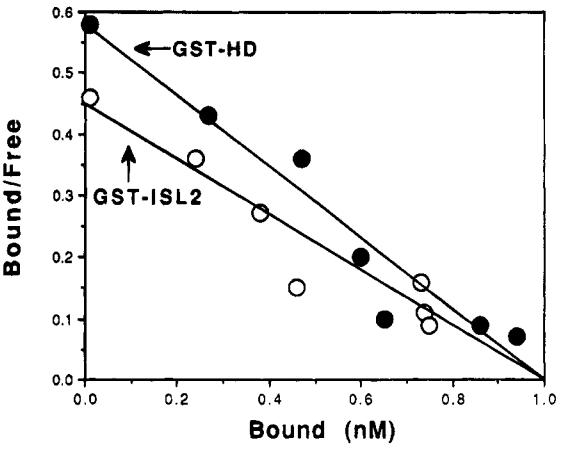

FIGURE 7: Measurements of binding constants of GST-HD and GST-ISL2. Binding affinities were determined under conditions of limiting protein concentrations with increasing concentrations of target DNA (H5.3). The ratio of bound DNA over free DNA was plotted against the amounts of bound DNA. Equilibrium dissociation constants (K_D) were determined from the slopes: 1.7×10^{-9} M for GST-HD (closed circles) and 2.2×10^{-9} M for GST-ISL2 (open circles).

(German et al., 1992). Therefore, the inhibitory role of LIM domains of rat Isl-1 in DNA binding may require further examination. Alternatively, the different behaviors of the LIM domains in the rat Isl-1 and salmon Isl-2 may be due to their intrinsic structures and can be clarified by site directed mutagenesis. If this is the case, the LIM domains may have different functions in different proteins. A functional difference may be localized in the highly divergent region between the second LIM domain and the homeodomain or in the relatively divergent second LIM domain (Figure 1A).

The binding consensus defined in the present work is consistent with the identified binding sites for the rat Isl-1 (Karlsson et al., 1990; Ohlsson et al., 1991; Sanchez-Garcia et al., 1993). This may not be surprising since the homeodomains of the two proteins differ only at a single residue (Figure 1A). Most homeodomain proteins bind to a common TAAT core sequence, and the two bases immediately following the core sequence are important for binding specificity (Treisman et al., 1992). The Class I consensus selected by GST-HD and GST-ISL2 is consistent with this notion. The first two bases following the TAAT core sequence almost invariably lack cytosine, with only two exceptions for the second base. This is consistent with an early report that the binding activity to the rat Isl-1 was abolished when these two bases were mutated from guanine to cytosine (Ohlsson et al., 1991). The Class II consensus, $^{C}/_{T}$TAAGTG, does not contain a TAAT core, but is highly close to the Class I consensus. Competition experiments indicate that both classes of binding sites interact with the same region of the homeodomain. Therefore, it is likely that the first G of the second class represents a variant of $^{C}/_{T}$TAAT$^{G}/_{T}{}^{G}/_{A}$ of the first class. When G is present instead of T, the requirement for following two bases would be higher; thus, the selected sequence is $^{C}/_{T}$TAAGTG rather than $^{C}/_{T}$TAAT$^{G}/_{T}{}^{G}/_{A}$. In fact, the two classes of consensus can be combined into a unified consensus as $^{C}/_{T}$TAA$^{T}/_{G}{}^{G}/_{T}{}^{G}/_{A}$, with only a partial change at one base compared to the Class I consensus.

The finding that the LIM domain neither binds to specific DNA nor affects the binding specificity and affinity is

surprising. Thus, the LIM homeodomain proteins behave neither like the paired homeodomain proteins nor like the POU homeodomain proteins, both of which contain a second conserved domain involved in DNA binding. The paired domain recognizes a DNA target independent of the homeodomain (Treisman et al., 1989) while the POU-specific domain increases the binding specificity and affinity of the homeodomain for the target sequence (Ingraham et al., 1990). A recent report indicates that the POU-specific domain of Oct-1 can select and bind its own target DNA independently but with a weak affinity (Verrijzer et al., 1992). The widespread assumption that a DNA binding role exists for the LIM domain is based on the similarity of the protein sequences of the LIM domain and a zinc finger, the well characterized DNA binding motif. Indeed, zinc has been found in many LIM proteins (Li et al., 1991; Michelsen et al., 1993; Archer et al., 1994; Kosa et al., 1994), including salmon Isl-2 from the present study. However, zinc binding per se does not necessarily ensure a role in DNA binding. There are numerous metalloproteins, such as alcohol dehydrogenase and metallothionein which are capable of binding zinc but not DNA, despite the fact that they contain similar cysteine-rich regions (Vallee & Auld, 1990). In some transcription factors with multiple zinc fingers, only certain fingers contribute to specific DNA binding (Keller & Maniatis, 1992). Therefore, although the LIM domain resembles zinc fingers and binds zinc, it is not necessarily a DNA binding domain. Further structural studies by nuclear magnetic resonance or X-ray crystallography will be required to provide more insight into the structure and function of the LIM domain.

The presence of conserved LIM domain in distinct proteins from diverse species strongly suggests that it has an important role. So far, two major groups of LIM domain proteins have been identified. The first group of LIM proteins also contains a homeodomain, which is characteristic of a family of well documented transcription factors. Thus, these LIM proteins are undoubtedly involved in the regulation of gene transcription. The other group of LIM proteins consist almost exclusively of the LIM domain(s) such as the rat cysteine-rich intestine protein CRIP, the human cysteine-rich protein CRP, and rhombotin-1 and -2. CRIP is an intracellular zinc carrier involving in transmucosal zinc transport (Hempe & Cousins, 1991); rhombotin-1 and -2 are T-cell oncogenes resulting in leukemia with chromosomal translocation (Boehm et al., 1991; McGuire et al., 1989); the actual function of CRP is not clearly defined, but it was reported to be induced with c-*myc* as the primary response to serum in quiescent cells (Wang et al., 1992). These three LIM proteins in the second group, including the human CRP and rhombotin-1 and -2, may also be involved in transcriptional regulation either by protein−DNA interaction or by protein−protein interaction. Our investigation argues against protein−DNA interaction for the LIM domain. Therefore, protein−protein interaction via the LIM domain is more likely and is supported by several recent reports. For example, the insulin gene can be synergically activated by an interaction between the LIM homeodomain protein *lmx-1* and a basic helix−loop−helix protein. This interaction requires complete LIM domains of *lmx-1*, and these LIM domains cannot be replaced by the LIM domains of Isl-1 (German et al., 1992). Two chicken LIM domain proteins CRP and zyxin form heterodimers associated with cytoskeleton (Sadler et al., 1993). Finally, the LIM domain also resembles some iron−sulfur structures. Li et al. (1991) have determined that the LIM domain purified from inclusion bodies contained iron and suggested that LIM domain may be receptive to changes in intracellular redox potential providing a mechanism to modulate the transcriptional activity in response to a variety of redox signals (Li et al., 1991). However, a recent study indicated that the LIM domain expressed in the soluble fraction contained only zinc (Archer et al., 1994), thus the redox role of LIM domains remains to be established.

ACKNOWLEDGMENT

We thank Dr. K. V. Ewart for comments and proofreading of the manuscript.

REFERENCES

Archer, V. E. V., Breton, J., Sanchez-Gorcia. I., Osada, H., Forster, A., Thomson, A. J., and Rabbitts, T. H. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 316−320.

Baltz, R., Domon, C., Pillay, D. T. N., and Steinmetz, A. (1992) *Plant J. 2*, 713−721.

Berg, J. M. (1990) *Annu. Rev. Biophys. Biophys. Chem. 19*, 405−421.

Birkenmeier, E. H., and Gordon, J. I. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 2516−2520.

Boehm, T., Foroni, L., Kaneko, Y., Perutz, M. F., and Rabbitts, T. H. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 4367−4371.

Bopp, D., Burri, M., Baumgartner, S., Frigerio, G., and Noll, M. (1986) *Cell 47*, 1033−1040.

Bourgouin, C., Lundgren, S. E., and Thomas, J. B. (1992) *Neuron 9*, 549−561.

Chittenden, T., Livingston, D. M., and Kaelin, W. G., Jr. (1991) *Cell 65*, 1073−1082.

Cohen, B., McGuffin, M. E., Pfeifle, C., Segal, D., and Cohen., S. M. (1992) *Genes Dev. 6*, 715−729.

Freyd, G., Kim, S. K., and Horvitz, H. R. (1990) *Nature 344*, 876−879.

Gehring, W. J., Affolter, M., and Burglin, T. (1994) *Annu. Rev. Biochem. 63*, 487−526.

German, M. S., Wang, J., Chadwick, R. B., and Rutter, W. J. (1994) *Genes Dev. 6*, 2165−2176.

Gong, Z., Hew, C. L., and Vielkind, J. R. (1991) *Mol. Marine Biol. Biotech. 1*, 64−72.

Gong, Z., Hu, Z., Gong, Z. Q., Kitching, R., and Hew, C. L. (1994a) *Mol. Marine Biol. Biotech.* (in press).

Gong, Z., Hui, C.-c., and Hew, C. L. (1994b) *J. Biol. Chem.* (submitted).

Hempe, J. M., and Cousins, R. J. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 9671−9674.

Herr, W., Sturm, R. A., Clerc, R. G., Corcoran, L. M., Baltimore, D., Ingraham, H. A., Rosefeld, M. G., Finney, M., Ruvkun, G., and Horvitz, H. R. (1988) *Genes Dev. 2*, 1513−1516.

Ingraham, H. A., Flynn, S. E., Voss, J. W., Albert, V. R., Kapiloff, M. S., Wilson, L., and Rosenfeld, M. G. (1990) *Cell 61*, 1021−1033.

Karlsson, O., Thor, S., Norberg, T., Ohlsson, H., and Edlund, T. (1990) *Nature 344*, 879−882.

Keller, A. D., and Maniatis, T. (1992) *Mol. Cell. Biol. 12*, 1940−1949.

Kosa, J. L., Michelsen, J. W., Louis, H. A., Olsen, J. I., Davis, D. R., Beckerle, M. C., and Winge, D. R. (1994) *Biochemistry 33*, 468−477.

Laemmli, U. K. (1970) *Nature 227*, 680−685.

Li, P. M., Reichert, J., Freyd, G., Horvitz, H. R., and Walsh, C. T. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 9210−9213.

McGuire, E. A., Hockett, R. D., Pollock, K. M., Bartholodi, M. F., O'Brien, S. O., and Korsmeyer, S. J. (1989) *Mol. Cell. Biol. 9*, 2124−2132.

Michelsen, J. W., Schmeichel, K. L., Beckerle, M. C., and Winge, D. R. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 4404−4408.

Mizuno, K., Okano, I., Ohashi, K., Nunoue, K., Kuma, K.-i., Miyata, T., and Nakamura, T. (1994) *Oncogene 9*, 1605−1612.

Ohlsson, H., Thor, S., and Edlund, T. (1991) *Mol. Endocrinol. 5*, 897−904.

Roberson, M. S., Schoderbek, W. E., Tremml, G., and Maurer, R. A. (1994) *Mol. Cell. Biol. 14*, 2985−2993.

Sadler, I., Crawford, A. W., Michelsen, J. W., and Beckerle, M. C. (1992) *J. Cell Biol. 119*, 1573−1587.

Sanchez-Garcia, I., Osada, H., Forster, A., and Rabbitts, T. H. (1993) *EMBO J. 12*, 4243−4250.

Schiff, L. A., Nibert, M. L., and Field, B. N. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4195−4199.

Smith, D. B., and Johnson, K. S. (1988) *Gene 67*, 31−40

Taira, M., Jamrich, M,Good, P. J., and Dawid, I. B. (1992) *Genes Dev. 6*, 356−366.

Taira, M., Hayes, W. P., Otani, H., and Dawid, I. B. (1993) *Dev. Biol. 159*, 245−256.

Thiesen, H.-J., and Bach, C. (1990) *Nucleic Acids Res. 18*, 3203−3209.

Treisman, J., Gonczy, P., Vashishtha, M., Harris, E., and Desplan, C. (1989) *Cell 59*, 553−24.

Treisman, J., Harris, E., Wilson, D., and Desplan, C.. (1992) *BioEssays 14*, 145−150.

Vallee, B. L., and Auld, D. S. (1990) *Biochemistry 29*, 5647−5659.

Vallejo, M., Penchuk, L., and Habener, J. F. (1992) *J. Biol. Chem. 267*, 12876−12884.

Verrijzer, C. P., Alkema, M. J., van Weperen, W. W., Van Leeuwen, H. C., Strating, M. J. J., and van der Vliet, P. C. (1992) *EMBO J. 11*, 4993−5003.

Wang, X., Lee, G., and Liebhaber, S. A. (1992) *J. Biol. Chem. 267*, 9176−9184.

Xu, Y., Baldassare, M., Fisher, P., Rathbun, G., Oltz, E. M., Yancopoulos, G. D., Jessell, T. M., and Alt, F. W. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 227−231.