

Molecular cloning and characterization of interferon α/β response element binding factors of the murine (2'-5')oligoadenylate synthetase ME-12 gene

[interferon signal transduction pathways/(2'-5')oligoadenylate synthetase gene regulation/transcription factors/amphipathic protein]

CONG YAN AND IGOR TAMM*

The Rockefeller University, 1230 York Avenue, New York, NY 10021

Contributed by Igor Tamm, September 12, 1990

ABSTRACT Seven clones encoding interferon response element binding factors have been isolated from a mouse fibroblast λ gt11 cDNA library by using a ^{32}P end-labeled tandem trimer of the mouse (2'-5')oligoadenylate synthetase gene interferon response element as a probe. Clone 16 shares strong similarity (95%) at both DNA and amino acid level with YB-1, a human major histocompatibility complex class II Y-box DNA-binding protein, and with dbpB, a human epidermal growth factor receptor gene enhancer region binding protein. The product of the gene represented by clone 16 may represent a factor that regulates multiple genes by binding to a variety of 5' regulatory elements. Clone 25 is a 2407-base-pair-long cDNA and contains a putative 311-amino acid open reading frame corresponding to an estimated mass of 35.5 kDa. This putative protein, designated as interferon response element binding factor 1 (IREBF-1), contains an acidic domain, three heptad repeat leucine arrays, and a region that shares similarity with the yeast transcriptional factor GAL4 DNA-binding domain. Furthermore, the C terminus of IREBF-1 shows an unusual amphipathic property: within a 79-amino acid range, one side of the α -helical region contains a preponderance of hydrophobic amino acids and the other side contains hydrophilic amino acids. This type of structure provides a strong hydrophobic force for protein-protein interaction.

Interferon (IFN) signal transduction pathways and IFN-inducible gene activation are complex processes and involve many biochemical events (for review, see refs. 1 and 2). We have previously shown that the well-defined signal transduction pathways in which diacylglycerol, cAMP, and Ca^{2+} serve as second messengers are not capable, alone or in combination, of activating (2'-5')oligoadenylate synthetase (2',5'-OAS) gene expression in murine BALB/c 3T3 cells (3). By a gel electrophoresis mobility assay using the mouse 2',5'-OAS gene's AB regulatory region [which contains an IFN response element (IRE) designated element B and a constitutive element A; see ref. 4] as a radiolabeled probe, we detected the formation of a set of specific protein-DNA complexes with nuclear and cytoplasmic extracts of BALB/c 3T3 cells (5, 6).

Among the DNA-binding factors we have detected, of special interest is the complex 4-forming protein factor that is IRE-specific and preexists in the cytoplasm (5, 6). This IFN-stimulated response factor (ISRF) appears to be complexed to an inhibitor (ISRFI) in the cytoplasm, dissociate from the inhibitor, and translocate into the nucleus upon treatment of cells with IFN (5, 6). A similar phenomenon has been observed in some other cell lines with other IFN-inducible genes (ref. 7 and references therein). ISRF appears to be a critical factor in the stimulation of IFN-inducible

2',5'-OAS gene expression by IFN. Molecular cloning of this factor should play a key role in studies of IFN signal transduction pathways and the activation of the IFN-inducible 2',5'-OAS gene.

We note that, in the case of the 2',5'-OAS gene as well as in other IFN-inducible gene systems, more than one protein complex is detected with IRE as the probe. However, only one protein factor represented by complex 4 appears to respond to IFN treatment in the 2',5'-OAS gene system (5), whereas in the IFN-stimulated gene 54 and gene 15 (ISG54 and ISG15) systems, two protein factors (ISGF2 and ISGF3) respond. ISGF2 is a slow IFN response factor and its appearance in the nucleus depends on protein synthesis (8). ISGF2, also known as IFN response factor 1 (IRF-1), has been cloned recently by several groups (9–11). Other protein factors of the IRF-1 family have been cloned as IRF-2 and the IFN consensus sequence binding protein (ICSBP) (12, 13). In contrast to ISGF2, ISGF3 is a rapid IFN response factor (7). Like ISRF (5, 6), it preexists in the cytoplasm and rapidly translocates into the nucleus upon IFN treatment. It is a primary factor that mediates activation of ISG54 and ISG15 expression by IFN (7).

Here we report that, by using the synthetic tandem trimer of the mouse 2',5'-OAS gene IRE (element B) as a probe, seven positive clones have been isolated by screening a mouse λ gt11 cDNA expression library. Clone 25 has been fully sequenced and shows no homology to any of the above-mentioned clones and therefore represents a member of a different gene family.[†] The product of this clone contains a region of shared similarity with GAL4 DNA-binding domain—an acidic domain and an amphipathic region that contains three heptad repeat leucine arrays.

MATERIALS AND METHODS

Library Screening. A mouse fibroblast λ gt11 cDNA expression library (3T3 Swiss albino strain, from Clontech) was screened by using the synthetic tandem trimer of the mouse ME-12 gene IRE (element B; ref. 4) by the method of Vinson *et al.* (14). The DNA probe was end labeled with [γ - ^{32}P]ATP by T4 kinase. The exact oligonucleotide sequence (trimer of element B with agct tail) of the probe is as follows:

```
5'-tcgaCGGGAAATGGAAACTGCGGGAAATG-  
3'-GCCCTTTACCTTTGACGCCCTTTAC-  
GAAACTGCGGGAAATGGAAACTG-3'  
CTTTGACGCCCTTTACCTTTGACagct-5'.
```

Abbreviations: IFN, interferon α/β ; 2',5'-OAS, (2'-5')oligoadenylate synthetase; IRE, IFN response element (also known as element B); IREBF, IRE binding factor; USE, upstream stimulating element; ISRF, IFN-stimulated response factor.

*To whom reprint requests should be addressed.

[†]The sequence reported in this paper has been deposited in the GenBank data base (accession no. M55290).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

The upstream stimulating element (USE) DNA fragment (39 base pairs) was a kind gift from Hong Du and Robert G. Roeder (The Rockefeller University, New York). The DNA-binding buffer contained 10 mM Tris-HCl (pH 7.5), 50 mM NaCl, and 10 mM MgCl₂.

Analysis of the DNA-Binding Protein Produced in *Escherichia coli*. Two methods have been used to prepare protein extracts for analyzing the product of clone 25 cDNA. The first method is to generate λ phage lysogen as described by Huynh *et al.* (15). We also used a second method to prepare protein extracts. Briefly, the DNA insert from λ gt11 phage was released by the restriction enzyme *EcoRI* and ligated into the *EcoRI* site of the polylinker region in the pBluescript SK plasmid vector (Stratagene). After transformation, the *E. coli* bacteria (XL1-Blue), bearing the inserted pBluescript plasmid, were grown in 10 ml of LB medium containing 500 μ g of ampicillin and 120 μ g of tetracycline overnight at 37°C. The sedimented cells were resuspended in 1 ml of solution containing 10 mM Tris-HCl (pH 7.4) and 5% glycerol. The cell suspensions were frozen-thawed three times and sonicated for 2 min. This solution was ready for the electrophoresis mobility assay.

Electrophoresis Mobility Assay. Protein-DNA binding reactions were carried out in a final volume of 20 μ l of a solution containing various amounts of protein, 1 μ g of poly(dI-dC) and 4×10^4 cpm of DNA probe in the presence of 10 mM Hepes, 50 mM NaCl, 1 mM MgCl₂, 0.1 mM EDTA, and 5% glycerol. Reaction mixtures were incubated at 37°C for 30 min and loaded immediately onto a 6% nondenaturing polyacrylamide gel. Gels were run 3–4 hr in TBE buffer (89 mM Tris borate, 89 mM boric acid, and 2 mM EDTA) and dried, and autoradiograms were prepared on x-ray film.

DNA Sequencing. The inserted DNA sequences of the pBluescript SK plasmid were determined by the dideoxy chain-termination method using Sequenase (United States Biochemical) and synthetic primers.

RESULTS

Mouse λ gt11 cDNA Library Screening. For primary screening, 5×10^5 plaque-forming units of phages of a mouse fibroblast cDNA library (Clontech) were screened using a ³²P end-labeled tandem trimer of the mouse 2',5'-OAS gene IRE. From this primary screening, 50 positive clones were isolated for further characterization. Seven clones were confirmed to be positive through subsequent screening. After the fourth round of purification, all 7 clones showed 100% positivity. For each purified clone, one well-isolated plaque was subjected to amplification. The amplified phage stock solutions were titered (all were $\approx 10^{10}$ plaque-forming units per ml) and stored at -70°C. All 7 clones showed IRE trimer specificity, as they did not bind to the USE fragment of the adenovirus major later promoter (Fig. 1). All DNA inserts were released by the *EcoRI* restriction enzyme, except for clone 12 (we suspect that it does not contain an *EcoRI* site, at least not at one end of inserted DNA). Clone 15 consists of two bands, 0.7 kilobases (kb) and 0.9 kb; clone 16 and clone 45 each show one 0.6-kb band; clone 25 shows one 2.4-kb band; clone 31 consists of bands of 2.4 kb and 0.7 kb; clone 38 shows one 1.2-kb band. Clones 25 and 16 were chosen for further studies.

Gel Electrophoresis Mobility Assay and Competition Study of the Clone 25 Gene Product. As described in *Materials and Methods*, we used two methods to prepare protein extracts for characterizing the clone 25 gene product. First, we generated λ phage lysogens as described by Huynh *et al.* (15) for a gel electrophoresis mobility assay. As shown in Fig. 2A, protein extract from bacteria that harbor recombinant phage λ 25 formed a complex with radiolabeled trimer of the mouse 2',5'-OAS gene IRE DNA fragment. This complex is specif-

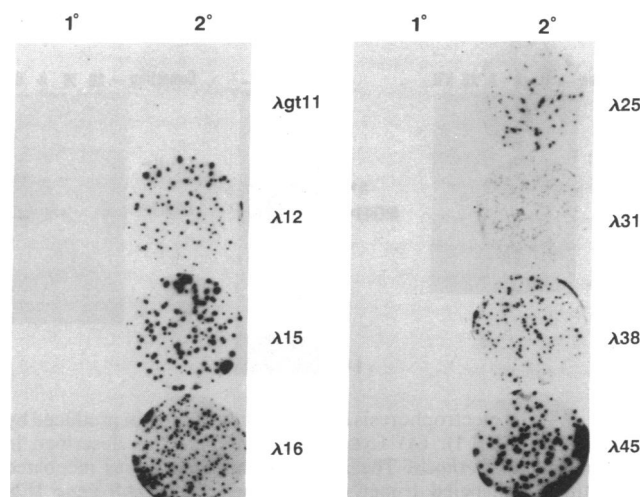


FIG. 1. Molecular cloning of IREBFs from a mouse fibroblast λ gt11 cDNA library. After four rounds of purification of seven positive clones, ≈ 100 phages were plated on a 10-cm dish (containing LB medium and ampicillin) for each clone (clones 12, 15, 16, 25, 31, 38, and 45), and the parent λ gt11 phage was plated as a control. Two nitrocellulose membrane filters were lifted from each original dish. One filter was hybridized with a ³²P end-labeled trimer of mouse 2',5'-OAS gene element B, which is the IRE, and the other was hybridized with ³²P end-labeled USE. After blotting and washing with hybridization buffer, the two filters were subjected to autoradiography. The left lanes (1°) represent filters hybridized with USE; the right lanes (2°) represent filters hybridized with the trimer IRE of the 2',5'-OAS gene.

ically displaced by unlabeled trimer IRE and the synthetic 5' regulatory region AB of the 2',5'-OAS gene but not by either regulatory element A or B used alone and not by the USE fragment of the adenovirus major late promoter. This agrees with our previous finding that only regulatory region AB forms stable complexes with its specific binding factors (5).

Second, we inserted DNA insertion fragment from clone 25 into pBluescript SK plasmid at its *EcoRI* site of the polylinker region in two orientations. The constructed recombinant plasmids were transfected into XL1-Blue bacteria to allow expression of inserted DNA gene product. Fig. 2B shows that by using the synthetic regulatory region AB as a radiolabeled probe, protein extract from bacteria that harbor the parent pBluescript plasmid generates a nonspecific band. Protein extract from bacteria that harbor plasmid-containing insertion DNA in the wrong orientation (λ 25R) generates the same nonspecific band. In contrast, protein extract from bacteria that harbor the plasmid containing insertion DNA in the correct orientation (λ 25) generates a new band, with the nonspecific band being either undetectable or weak. Fig. 2C shows that this newly formed complex is specifically displaced by unlabeled regulatory region AB fragment or trimer IRE. Element A competes slightly, whereas element B does not. The nonspecific band could not be displaced regardless of whether the trimer IRE or the AB, A, or B fragment was used as the competitor.

These two lines of evidence show that the insertion DNA from clone 25 expresses a protein product that specifically binds to IRE used either as a trimer or linked to element A.

Structure of the cDNA Encoding IFN Response Element Binding Factor 1 (IREBF-1). As shown in Fig. 3, the cDNA insert of clone 25 is 2407 base pairs long and contains a long open reading frame encoding a putative protein of 311 amino acids (nucleotides 542–1477). A GenBank homology search (release 65.0) revealed no identity with any previously reported cDNA sequence. The first AUG codon in the open reading frame is located at nucleotide position 542, but the

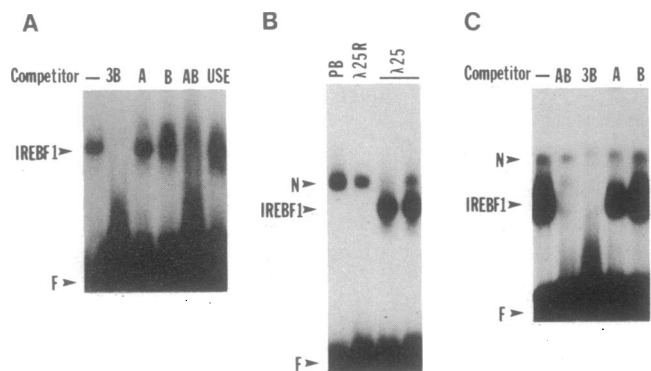


FIG. 2. Gel electrophoresis mobility study of protein produced by clone 25 (IREBF-1). (A) Lysogen was prepared as described in *Materials and Methods*. The protein from lysogen was incubated with the radiolabeled trimer of the mouse 2',5'-OAS gene IFN response element B (IRE) and subjected to 6% polyacrylamide gel electrophoresis in TBE buffer. Oligonucleotide competitor was added to incubation reaction mixtures in a 100-fold excess as follows (from left to right): no competitor, trimer of mouse 2',5'-OAS gene IFN response element B, mouse 2',5'-OAS gene constitutive element A, mouse 2',5'-OAS gene IFN response element B (IRE), mouse 2',5'-OAS gene regulatory region AB, and the USE of the adenovirus major later promoter. F, free DNA. (B) Insertion DNA from phage clone 25 was released by *EcoRI* digestion and inserted into the pBluescript SK plasmid *EcoRI* site in both orientations. The reconstructed plasmids were transfected into the XL1-Blue bacteria strain. Protein extracts from the bacteria-harboring plasmid were incubated with radiolabeled mouse 2',5'-OAS gene regulatory region AB and subjected to 6% polyacrylamide gel electrophoresis in TBE buffer. Lanes from left to right represent protein from bacteria containing pBluescript plasmid (PB), insertion DNA from clone 25 in the pBluescript plasmid in the wrong orientation, insertion DNA from clone 25 in the pBluescript plasmid in the correct orientation (duplicate lanes). F, free DNA; N, nonspecific band. (C) Protein from bacteria containing insertion DNA from clone 25 in pBluescript in the correct orientation was incubated with radiolabeled mouse 2',5'-OAS gene regulatory region AB and subjected to 6% polyacrylamide gel electrophoresis in TBE buffer. Competitor was added to incubation reaction in a 100-fold excess as follows (from left to right): no competitor, mouse 2',5'-OAS gene regulatory region AB, trimer of the mouse 2',5'-OAS gene enhancer element B, mouse 2',5'-OAS gene constitutive element A, mouse 2',5'-OAS gene IFN response element B. F, free DNA; N, nonspecific band.

nucleotide sequence (GCCACAATGA) around the AUG codon at position 629 matches well with the Kozak consensus sequence (GCCA/GCCATGG) for the translation initiation site (16). A poly(A) signal (AATAAA) is located at nucleotide position 2224.

In the 5' untranslated DNA region, there are several small upstream open reading frames and stop codons. The same phenomenon has been observed in many genes, including a yeast transcriptional activator, GCN4. Experimental data have indicated that these upstream open reading frames and stop codons are important in the regulation of translational initiation of GCN4 (refs. 17 and references therein).

Structure Analysis of IREBF-1. The putative protein encoded by clone 25 is 311 amino acids long with an estimated mass of 35.5 kDa. We shall designate it the IFN response element binding factor 1 (IREBF-1). Fig. 3 shows that it contains three potential N-linked glycosylation sites (Asn-Xaa-Tyr/Ser/Thr) and two potential casein kinase II phosphorylation sites (Ser/Thr-Xaa-Xaa-Asp/Glu). The hydrophathy plot of this protein (data not shown) indicates that there is a hydrophilic region in the middle of the protein (amino acids 110–200). The predicted secondary structure of the protein reveals three distinct regions by Robson's method (18). (i) Amino acids 1–100 have the potential to form an α -helical structure. (ii) Amino acids 101–220 have the poten-

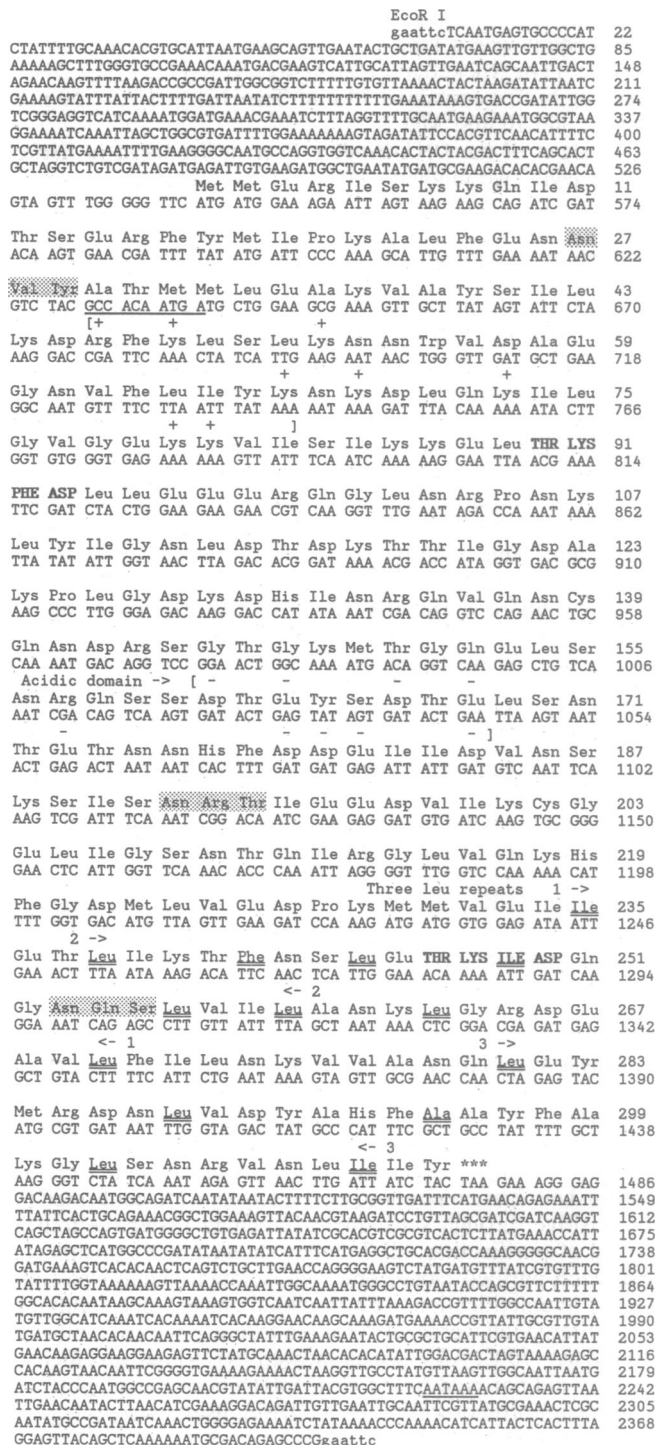


FIG. 3. Nucleotide and deduced amino acid sequence of clone 25 (IREBF-1). The Kozak consensus sequence and the poly(A) signal sequence are underlined. The potential N-linked glycosylation sites are stippled. The potential casein kinase II (CK II) phosphorylation sites are in boldface capital letters. The three arrays of heptad repeat leucines are defined by numbers and arrows above the amino acid sequence, and the leucine residues in the repeats are doubly underlined. The GAL4 DNA-binding domain similarity region and the acidic domain are defined by square brackets above the amino acid sequence. The "+" and "-" refer to positively and negatively charged amino acid side chains, respectively.

tial to form turn/coil structures, and, as already indicated, this region is hydrophilic. (iii) Amino acids 220–311 have the potential to form an α -helical structure with a turn/coil break at amino acid positions 243–260.

46	RFKLSLKNWVDAEAGNVF-LIYKNKDLQKILGVGEKKVI	83	λ 25
		
	RAHLTEVESRLERLEQLFLLIFPREDLDMILKMSLQDI	89	GAL4
51		89	

FIG. 4. Similarity between the GAL4 DNA-binding domain and amino acids 46–83 of IREBF-1. Identical residues are indicated by two dots, and conservative substitutions are indicated by a single dot.

The region of amino acids 46–83 of the protein shares a similarity with the DNA-binding domain of the yeast transcriptional activation factor GAL4 (26% identity in a 38-amino acid overlap) revealed by a search through the NEWAT library (Fig. 4). In this region, the percentage of basic amino acids is 21% (compared to an overall percentage of basic amino acids in the entire protein of 13.5%). More interestingly, secondary structure analysis shows that all the basic amino acids are heavily concentrated in certain topographic locations of the α -helix (positions 1, 3, and 7 in Fig. 5A), where the percentage of basic amino acids is 47% (8 out of 17). If this is the DNA-binding site of IREBF-1, these highly concentrated basic amino acids can be expected to play an important role in interacting with the negatively charged DNA backbone.

The study of the hydrophilic region reveals an acidic domain (amino acids 161–184; Fig. 3). In this region of 24 amino acids, there are 9 acidic amino acids (37.5%) with only 1 basic amino acid at position 177 (histidine). The acidic domain has been studied extensively in some transcriptional activators, such as GCN4, GAL4, and VP16 (for review, see refs. 19 and 20). It is a transcriptional activation domain in these factors, because deletion or mutation of these acidic amino acids greatly decreases the gene activation activities of these factors.

There is a region at the C terminus that contains three heptad repeat leucines (i.e., there is one leucine in every seven amino acids) in amino acids 235–270, 238–259, and 281–310 (Figs. 3 and 5B). McKnight and colleagues (21) have described a leucine zipper motif containing heptad repeat leucines as a structure common to a class of DNA-binding proteins. Its important function is to mediate dimerization with other protein factors, with the formation of either homo- or heterodimers. Many important transcriptional activators, such as GCN4, c-fos, c-jun, and USF, contain leucine zipper structures (for review, see ref. 20). It remains to be deter-

mined whether the heptad repeat leucines in IREBF-1 represent leucine zippers and are involved in protein dimerization. More strikingly, Fig. 5B shows that helical wheel (3.5 residues per turn) analysis of amino acids 233–311 of IREBF-1 identifies a remarkable amphipathic feature in this long α -helical stretch. One side of the helix (positions 3, 6, and 7 of the wheel in Fig. 5B) is hydrophobic and contains three arrays of heptad repeat leucines. The other side of the helix (positions 1, 2, 4, and 5) is hydrophilic and contains charged amino acids and hydroxyl groups (Ser, Thr, and Tyr). This structure provides a very strong hydrophobic force favoring formation of protein–protein complexes while exposing hydroxyl groups to the environment for modification, such as phosphorylation or glycosylation. Such modifications could play an important role in modulating protein structure, which could affect protein–protein interaction. There are at least five or six salt bridges in this region to stabilize the α -helical structure as originally described by McKnight (21).

DNA Sequence Analysis of Clone 16. DNA sequencing and an homology search with GenBank (release no. 65.0) revealed that clone 16 is a partial gene sequence (648 base pairs and 216 amino acids) and shares >95% homology at both DNA and amino acid levels with human YB-1, a human major histocompatibility complex class II Y-box DNA-binding protein (22), and human dbpB, a human epidermal growth factor receptor gene enhancer region binding protein (23). YB-1 is thought to be a negative regulatory factor (22). The high degree of similarity between the mouse gene (clone 16) and human genes (YB-1 and dbpB), at both amino acid and DNA levels, suggests that this gene is highly conserved in different species.

DISCUSSION

IRE (also referred to as element B) is a key DNA segment that mediates activation of the IFN-inducible 2',5'-OAS gene upon IFN treatment of cells (ref. 4 and references therein). In the consensus sequence shared by IREs in different IFN-inducible genes, including both the murine and human 2',5'-OAS gene, the GAAAC element appears to be the most conserved (4). We (5, 6) and other investigators (4) have demonstrated that this element can form complexes with multiple protein factors in a variety of mammalian cell lines.

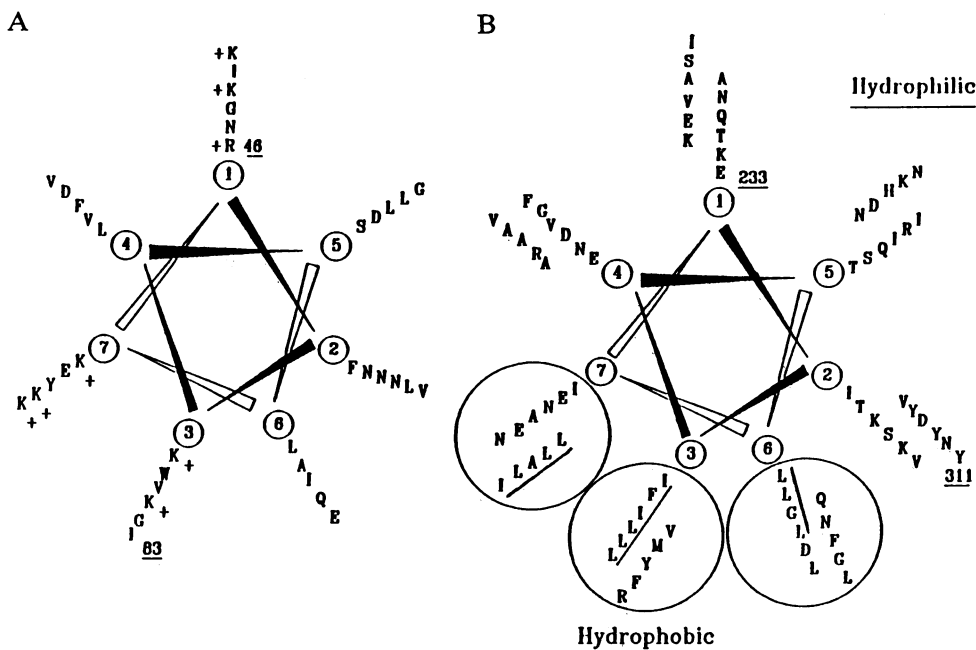


FIG. 5. Helical wheel analysis of IREBF-1. Amino acids are indicated by single letters. Underlined numbers represent the starting and ending amino acids. The circled numbers represent amino acid positions in the α -helix. (A) Positively charged amino acid locations in the 46–83 segment containing 38 amino acids. The “+” represents a positive charge on the amino acid side chains. (B) Amphipathic property of amino acids in the 233–311 segment containing 79 amino acids. The leucine heptad repeats are indicated by a solid line. The hydrophobic force is indicated by circles.

