Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 7958–7962, October 1990
Biochemistry

Cloning of the cDNA encoding an RNA regulatory protein—the human iron-responsive element-binding protein

(posttranscriptional gene regulation/iron)

TRACEY A. ROUAULT*, CAREEN K. TANG*, STAMATINA KAPTAIN*, WILSON H. BURGESS†, DAVID J. HAILE*, FELIPE SAMANIEGO*, O. WESLEY MCBRIDE‡, JOE B. HARFORD*, AND RICHARD D. KLAUSNER*

*Cell Biology and Metabolism Branch, National Institute of Child Health, Bethesda, MD 20892; †Laboratory of Molecular Biology, Jerome H. Holland Laboratory, American Red Cross, Rockville, MD 20855; and ‡Laboratory of Biochemistry, Division of Cancer Biology and Diagnosis, National Cancer Institute, Bethesda, MD 20892

Communicated by G. D. Aurbach, July 23, 1990 (received for review May 20, 1990)

ABSTRACT **Iron-responsive elements (IREs) are stem-loop structures found in the mRNAs encoding ferritin and the transferrin receptor. These elements participate in the iron-induced regulation of the translation of ferritin and the stability of the transferrin receptor mRNA. Regulation in both instances is mediated by binding of a cytosolic protein to the IREs. High-affinity binding is seen when cells are starved of iron and results in repression of ferritin translation and inhibition of transferrin receptor mRNA degradation. The IRE-binding protein (IRE-BP) has been identified as an ≈90-kDa protein that has been purified by both affinity and conventional chromatography. In this report we use RNA affinity chromatography and two-dimensional gel electrophoresis to isolate the IRE-BP for protein sequencing. A degenerate oligonucleotide probe derived from a single peptide sequence was used to isolate a cDNA clone that encodes a protein containing 13 other sequenced peptides obtained from the IRE-BP. Consistent with previous characterization of the IRE-BP, the cDNA encodes a protein of 87 kDa with a slightly acidic pI, and the corresponding mRNA of ≈3.6 kilobases is found in a variety of cell types. The encoded protein contains a nucleotide-binding consensus sequence and regions of cysteine and histidine clusters. This mRNA is encoded by a single gene on human chromosome 9, a finding consistent with previous localization by functional mapping. The protein contains no previously defined consensus motifs for either RNA or DNA binding. The simultaneous cloning of a different, but highly homologous, cDNA suggests that the IRE-BP is a member of a distinct gene family.**

Studies on the regulation of cellular iron metabolism in higher eukaryotes have revealed an intriguing system of posttranscriptional gene control (1, 2). The mRNAs encoding two of the major proteins that participate in iron homeostasis contain regulatory structures that have been termed iron-responsive elements (IREs). These stem-loop structures, first identified in the 5′-untranslated regions (UTRs) of ferritin mRNAs, were subsequently identified in the 3′ UTR of the transferrin receptor (TfR) mRNA (3–5). In ferritin, a single IRE is responsible for iron-dependent control of translation, whereas multiple IREs in the TfR 3′ regulatory region mediate iron-dependent control of mRNA half-life (6, 7). IREs function as the binding site for a cytosolic protein as demonstrated in RNA mobility shift and UV crosslinking experiments (8, 9). It appears that both ferritin and TfR IREs bind to the same protein *in vitro* (10). Experiments conducted both *in vivo* and *in vitro* demonstrate that binding of the protein is responsible for repressing ferritin mRNA translation (9, 11). In TfR mRNA, the IREs are part of a more complex RNA instability element. Binding of the protein to the TfR 3′ UTR inhibits degradation of this otherwise rapidly degraded mRNA (6, 7). IREs and a specific cytosolic binding protein have been detected across a wide number of species of higher eukaryotes (12, 13). Several protocols have been reported for use in isolating proteins of similar activity and apparent molecular weight, and these isolated proteins have been named the ferritin repressor protein (11), the iron-regulatory factor (14), and the IRE-binding protein (IRE-BP) (9). We favor the latter name, as it encompasses a role in both ferritin and TfR regulation and more directly reflects the known function of the protein. Based upon the difference in mobility of human and rodent IRE–IRE-BP complexes, we have localized the gene encoding human IRE-BP to chromosome 9 (15).

After initial characterization of IRE-BP, the amount of binding activity in a cell lysate was found to increase under conditions of iron deprivation and to decrease in response to an iron load (9, 16). The magnitude of the change in the binding activity in cell lysates after manipulations of intracellular iron was reflected in the quantitative change of ferritin translation in response to iron perturbations. A time course showing progressive changes in the rates of ferritin biosynthesis demonstrated concomitant changes in the extent of binding activity (17). Thus, the level of IRE-BP activity provided a stable record of the iron status of the cell. The lower binding activity seen in iron-replete cells was demonstrated to be entirely recovered *in vitro* by reduction with thiol agents (17). This observation led us to propose a "sulfhydryl switch" mechanism, whereby the level of cellular iron switched the protein from a lower-affinity (oxidized) form to a high-affinity (reduced) form(s). Heme has been proposed to mediate the inhibition of the interaction of IRE-BP with its target RNA (18, 19). However, heme can nonspecifically inhibit numerous nucleic acid–protein interactions, and the significance of inhibition of binding by heme in this setting is unclear (20, 21).

We have used an IRE RNA affinity-resin purification from a human liver cytosolic extract and two-dimensional gel electrophoresis to isolate the IRE-BP for protein sequencing. After sequencing of peptides, a 3.4-kilobase (kb) cDNA clone was isolated by hybridization with a degenerate oligonucleotide corresponding to one partial peptide sequence. An open reading frame (ORF)§ of 790 amino acids was identified that contained 14 of the sequenced peptides. A second, distinct cDNA clone selected with the same degenerate oligonucleotide was found to encode a closely related protein, although none of the sequences predicted by DNA sequencing§ of this

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.

Abbreviations: IRE, iron-responsive element; IRE-BP, IRE-binding protein; UTR, untranslated region; TfR, transferrin receptor; RFLP, restriction fragment length polymorphism; ORF, open reading frame.

§The sequences reported in this paper have been deposited in the GenBank data base (accession nos. M37836 for IRE-BP and M37835 for clone 10.1).

second clone perfectly matched any of the peptides derived from protein isolated by the IRE-affinity resin.

MATERIALS AND METHODS

IRE-BP Purification. Purification by means of a single round of IRE-affinity chromatography was accomplished using described methods (22). Material bound to the affinity column was eluted in 9.5 M urea/0.1% SDS/0.5 mM dithiothreitol/0.2% Nonidet P-40/4% Ampholytes, pH 3–10 (Serva), incubated for 2 hr at 42°C, and subjected to isoelectric focusing in a tube gel (23). Proteins were subjected to electrophoresis in the second dimension according to the methods of Laemmli (24). Proteins were transferred electrophoretically to nitrocellulose (Schleicher & Schuell). After transfer, the membrane was immersed for 60 sec in 0.1% Ponceau S dye (Sigma)/1% aqueous acetic acid. The specific region corresponding to the IRE-BP was identified. This region was cut from the nitrocellulose and incubated for 30 min at 37°C in 1.2 ml of 0.5% polyvinylpyrrolidone-40 dissolved in 100 mM acetic acid to prevent subsequent adsorption of protease to the nitrocellulose during digestion. Excess polyvinylpyrrolidone-40 was removed by extensive washing with water (25, 26). The protein (100–200 pmol) was digested with either Asp-N or Lys-C sequencing grade proteases (Boehringer Mannheim) at an estimated enzyme-to-protein ratio of 1:10. Asp-N digestion was done in 50 mM $NaHPO_4$, pH 8, at 37°C for 18 hr, and Lys-C digestion was done in 25 mM Tris/1 mM EDTA/5% CH_3CN, pH 8.5, at 37°C for 18 hr. Resulting peptides were separated on an Applied Biosystems model 130A microbore HPLC system with a 2.1 × 210-mm Applied Biosystems RP-300 column equilibrated in 98% solvent A (0.1% trifluoroacetic acid in water)/2% solvent B [70% (vol/vol) acetonitrile and 0.09% trifluoroacetic acid in water] using a linear gradient from 2% solvent B to 80% solvent B over 55 min. Amino acid sequences were determined by using an Applied Biosystems model 477A sequencer and modified Edman chemistry. Cleaved amino acid derivatives were identified with an on-line Applied Biosystems model 120A phenylthiohydantoin analyzer. Peaks containing double peptide sequences were reduced with 10 mM dithiothreitol at 37°C for 60 min and rechromatographed as described above.

Library Screening. A 64-fold degenerate 20-mer derived from a peptide containing the amino acid sequence NMCPEYG was synthesized on an Applied Biosystems 381A DNA synthesizer and subsequently labeled with [γ-^{32}P]ATP (ICN) and polynucleotide kinase (Boehringer Mannheim) (27). A cDNA λ Zap II library (Stratagene) derived from human YT cells (28) was probed with a probe of 500,000 cpm/ml. Duplicate filters (DuPont/NEN) from plates containing a total of 400,000 phage plaques were hybridized after prehybridization in 6× SSC (1× SSC is 0.15 M sodium chloride/0.015 M sodium citrate, pH 7), 2× Denhardt's solution (1× Denhardt's solution is 0.02% polyvinylpyrrolidone/0.02% Ficoll/0.02% bovine serum albumin)/1% SDS/salmon sperm DNA at 100 μg/ml. After 16-hr hybridization, filters were washed with 2× SSC/1% SDS at room temperature, followed by two washes with 2× SSC/1% SDS at 53°C for 10 min. By using this approach, 21 positive clones were identified that were then transduced into Bluescript plasmid vectors by using the helper phage R408 (Stratagene).

DNA Sequencing. Standard Sanger dideoxynucleotide chain-termination sequencing reactions were performed with Sequenase (Version 2.0; United States Biochemical). Inosine lanes were electrophoresed in parallel to clarify sequence in regions containing sequence ambiguities.

Chromosomal Localization. DNA samples from 97 independent human–mouse or human–hamster somatic cell hybrids and subclones (28, 29) were digested with *Eco*RI, and the fragments were resolved on 0.7% agarose gels. Southern blots were hybridized with the probe used in the analyses of Northern (RNA) and Southern analyses. A high-stringency wash in 0.1× SSC/0.2% SDS at 55°C was performed. The presence of the hybridizing human sequences in the DNA samples was correlated with the specific human chromosomes retained in each somatic cell hybrid.

RESULTS AND DISCUSSION

Purification of IRE-BP for Protein Sequencing. We had demonstrated (22) that an IRE-containing RNA affinity resin can be used to purify specifically what appears to be a single protein by using two rounds of affinity isolations starting with unfractionated human liver cytosolic extract. The purification was assessed with a gel-mobility-shift assay for IRE-BP activity. A single protein of ≈90 kDa was isolated in the purification. This protein approximated in size to the single band on SDS/PAGE that could be affinity labeled in total liver cytosol by UV-crosslinking to a labeled IRE-containing RNA. Although two rounds of affinity resin were required to achieve a purified protein, the IRE-BP could be detected by one-dimensional SDS/PAGE after only a single round of purification (22). To obtain enough protein for sequencing, we chose to replace the second round of affinity purification with two-dimensional isoelectric focusing–SDS/PAGE. Results of a typical round of purification followed by two-dimensional gel are shown in Fig. 1. As with two-step affinity purification, the results of an IRE-binding resin are compared with the material isolated from a control resin. This procedure was followed nine times and was reproducible. Essen-

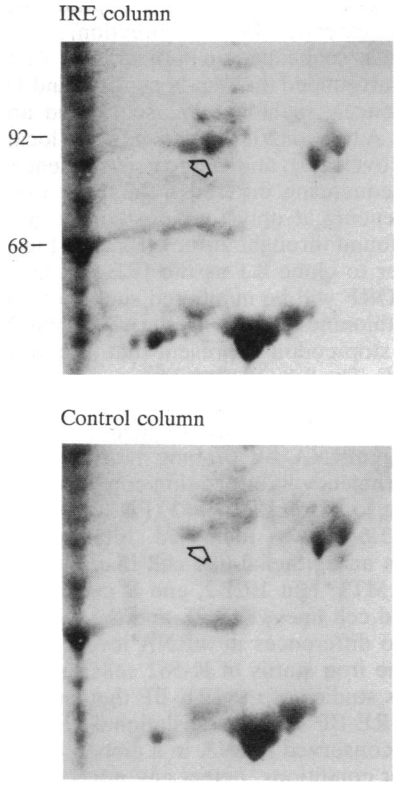

FIG. 1. Isolation of the IRE-BP. After affinity purification and two-dimensional gel electrophoresis of eluates from resins known to bind IRE-BP and from control resins, proteins were electrophoretically transferred to nitrocellulose and stained with Ponceau S. Molecular masses are marked in kDa, and arrows point to the gel region from which the IRE-BP was excised. Stainable protein was notably absent from this region of the control gel, but this corresponding region was excised and processed similarly.

tially every spot on the IRE gel appeared on the control gel, except one prominent protein that migrated with an apparent molecular mass of ≈90 kDa and a pI between 6 and 7 (Fig. 1).

Several independent isolations were done, and the corresponding regions of both control and IRE gels, after transfer to nitrocellulose, were digested with either of two endoproteinases. The resulting peptides were resolved by HPLC. The high yield and specificity of the IRE-binding protein purification was shown by comparing the peptide maps for the two gels after digestion with Lys-C (data not shown). Eleven different peptide sequences were obtained from Lys-C digestions, and four peptide sequences were obtained from peptides derived from digestion with Asp-N protease. Three of the four sequences obtained from the latter enzyme digestion overlapped with Lys-C-derived peptides.

Isolation of a cDNA Clone. A 7-amino acid sequence within a peptide of 15 amino acids was used to construct a 20-mer with 64-fold degeneracy. By using this probe, 21 positive clones were identified from an initial screening of 400,000 phage plaques. The positive clones were plaque-purified and compared by restriction endonuclease mapping and Southern hybridization with the screening oligonucleotide. Four clones have thus far been identified that have unique restriction maps and do not appear to represent different regions of the same cDNA. The screening oligonucleotide was used as a sequencing primer, and two of the clones have been completely sequenced. With additional sequencing primers that would hybridize to both strands of the cDNA, one of the clones, initially called 8.1, yielded the identical predicted amino acid sequence for the region containing the oligonucleotide used for screening as well as the flanking amino acids known to be present at the amino and carboxyl termini of the peptide as determined by the direct peptide analysis. This sequence was extended in both directions, and the entire 3.4-kb clone was sequenced on both strands. An ORF of 790 amino acids surrounded the Asp-N peptide, and 13 remaining peptide sequences, including 192 sequenced amino acids, were located. A total of 208 amino acids was located (including sequence overlaps), and amino acid sequences predicted from DNA sequencing differed from those obtained from peptide sequencing at only four positions. The sequenced peptides are found throughout the ORF (see Fig. 3). We will hereafter refer to clone 8.1 as the IRE-BP, and the amino acids in this ORF will be numbered such that the proposed initiating methionine shown in Fig. 3 is number 48. At position 1, a stop codon is present that is in-frame with an upstream ORF. The full ORF could even extend further 5′ of the proposed initiating methionine at position 48.

The cDNA clone was used to identify the size of the corresponding mRNA by probing Northern (RNA) blots under high-stringency hybridization conditions with a probe corresponding to bases 1262–1863 (Fig. 2). A single mRNA species of ≈3.6 kb was identified. Identical results were obtained from an epithelial-like cell line (HeLa), lymphoid lines (T-cells MT1, Hut 102B2, and B cells Raji, 4672, and 8221), myeloid cell lines (K-562), and the gibbon T-cell line MLA 144. No differences in mRNA levels were seen as a function of the iron status of K-562 cells, as was expected from previous studies of the IRE-BP that suggested regulation of the IRE-BP is posttranslational (16). The cDNA recognized a conserved mRNA in a gibbon cell line but did not, under our conditions, detect any mRNA in murine cell lines, although murine IRE-BP activity is readily detectable in gel-shift assays (15). The clone isolated contained 3355 bases [excluding poly(A)] as determined by sequencing. Identification of a 40-base poly(A) tail ensured that the entire 3′ UTR was present in the clone. The ORF ends at base 2630, and the 3′ UTR is 725 bases. From the estimated message size of 3.6 kb, we are probably not missing >150–200 bases in the 5′ region of this cDNA clone.

```
      |  10      |  20      |  30      |  40      |  50      |  60
    1 cccacacact gacatgacag ggcttgggca agactcctgt tctacttatc cttttgaaat 60
   61 acctcaccct gccactccac catgtatgat cattccagag atctttgtga ctagagttag 120
  121 tgtcctagga aaaccagaac tcagaacttg cctccatggt tgagtaacaa gctgtacaag 180
  181 aacatagaag tgccatttaa gcctgctcgt gtcatcctgc aggactttac gggtgtgccc 240
  241 gctgtggttg actttgctgc aatgcgtgat gctgtgaaaa agttaggagg agatccagag 300
  301 aaaataaacc ctgtctgccc tgctgatctt gtaatagatc attccatcca ggttgatttc 360
  361 aacagaaggg cagacagctt acagaagaat caagacctgg aatttgaaag aaatagagag 420
  421 cgatttgaat ttttaaagtg gggttcccag gcttttcaca acatgcggat tattcccact 480
  481 ggctcaggaa tcatccacca ggtgaatttg gaatatttgg caagagtggt atttgatcag 540
  541 gatggatatt attacccaga cagcctcgtg ggcacagact cgcacactac catgattgat 600
  601 ggcttgggca ttcttggttg gggtgtcggt ggtattgaag cagaagctgt catgctgggt 660
  661 cagccaatca gtatggtgct tcctcaggtg attggctaca ggctgatggg gaagccccac 720
  721 cctctggtaa catccactga catcgtgctc accattacca agcacctccg ccaggttggg 780
  781 gtagtgggca aatttgtcga gttcttcggg cctggagtag cccagttgtc cattgctgac 840
  841 cgagctacga ttgctaacat gtgtccagag tacggagcaa ctgctgcctt tttcccagtt 900
  901 gatgaagtta gtatcacgta cctggtgcaa acaggtcgtg atgaagaaaa attaaagtat 960
  961 attaaaaaat atcttcaggc tgtaggaatg tttcgagatt tcaatgaccc ttctcaagac 1020
 1021 ccagacttca cccaggttgt ggaattagat ttgaaaacag tagtgccttg ctgtagtgga 1080
 1081 cccaaaaggc ctcaggacaa agttgctgtg tccgacatga aaaaggactt tgagagctgc 1140
 1141 cttggagcca agcaaggatt taaaggattc caagttgctc ctgaacatca taatgaccat 1200
 1201 aagacctttа tctatgataa cactgaattc accettgctc atggttctgt ggtcattgct 1260
 1261 gccattacta gctgcacaaa caccagtaat ccgtctgtga tgttaggggc aggattgtta 1320
 1321 gcaaagaaag ctgtggatgc tggcctgaac gtgatgcctt acatcaaaac tagcctgtct 1380
 1381 cctgggagtg gcgtggtcac ctactaccta caagaaagcg gagtcatgcc ttatctgtct 1440
 1441 cagcttgggt ttgacgtggt gggctatggc tgcatgacct gcattggcaa cagtgggcct 1500
 1501 ttacctgaac ctgtggtaga agccatcaca cagggagacc ttgtagctgt tggagtcatc 1560
 1561 attggaacag gaattttgaa ggctgagctt taccccaaca cccgggccaa ctatttagcc 1620
 1621 tctcccccct tagtaatagc atatgcaatt gctggaacca tcagaatcga ctttgagaaa 1680
 1681 gagccattgg gagtaaatgc aaagggacag caggtatttc tgaaagatat ctggccgact 1740
 1741 agagacgaga tccaggcagt ggagcgtcag tatgtcatcc cggggatgtt taaggaagtc 1800
 1801 tatcagaaaa tagagactgt gaatgaaagc tggaatgcct tagcaacccc atcagataag 1860
 1861 ctgtttttct ggaattccaa atctacgtat atcaaatcac caccattctt tgaaaacctg 1920
 1921 actttggatc ttcagcccc taaatctata gtggatgcct atgtgctgct aaatttggga 1980
 1981 gattcggtaa caactgacca catctcccca gctggaaata ttgcaagaaa cagtcctgct 2040
 2041 gctcgctact taactaacag aggcctaact ccacgagaat tcaactccta tggctccccc 2100
 2101 gaggtaatga cgccgtcatg gcacggggaa catttgccaa cattgcgctt gttaaacaga 2160
 2161 tttttgaaca agcaggcacc acagactatc catctgcctt ctggggaaat ccttgatgtg 2220
 2221 tttgatgctg ctgagcggta ccagcaggca ggccttcccc tgatcgttct ggctggcaaa 2280
 2281 gagtacggtg caggcagctc ccgagactgg gcagctaagg gccctttcct gctgggaatc 2340
 2341 aaagccgtcc tggccgagag ctacgagcgc attccaccgca gtaacctggt tgggatgggt 2400
 2401 gtgatcccac ttgaatatct ccctggtgag aatgcagatg ccctggggct cacagggcaa 2460
 2461 gaacgataca ctatcattat tccagaaaac ctcaaaccac aaatgaaagt ccaggtcaag 2520
 2521 ctggatactg gcaagacctt ccaggctgtc atgaggtttg acactgatgt ggagctcact 2580
 2581 tatttcctca acgggggcat cctcaactac atgatccgca agatggccaa gtaggagacg 2640
 2641 tgcacttggt cgtgcgccca gggaggaagc cgcaccacca gccagcgcag gccctggtgg 2700
 2701 agaggcctcc ctggctgcct ctgggagggg tgctgccttg tagatggagc aagtgagcac 2760
 2761 tgagggtctg gtgccaatcc tgtaggcaca aaaccagaag gttctacatt ctctattttt 2820
 2821 gttaatcatc ttctcttttt ccagaatttg gaagctagaa tggtgggaat gtcagtagtg 2880
 2881 ccagaaagag agaaccaagc ttgtctttaa agttactgat cacaggacgt tgctttttca 2940
 2941 ctgtttccta ttaatcttca gctgaacaca agcaaacctt ctcaggaggt gtctcctacc 3000
 3001 ctcttattgt tcctcttacg ctctgctcaa tgaaaccttc ctcttgaggg tcatttтcct 3060
 3061 ttctgtatta attatacaag tgttaagtga catagataag aactttgcac acttcaaatc 3120
 3121 agagcagtga ttctctcttc tctccccttt tccttcagag tgaatcatcc agactcctca 3180
 3181 tggataggtc gggtgttaaa gttgttttga ttatgtacct tttgatagat ccacataaaa 3240
 3241 agaaatgtga agttttcttt tactatcttt tcatttatca agcagagacc tttgttggga 3300
 3301 ggcggtttgg gagaacacat ttctaatttg aatgaaatga aatctatttt cagtg      3355
      |  10      |  20      |  30      |  40      |  50      |  60
```

FIG. 2. The DNA sequence of the IRE-BP. The putative translation start site is at base 262, and upstream stop codons are at 113 and 119. The translation stop site is at base 2631, and the poly(A) sequence begins at 3356.

Chromosomal Localization. In a previous study (15) we had localized the gene encoding the human IRE-BP to chromosome 9. This study used the gel-mobility-shift assay to localize binding activity to a specific chromosome in human–rodent hybrid cell lines and was based on a distinguishable difference in migration of an IRE-containing probe when it was complexed to human or rodent IRE-BP. A 647-base-pair (bases 1262–1863) fragment contained within the reading frame of clone 8.1 was isolated and used to hybridize to human genomic DNA derived from a panel of 97 rodent–human hybrid cell lines. Analysis of the Southern hybridizations from these hybrids localized the DNA sequence to chromosome 9. In contrast, the IRE-BP sequence segregated discordantly (≥24%) with other human chromosomes. Two discordancies (2% discordancy) were noted with chromosome 9 that may represent retention of only a portion of this chromosome in the discordant hybrids.

The same probe was used to assess the presence of restriction fragment length polymorphisms (RFLPs). A high-frequency RFLP was found in *Eco*RI digests, and the IRE-BP gene was detected as 11- to 12-kb or 5-kb allelic bands that displayed Mendelian segregation in families (data not shown). This RFLP will permit regional localization of the IRE-BP gene on chromosome 9 by genetic linkage analysis. No RFLPs were detected with 11 other enzymes (*Bam*HI, *Hin*dIII, *Xba* I, *Sac* I, *Taq* I, *Pvu* II, *Bgl* II, *Pst* I, *Msp* I, *Eco*RV, *Kpn* I) on examination of 10 unrelated individuals.

Cloning of a Second, Closely Related Protein. One of the striking findings of the library screening came when we examined a second distinct clone, 10.1, for which the degenerate screening oligonucleotide also served as a successful sequencing primer. Despite differences in the nucleotide sequence between clones 10.1 and 8.1, we could locate the identical amino acid sequence (NMCPEYG) used to generate the screening oligonucleotide. The surrounding amino acid sequence matched the peptide from clone 8.1 at 13 of 15 positions. Further sequencing revealed that this second peptide is contained within an ORF of >800 amino acids. Clone 10.1 clearly encoded a protein very closely related to the IRE-BP (8.1). Overall, the two proteins are 57% identical and 75% similar when conservative amino acid changes are considered (30, 31). Fig. 3 shows that similarity extends over the entire ORFs, although certain regions show significant identity, whereas others are less conserved. Each protein contains a single insertion that interrupts the aligned similarities. The computer alignment (32) suggests that clone 10.1 contains an inserted stretch of 73 amino acids between amino acids 37 and 38 of the IRE-BP, whereas clone 10.1 lacks a region homologous to the sequence between amino acids 436 and 470 of the IRE-BP. At the 3′ end of the ORFs, the two proteins end at homologous positions. However, the correspondence at the 5′ end is less clear. We have placed the initiating methionine for the IRE-BP as the first methionine that follows two stop codons in-frame with the ORF of the protein. However the ORF from clone 10.1 continues in the 5′ direction through this region. Perhaps most suggestive that the coding region of clone 8.1 actually extends further 5′ is the clear continuation of amino acid similarity for an additional 27 amino acids 5′ of the position of the putative IRE-BP initiating methionine. In this region of amino acid conservation, the third bases of the codons are not conserved, a feature suggestive of selective retention of amino acid sequence. We have extensively sequenced this region of the IRE-BP clone to check for sequencing errors and have found none. Even so, the possibility exists that the true start of the IRE-BP ORF is further 5′, although the size of the protein and the mRNA would not allow for a very significant amino-terminal extension of the protein.

Despite the extensive sequence similarity between the two proteins, there is no evidence that the IRE affinity purification isolates any of the 10.1 protein. None of the peptides sequenced by Edman degradation is represented exactly in the 10.1 protein, whereas 14 peptides were encoded by the clone 8.1 ORF. Whether these two proteins are derived from a larger family of genes remains to be determined. Clone 10.1 has been localized to chromosome 15 (data not shown).

Structural Characteristics of the Cloned Protein. The ORF of the IRE-BP (8.1) clone encodes 790 amino acids corresponding to a protein with a molecular mass of 87 kDa and a pI, in the absence of any posttranslational modifications, of 6.1. Search of available data banks failed to find any previously reported protein or gene with significant sequence similarity at either the amino acid or nucleotide level (32). We have not yet directly shown that a protein expressed from clone 8.1 has IRE-binding activity (see *Note*). However, the fact that the purification protocol was based on the capacity of the isolated protein to function as an IRE-BP, and 14 different peptide sequences, found in the clone, were obtained from separate preparations and digests indicates that this is indeed the 90-kDa IRE-BP. Size fractionation, UV-crosslinking, and three published purification protocols (22, 33, 34) have shown that 90 kDa is the approximate molecular mass of the IRE-BP. The affinity purification scheme used in our study has clearly been shown to specifically purify both a functional IRE-BP and isolate a single protein species (22). A similar protocol recently reported (34) confirms the validity of this approach. As predicted by the specificity of the

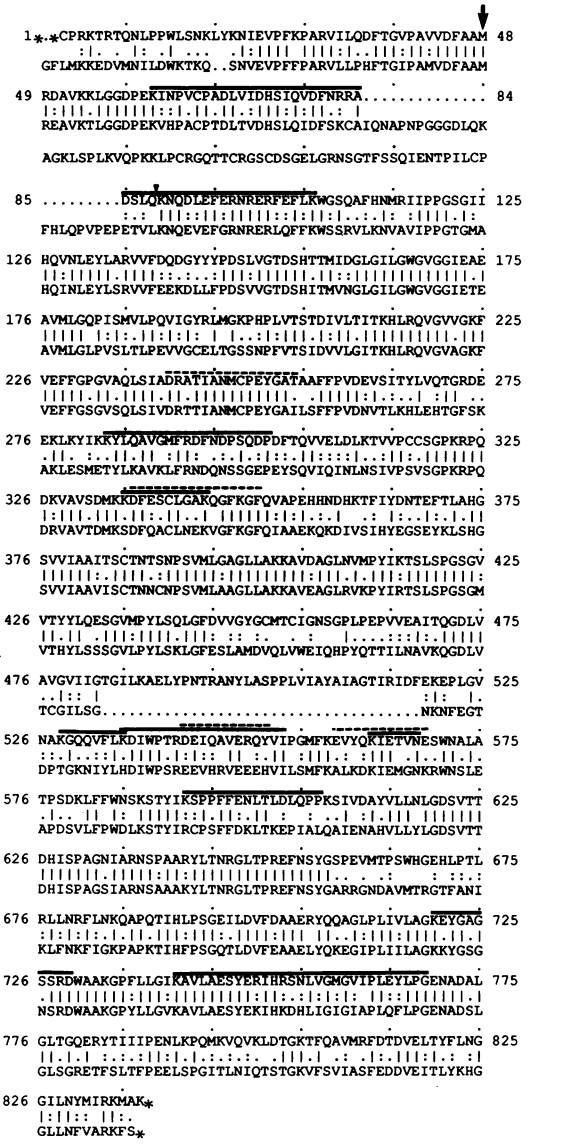

FIG. 3. Amino acid sequence of the IRE-BP (using single-letter amino acid code). An arrow at position 48 denotes the putative IRE-BP translation start site, and numbering of amino acids begins after an upstream in-frame stop codon. Solid black lines denote peptides obtained from Lys-C digestions, and dashed lines represent peptide sequences obtained from N-Asp digestion. Three sequence overlaps are present. An inverted triangle denotes a division between two separate Lys-C peptide sequences. Also shown is the aligned protein sequence derived from the related clone 10.1. Identities are represented with vertical lines, and double and single dots bridging the sequences represent scores derived from a relatedness odds matrix (30, 31) that primarily reflects chemical relatedness of amino acids. The top numbered rows of each pair represent the IRE-BP. Stop codons are represented by asterisks, and gaps are represented by single dots positioned above or below the letters for amino acids of the related protein. Between amino acids 84 and 85 of the IRE-BP, a large 73-amino acid unnumbered insertion of clone 10.1 is represented.

purification scheme, we could locate sequences from peptides generated with two different endoproteases within a single clone. Overlapping sequence found between different preparations confirms the reproducibility of the isolation of a single protein, and none of the peptides identified from the IRE-containing resin was found in the control resin eluate. The fact that clone 8.1 contains 14 of the sequenced peptides and the added fact that the cloned cDNA is encoded by a single gene present on chromosome 9, as predicted from

functional analysis (15), leads us to conclude that clone 8.1 must encode the IRE-BP. We cannot rule out that the clone 10.1 protein may also bind IREs.

Whether the IRE-BP defined functionally truly consists of a single protein can be questioned. Some published reports of purifications have suggested a doublet on SDS/PAGE (34), and UV-crosslinking of IRE-containing RNA to rat lysates has identified two bands at ≈90 kDa (8). Further, we often observed several bands in the gel-mobility-shift assay (21). Whether these observations result from modifications of a single protein or are better explained by the existence of additional IRE-BP polypeptides needs to be addressed.

Absent from the clone 8.1 sequence are predicted ribonucleoprotein consensus sequences (35, 36), zinc finger motifs (37), or very basic domains. The IRE-BP is a relatively large protein compared with many characterized nucleic acid-binding regulatory proteins. Several interesting motifs can be identified that may have important functional implications. Beginning at amino acid 161 (depicted in Fig. 3) is a consensus sequence for the binding of adenine nucleotides (GXGXXG XXXXG) found in ATP-dependent kinases and adenine dinucleotide-binding proteins (38, 39). The 8.1 sequence lacks the AXK sequence downstream that is present in kinases. The exact sequence within this consensus motif most closely resembles the NADH binding site for an *Escherichia coli* oxidoreductase (39). Interestingly, the region containing this motif is among the most highly conserved between the IRE-BP and the 10.1 protein (Fig. 3). The presence of a potential adenine dinucleotide-binding site is particularly intriguing in terms of our model, suggesting the use of a redox sulfhydryl switch in the regulation of the affinity state of the IRE-BP (16). NADH might serve as a potential electron donor in IRE-BP reduction.

Further intriguing features of the primary sequence are cysteine and histidine clusters that might function as metal-binding sites. Beginning at amino acid 356, the sequence EHHNDH is similar to proposed metal-binding sites (40). In addition, two cysteine clusters are seen near amino acids 318 and 452. The latter contains the sequence CXXC, a sequence often seen in iron-containing proteins, generally as part of an iron–sulfur center (41). Such closely spaced cysteine pairs may serve as either metastable iron interaction sites or sites of formation of stable iron–sulfur complexes. High potential iron–sulfur clusters are generally coordinated by two cysteines and two histidines (42).

Note. cDNA clones corresponding to portions of the rabbit IRE-BP (also termed the ferritin repressor protein or FRP) have been identified recently in an expression library with antibodies capable of immunoprecipitating the rabbit IRE-binding activity (43). The partial sequence of the protein predicted from these rabbit cDNA clones is nearly identical to that of the human IRE-BP encoded by our clone 8.1.

We thank J. Gnarra, J. Casey, D. Koeller, V. Zenger, and M. Rice for helpful discussions and input. We thank Warren Leonard for the gift of the YT human T-cell cDNA library and for RNA samples. We gratefully acknowledge Dr. William Walden (University of Illinois) for making his rabbit ferritin repressor protein sequence available before publication for this comparison. S.K. is supported by the Howard Hughes Medical Institute–National Institutes of Health Research Scholars Program, and F.S. is a Fellow of the Robert Wood Johnson Foundation. This work was supported in part by National Institutes of Health Grant HL35762 to W.H.B.

1. Klausner, R. D. & Harford, J. B. (1989) *Science* **246,** 870–872.
2. Theil, E. C. (1990) *J. Biol. Chem.* **265,** 4771–4774.
3. Hentze, M. W., Caughman, S. W., Rouault, T. A., Barriocanal, J. G., Dancis, A., Harford, J. B. & Klausner, R. D. (1987) *Science* **238,** 1570–1573.
4. Aziz, N. & Munro, H. N. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 8478–8482.
5. Casey, J. L., Hentze, M. W., Koeller, D. M., Caughman, S. W., Rouault, T. A., Klausner, R. D. & Harford, J. B. (1988) *Science* **240,** 924–928.
6. Mullner, E. W. & Kuhn, L. C. (1988) *Cell* **53,** 815–825.
7. Casey, J. L., Koeller, D. M., Ramin, V. C., Klausner, R. D. & Harford, J. B. (1989) *EMBO J.* **8,** 3693–3699.
8. Leibold, E. A. & Munro, H. N. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2171–2175.
9. Rouault, T. A., Hentze, M. W., Caughman, S. W., Harford, J. B. & Klausner, R. D. (1988) *Science* **241,** 1207–1210.
10. Koeller, D. M., Casey, J. L., Hentze, M. W., Gerhardt, E. M., Chan, L. N., Klausner, R. D. & Harford, J. B. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 3574–3578.
11. Brown, P. H., Daniels-McQueen, S., Walden, W. E., Patino, M. M., Gaffield, L., Bielser, D. & Thach, R. E. (1989) *J. Biol. Chem.* **264,** 13383–13386.
12. Hentze, M. W., Caughman, S. W., Casey, J. L., Koeller, D. M., Rouault, T. A., Harford, J. B. & Klausner, R. D. (1988) *Gene* **72,** 201–208.
13. Rothenberger, S., Müllner, E. W. & Kühn, L. C. (1990) *Nucleic Acids Res.* **18,** 1175–1179.
14. Mullner, E. W., Neupert, B. & Kuhn, L. C. (1989) *Cell* **58,** 373–382.
15. Hentze, M. W., Seuanez, H. N., O'Brien, S. J., Harford, J. B. & Klausner, R. D. (1989) *Nucleic Acids Res.* **17,** 6103–6108.
16. Haile, D. J., Hentze, M. W., Rouault, T. A., Harford, J. B. & Klausner, R. D. (1989) *Mol. Cell. Biol.* **9,** 5055–5061.
17. Hentze, M. W., Rouault, T. A., Harford, J. B. & Klausner, R. D. (1989) *Science* **244,** 357–359.
18. Ward, J. H., Jordan, I., Kushner, J. P. & Kaplan, J. (1984) *J. Biol. Chem.* **259,** 13235–13240.
19. Lin, J. J., Daniels-McQueen, S., Patino, M. M., Gaffield, L., Walden, W. E. & Thach, R. E. (1990) *Science* **247,** 74–77.
20. Rouault, T., Rao, K., Harford, J., Mattia, E. & Klausner, R. D. (1985) *J. Biol. Chem.* **260,** 14862–14866.
21. Haile, D. J., Rouault, T. A., Harford, J. B. & Klausner, R. D. (1990) *J. Biol. Chem.* **265,** in press.
22. Rouault, T. A., Hentze, M. W., Haile, D. J., Harford, J. B. & Klausner, R. D. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 5768–5772.
23. O'Farrell, P. H. (1975) *J. Biol. Chem.* **250,** 4007–4021.
24. Laemmli, U. K. (1970) *Nature (London)* **227,** 680–685.
25. Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E. & Kent, S. B. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 6970–6974.
26. Hampton, B., Burgess, W. H., Marshak, D. R., Cullen, K. J. & Perdue, J. F. (1989) *J. Biol. Chem.* **264,** 19155–19160.
27. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
28. Gnarra, J. R., Otani, H., Wang, M. G., McBride, O. W., Sharon, M. & Leonard, W. J. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 3440–3444.
29. McBride, O. W., Hieter, P. A., Hollis, G. F., Swan, D., Otey, M. C. & Leder, P. (1982), *J. Exp. Med.* **155,** 1480–1490.
30. Dayhoff, M., Schwartz, R. M. & Orcutt, B. C. (1978) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. (Natl. Biomed. Res. Found., Silver Spring, MD), Vol. 5, Suppl. 3, pp. 345–352.
31. Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2444–2448.
32. Devereux, J. R., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395.
33. Walden, W. E., Patino, M. M. & Gaffield, L. (1989) *J. Biol. Chem.* **264,** 13765–13769.
34. Neupert, B., Thompson, N. A., Meyer, C. & Kuhn, L. C. (1990) *Nucleic Acids Res.* **18,** 51–55.
35. Mattaj, I. W. (1989) *Cell* **57,** 1–3.
36. Query, C. C., Bentley, R. C. & Keene, J. D. (1989) *Cell* **57,** 89–101.
37. Berg, J. M. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 99–102.
38. Hanks, S. K., Quinn, A. M. & Hunter, T. (1988) *Science* **241,** 42–52.
39. Scrutton, N. S., Berry, A. & Perham, R. N. (1990) *Nature (London)* **343,** 38–43.
40. Springman, E. B., Angleton, E. L., Birkedal-Hansen, H. & Van Wart, H. E. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 364–368.
41. Stout, C. D. & Robbins, A. H. (1987) in *Iron-Sulfur Protein Research*, eds. Matsubara, H., Katsube, Y. & Wada, K. (Springer, Berlin), pp. 139–148.
42. Gurbiel, R. J., Batie, C. J., Sivaraja, M., True, A. E., Fee, J. A., Hoffman, B. M. & Ballou, D. P. (1989) *Biochemistry* **28,** 4861–4871.
43. Patino, M. M., Beck, M. M. & Walden, W. E. (1990) *FASEB J.* **4,** A1998 (abstr.).