Heterogeneity of mRNA and Protein Products Arising from the Protein 4.1 Gene in Erythroid and Nonerythroid Tissues

Tang K. Tang,*‡ Zhi Qin,§ Thomas Leto,§ Vincent T. Marchesi,§ and Edward J. Benz, Jr.*‡

Departments of *Internal Medicine, ‡Human Genetics, and §Pathology, School of Medicine, Yale University, New Haven, Connecticut 06510

Abstract. Immunologically cross-reactive isoforms of the cytoskeletal element protein 4.1 have been identified in many tissues in which they exhibit heterogeneity of molecular weight, abundance, and intracellular localization. To examine the basis for isoform production in erythroid and nonerythroid tissues, we have compared the structure and expression of cDNAs isolated from human erythroid and nonerythroid sources. We have encountered cDNAs representing many distinct mRNA sequences. These exhibit complete nucleotide sequence homology along most of their lengths. Differences were confined to five sequence blocks designated Motifs I-V, which were present or absent in each mRNA moiety. Motif I was expressed only in erythroid cells; it encodes 21 amino acids in a well-characterized spectrin/actin binding domain. Motif II, located near the COOH terminus of the 80-kD "erythroid" protein 4.1 molecule is present in the vast majority of transcripts from both erythroid and nonerythroid cells. Motifs IV and V alter the 5' untranslated region: simultaneous insertion of Motif IV and deletion of Motif V in the untranslated region inserts a new initiator methionine and establishes a contiguous open reading frame encoding a novel 135-kD protein 4.1 molecule. By immunochemical analysis we have identified the longer isoform in cells. Our results are most consistent with tissue-specific alternative mRNA splicing of transcripts of the protein 4.1 gene to yield numerous isoforms. These isoforms exhibit tissue specificity and alter strategic portions of the molecule. Moreover, we describe a novel high molecular weight form of protein 4.1 that arises by splicing events which allow translation at an upstream site.

P ROTEIN 4.1 is an 80-kD sulfhydryl-rich phosphoprotein originally described on the basis of its electrophoretic mobility in cytoskeletal preparations of mature human red cells. In erythrocytes, protein 4.1 has been shown to be important for maintenance of the structural integrity and flexibility of the red cell membrane and its underlying cytoskeleton (for reviews see Marchesi, 1985; Bennett, 1985). Hereditary defects in the quantity or structure of protein 4.1 are associated with congenital hemolytic anemias (Tchernia et al., 1981; Conboy et al., 1986b). Limited chymotryptic digestion of erythroid protein 4.1 (Fig. 1) yields four fragments, two of which have been tentatively identified as supporting known functions of the molecule (Leto et al., 1986a). The amino-terminal 30-kD domain appears to bind the transmembrane protein glycophorin in the presence of triphosphoinositides (Leto et al., 1986a), whereas the 10-kD domain has been clearly shown to include the site for binding to complexes of erythroid spectrin and actin (Correas et al., 1986a,b). These dual binding sites imply that protein 4.1 in erythrocytes serves as a linking molecule, attaching the spectrin/actin cytoskeletal scaffold to the lipid bilayer by means of the cytoplasmic domains of transmembrane proteins.

Numerous immunochemical studies have identified cross-reactive forms of protein 4.1 in most tissues. These isoforms exhibit remarkable diversity with regard to the numbers of discrete molecular weight forms present in each tissue, and their intracellular localization (Cohen et al., 1982; Spiegel et al., 1984; Aster et al., 1984; Granger and Lazarides, 1984; Goodman et al., 1984; Davies and Cohen, 1985; Baines and Bennett, 1985; Leto et al., 1986b; Tang et al., 1988). In nonerythroid tissues, protein 4.1 appears to be prominent in the nucleus (Tang, T. K., E. J. Benz, Jr., and V. T. Marchesi, manuscript in preparation), in intracytoplasmic fibrillar networks (Cohen et al., 1982), or in the perinuclear region (Leto et al., 1986b) as well as in the subplasmalemma. These findings suggest that erythroid protein 4.1 is the prototype of either a large gene family or a large protein family arising from a single gene by alternative processing of mRNA and/or protein products. Recently, Conboy et al., (1986a) reported the cDNA sequence encoding a protein 4.1 isolated from a reticulocyte cDNA library. We (Tang et al., 1988) and others (Ngai et al., 1987; Conboy et al., 1988)

Tang K. Tang's present address is the Institute of Biomedical Sciences, Academia Sinica, Taipei 11529, Taiwan, ROC. Thomas Leto's present address is National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892. Address reprint requests to Edward J. Benz, Jr.
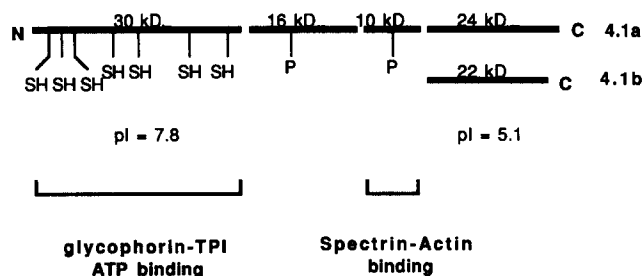
Figure 1. A molecular model of human erythrocyte protein 4.1 deduced from limited proteolytic digestions and chemical degradation studies (Leto et al., 1986a).

have subsequently presented evidence that nonerythroid and erythroid protein 4.1 exhibited discrete regional differences in mRNA structure that most likely arise by alternate mRNA splicing.

In this report, we describe results of a more detailed comparison of mRNAs encoded by erythroid and nonerythroid (lymphoid) cells. To our surprise, we encountered sequence motifs inserted or deleted not only into the protein coding regions of these mRNAs, but also into the 5' untranslated region of the prototypical reticulocyte mRNA. When two such splicing events occur in concert, a novel high molecular weight protein 4.1 isoform is generated; this isoform appears to localize both in the nucleus and along the submembranous sites at intercellular junctions (Tang, T. K., E. J. Benz, Jr., and V. T. Marchesi, manuscript in preparation), suggesting that this family of protein 4.1 isoforms can serve as a structural component within the nucleus as well as being part of the membranous cytoskeleton. In addition, we have found distinctly nonrandom accumulations of mRNAs containing or lacking specific motifs in different tissues. The tissue-specific expression of several motifs is consistent with their putative functions.

Materials and Methods

Cell Culture

MOLT 4, K562, HEL, uninduced and DMSO-induced MEL cells were maintained in the medium as previously described (Tang et al., 1988). HeLa cells were maintained in DME supplemented with 10% FCS, 2 mM L-glutamine, 100 μ/ml penicillin, and 100 Ug/ml streptomycin.

Isolation and DNA Sequence Analysis of Protein 4.1 cDNA Clones

Protein 4.1-specific clones were isolated from MOLT 4 (Tang et al., 1988), and human fetal liver and bone marrow cDNA libraries (kindly provided by Dr. B. Forget, Yale University). These libraries were screened with protein 4.1-specific cDNA probes (pTM-1 and -2) which have been reported

previously (Tang et al., 1988). cDNA inserts from the positive clones were subcloned into pGEM-7Zf(+) (Promega Biotec, Madison, WI) or bluescript (Stratagene, La Jolla, CA) vectors which permit direct dideoxynucleotide sequencing of inserted DNA. Double-strand DNA sequencing was performed using GemSeq K/RT sequencing kit (Promega Biotec). To identify the sequence motifs within the cDNA clone, a series of synthetic oligonucleotides that are flanked by motif sequences were made. The synthetic oligonucleotides used for this purpose are shown in Table I.

The synthetic oligonucleotides were assembled on a DNA synthesizer (Applied Biosystems, Inc., Foster City, CA) and purified as described (Tang et al., 1988). These short oligonucleotides were used as sequencing primers to obtain the nucleotide sequence of specific sequence motifs in the clones.

RNA Isolation and Blot Hybridization

Total and poly(A)$^+$ RNAs isolated from cultured cell lines, leukocytes of a chronic myelogenous leukemia patient, and reticulocytes were prepared as described (Tang et al., 1988). Poly(A)$^+$ RNAs were fractionated on 1.4% agarose formaldehyde gels, followed by transfer to Nytran filters (Schleicher & Schuell, Inc., Keene, NH) and hybridization to pTM-1 cDNA probe. Conditions for hybridization and washing were performed as previously described (Tang et al., 1988).

RNase Protection Analysis

30 μg total RNA was hybridized to the antisense probes (see Fig. 4), followed by digestion with RNase A and T1, using a modification of previously described methods (Zinn et al., 1983). Briefly, 30 μg dried RNA samples were dissolved in 30 μl of formamide hybridization buffer (80% formamide/40 mM Pipes, pH 6.7/0.4 M NaCl/1 mM EDTA) containing the antisense probe (5 × 10^5 cpm), heated at 95°C for 10 min, and incubated at 45°C overnight. After hybridization, 300 μl of RNase digestion buffer (10 mM Tris-HCl, pH 7.5/5 mM EDTA/300 mM NaCl) containing 40 μg of RNase A and 3 μg of T1 was added. RNase digestion was performed at 30°C for 1 h. The reaction was stopped by the addition of 20 μl of 10% SDS and 50 μg of proteinase K, followed by incubation at 37°C for 20 min. The reactions were phenol extracted, ethanol precipitated, and dissolved in the loading buffer. The protected fragments were denatured and analyzed on a 7% polyacrylamide/urea gel.

Preparation of Antibodies against Synthetic Peptides

Synthetic peptides 5'-1 (a mixture of three peptides a [EAENSQHQQKEE-GEEC], b [AINSGQQEPQQEESC], and ab' [KEEGEEAINSGQQEPQ-QEESC]) were derived from the predicted amino-terminal end of the larger 4.1 isoform (see Fig. 3 D). Synthetic peptides 30b (GSYTIQSELGDYD-PELHGVDYVSDC), 16a (TQAQTRQASALIDRPAPHFERC), and 24a (GVLLTAQTITSETPSSTTTTKITKC) were derived from the 30-, 16-, and 24-kD domain of erythroid protein 4.1. Synthetic peptides were made with additional cysteine residues in their COOH-terminal ends. Antibodies against these peptides were raised in rabbits after conjugation to keyhole limpet hemocyanin. The peptide-specific antibodies were purified by affinity chromatography on immobilized synthetic peptides as described (Correas et al., 1986b). Rabbit polyclonal antibodies to human erythrocyte protein 4.1 were prepared using standard techniques of immunization as described earlier (Tang et al., 1988). Human erythrocyte protein 4.1 was prepared and purified as described (Leto et al., 1986b).

In Vitro Translation and Immunoprecipitation

Two clones, pLym and pEry2, were constructed from pTM-1 and -2. pLym (−I, +II, −III, +IV, −V), spanning all five motifs, contains Motif II and a new potential translation initiation site which is located within Motif IV.

Table I. Oligonucleotides Used for Detecting Motifs in cDNA Libraries

Motif	Sense	Motif	Antisense
IS	5'AAGGAAGACGAGCCACCT3'	IAS	5'TTCTGGTACAGACTCCAT3'
IIS	5'AGACGTCCCTATTGTCCA3'	IIAS	5'CCTGTGATCACAATTCTC3'
IIIS	5'ATGCACTGCAAGGTTTCT3'	IIIAS	5'TTCCAAGGGACACCACGA3'
VS	5'AAGGTGAAGGAGGTCAG3'	VAS	5'TCTGGATCTTCTCTGAG3'
IVS	5'AGAACGCGGTCGGCCCG3'		

pEry2 (−I, +II, −III, −IV, −V) was constructed by deletion of the Eco RI and Hind III fragment which contains Motif IV from pLym, and contains only the interior translation initiation site. Synthetic sense-capped mRNAs were generated from the SP6 or T7 promoter within the vector. Briefly, these components were added in the following order: 6 μl diethylpyrocarbonate-treated water, 5 μl 5× transcription buffer (200 mM Tris-HCl, pH 7.5/30 mM MgCl₂/10 mM Spermidine/50 mM NaCl), 2.5 μl 0.1 M DTT, 40/U RNasin ribonuclease inhibitor, 5 μl rNTP (2.5 mM ATP/2.5 mM CTP/2.5 mM UTP/0.25 mM GTP), 1.25 μl 2.5 mM m7G(5')ppp(5')G (RNA cap structure analog), 0.2–1 μg linearized DNA, 2 U SP6 or T7 RNA polymerase (Promega Biotec). Incubation was performed at 40°C for 1 h. The sense mRNAs were translated in a rabbit reticulocyte system (Promega Biotec), in the presence of ^{35}S-methionine, to radiolabel newly synthesized proteins. In vitro translation products or immunoprecipitates obtained with antibodies against synthetic peptides were resolved on SDS–polyacrylamide gels and visualized by fluorography as described (Davis et al., 1986).

Immunoblot Analysis

Solubilized proteins in hot SDS from MOLT 4 cells, and human red cell ghosts were prepared as described (Tang et al., 1988). Solubilized proteins were separated by SDS-PAGE and blotted onto nitrocellulose, probed with primary antibodies, and then detected with ^{125}I-labeled *Staphylococcus* protein A (Tang et al., 1988) or an anti-rabbit IgG, which had been alkaline phosphatase conjugated (Promega Biotec).

Results

Diverse Structures of cDNA Clones Encoding Multiple Forms of Protein 4.1 in Erythroid and Lymphoid Cells

Our initial goal was to characterize the genes that encode different protein 4.1 isoforms in erythroid and nonerythroid forms. As discussed below, we isolated cDNAs whose nucleotide sequences were consistent with the notion that there is only one gene for protein 4.1. Multiple isoforms arise by alternative mRNA splicing. We present the characterization of five sequence regions or cassettes, called Motifs I–V, that distinguish different isoforms. We also show that a novel high molecular weight isoform predicted by these cDNA sequences can be detected in some cells by antibodies derived from deduced peptide sequences.

Fig. 1 shows a molecular model for the structure of the prototypical erythroid protein 4.1 species based on mRNA sequence, partial amino acid sequence data, and functional analysis (Leto et al., 1986a). The regions of the putative glycophorin and spectrin/actin binding sites are indicated, as are the distribution of phosphates and clustered sulfhydryl groups.

Fig. 2 diagrams the structure of several representative cDNAs obtained from erythroid and nonerythroid sources (details of these clones are not shown). All cDNAs were identical to one another along the bulk of sequence, and sequence differences were limited to five regions ranging in size from 17 to 105 nucleotides. These are indicated as open or closed blocks labeled Motifs I–V in Fig. 2 (a black box indicates that the sequence is present in the cDNA clones; an open box indicates that the sequence is absent). The complete nucleotide sequence of an mRNA in which all five of the sequence blocks are retained has been published elsewhere (Tang et al., 1988; Tang, T. K., E. J. Benz, Jr., and V. T. Marchesi, manuscript in preparation).

The patterns of variation outlined in Fig. 2 are most consistent with one or a few copies of a protein 4.1 gene; the diversity of mRNA transcripts would then be generated by alternative mRNA splicing. Southern blot analysis of ge-
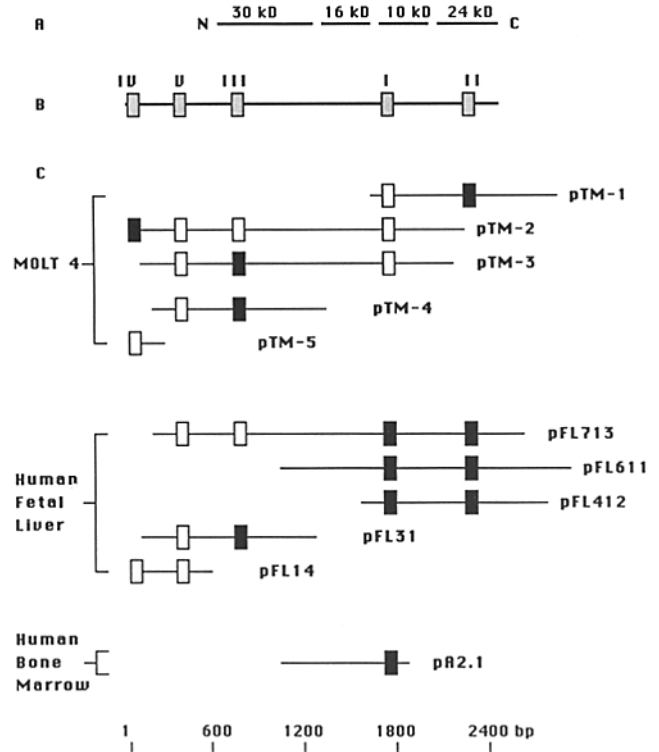

Figure 2. Summary of the diverse protein 4.1 cDNA clones isolated from the cDNA libraries of different source of tissues. (*A*) Structural domains of human erythrocyte protein 4.1. (*B*) Five different nucleotide sequence motifs (I–V) identified in protein 4.1 mRNAs. (*C*) Diverse structure of protein 4.1 cDNA clones isolated from human lymphoid (MOLT 4), fetal liver, and bone marrow cDNA libraries. An open block indicates the sequence is absent in the cDNA clones; a solid block indicates that the sequence is present. The detailed structures of the clones are not shown; the sequence data are being prepared for publication.

nomic DNA reveals one or at most a very few copies of the 4.1 gene (our unpublished results; Tchernia et al., 1981; Ngai et al., 1987; Conboy et al., 1988). The gene copy number is certainly not sufficient to account for the diversity of mRNA sequences observed.

The five sequence blocks that differentiate transcripts from one another have been provisionally named Motifs I–V, in the order of their discovery (Fig. 3 *D*). Motif I is 63 nucleotides long, encoding a 21-amino acid segment within a 10-kD spectrin/actin binding domain. Motif II (102 nucleotides) encodes a 34-amino acid segment near the COOH end of the 22/24-kD domain. Motif III (105 nucleotides) encodes a 35-amino acid peptide near the amino-terminal end of the 30-kD domain. Although this domain includes the glycophorin binding region, precise localization of this function in the 30-kD segment has not been established.

Motifs IV and V (Fig. 3 *D*) can exert more complex effects on the structure and function of protein 4.1 mRNA. The 5' untranslated sequence of the prototypical protein 4.1 mRNA contains a long open reading frame (ORF)[1] contiguous with the ORF-2 in the figure. ORF-2 is translatable from the indicated initiator methionine into an 80-kD polypeptide.

A   N ——— 30 kD ——— 16 kD 10 kD 22/24 kD ——— C

B   5' UT ——— Met ——— 3' UT
                Erythroid   protein 4.1 cDNA prototype

C   Met
    ORF-2
    ORF-1

D   N [IV [ ] V   III   I   II] C
        17nt   80nt | 105nt   63nt   102nt
        Met.   → Met
          a b
        ab' (5'-1)   30b   16a   24a
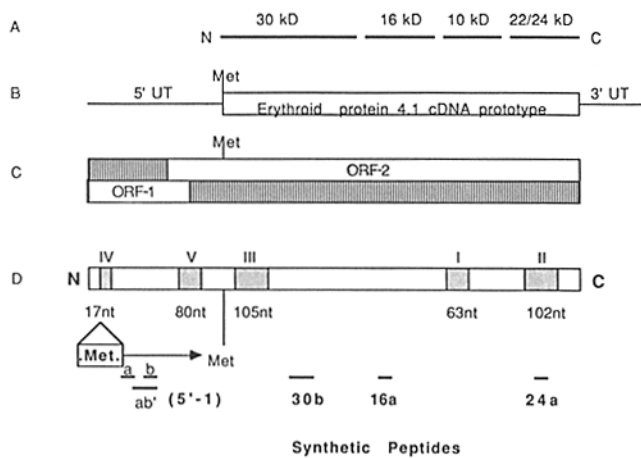
Synthetic Peptides

Figure 3. Diverse forms of erythroid and nonerythroid protein 4.1 and alignment of synthetic peptides used for generation of site-specific antibodies. (*A*) Structural domains of human erythrocyte protein 4.1. (*B*) Erythroid protein 4.1 cDNA clone reported by Conboy et al. (1986a). (*C*) ORF-1 and -2 found in erythroid 4.1 sequence, are indicated by the open boxes. The boxes with vertical lines indicate that there are no other long ORFs in ORF-1 or -2. ORF-1 begins by counting codons from base 1 of our most upstream cDNA; ORF-2, from base 2. (*D*) Variable motifs (I-V) defined by comparison of erythroid (Conboy et al., 1986a) and lymphoid cDNA sequence. Synthetic peptides, 5'-1, 30b, 16a, and 24a, are described in detail in Materials and Methods.

All three possible reading frames contain translation termination codons. Every possible reading frame is thus "closed"; that is, it cannot be translated. Moreover, there is no initiator methionine in the remaining known 5' untranslated sequence. However, we identified cDNAs that differed at the extreme 5' end by the addition of Motif IV and the deletion of Motif V. Removal of the 80 bases within Motif V deletes several termination codons and shifts the reading frame by 1 base. Inclusion of Motif IV in the mRNA introduces a methionine codon surrounded by a strong consensus translation initiation signal (Kozak, 1984); moreover, the combination of a 17-base insert (Motif IV) and an 80-base deletion (Motif V) shifts the reading frame in such a way that, counting from the A of the ATG of the new upstream methionine codon, there are no termination codons encountered between this methionine and the initiator methionine of the 80-kD protein. The open reading frame is thus continuous.

Inclusion of Motif IV and deletion of Motif V, when occurring in concert, have the potential to generate a new protein 4.1 isoform having a higher molecular weight. This additional coding sequence adds 209 amino acids to the amino-terminal end of the prototypical 4.1 (80-kD) molecule. Initiation at the upstream methionine would then generate a larger protein within which the COOH-terminal 80 kD represents the prototypical 4.1 molecule.

cDNAs outlined in Fig. 2 thus establish the structure of several distinctive protein 4.1 mRNAs and predict the existence of many heterogeneous isoforms. Moreover, the identification of these five sequence motifs provide for the structural bases for at least 32 mRNAs, from which 16 distinct protein isoforms can be generated. (Since Motif IV and Motif V muct act in concert to generate additional protein isoforms, there can be more mRNAs than proteins.)

Expression of Multiple Protein 4.1 mRNAs in Erythroid and Nonerythroid Cells

To detect protein 4.1 mRNAs expressed in various tissues and species, we developed nucleic acid probes for mRNAs containing or lacking specific motifs. Poly(A)$^+$ mRNA and total RNA were isolated from human erythroid and nonerythroid cells and analyzed either by Northern blot hybridization or RNase protection analysis (Fig. 4). Northern blots were analyzed with lymphoid 4.1 cDNA probe (pTM-1). RNase protection analysis was performed with synthetic RNA probes containing the appropriate motifs, thereby generating protected fragments of diagnostic size.

Fig. 4 shows the results of RNase protection analysis in erythroid and nonerythroid cells. As indicated by Fig. 4 *A*, the 156-nt (nucleotide) protected fragment indicated the presence of an mRNA containing Motif IV; the 117-nt and 22-nt fragments indicated expression of an mRNA lacking Motif IV. The weak 22-nt band, because of its small size and consequent lower content of radioactive label, was not detected. Fig. 4 *A* also shows that at least two types of protein 4.1 mRNAs (with or without Motif IV) were produced from MOLT 4 and HEL cells. About two thirds of the protein 4.1 mRNA contains Motif IV in MOLT 4 cells. This result is consistent with the finding of two independent cDNA clones, containing (pTM-2) or lacking (pTM-5) Motif IV, which were isolated from the MOLT 4 cDNA library (Fig. 2 *C*). In contrast, the dominant protected fragment (117nt) indicated that the major species of protein 4.1 mRNA produced from human reticulocytes lacks Motif IV (Fig. 4 *A*). However, upon long exposure (5 d; data not shown), a minute amount of reticulocyte RNA containing Motif IV could be detected. This weak protected band (156nt) accounted for <5% of total protein 4.1 mRNA; it may represent residual mRNA from earlier maturation stages or from contaminating RNA from residual nucleated cells in the original reticulocyte preparation. The multiple-banding pattern shown in Fig. 4 *A* could arise from nuclease digestion artifacts, or indicate potential additional motifs in that region.

In contrast to the distribution of Motif IV, mRNA containing Motifs III (the protected fragment is 345 nt) or II (340 nt) represent well over 90% of the mRNA found in both erythroid and nonerythroid cells (Fig. 4, *B* and *C*).

Expression of Protein Isoforms in Nonerythroid Cells

Analysis of nonerythroid protein 4.1 mRNAs (Fig. 4 *A*) suggested that the majority (two thirds) of the mRNA contained Motif IV, and thus possess the potential for translation into higher molecular weight isoforms. To determine whether this mRNA was expressed in the form of higher molecular weight protein, we isolated total protein from MOLT 4 cells. Purified erythroid protein 4.1 and synthetic peptides that were derived from the deduced amino acid sequences of the 30-, (30b), 16-, (16a), and 24-kD (24a) domain were prepared (Fig. 3 *D*). Polyclonal antisera against the erythroid 4.1 and peptides were raised and purified as described in Materials and Methods.

Fig. 5 shows immunoblot analysis of 4.1 proteins from MOLT 4 cells. At least two major protein 4.1 isoforms corresponding to peptide chains of ~135 and 80 kD were identified in MOLT 4 cells. Both proteins strongly react with antibodies against intact erythroid protein 4.1 and synthetic
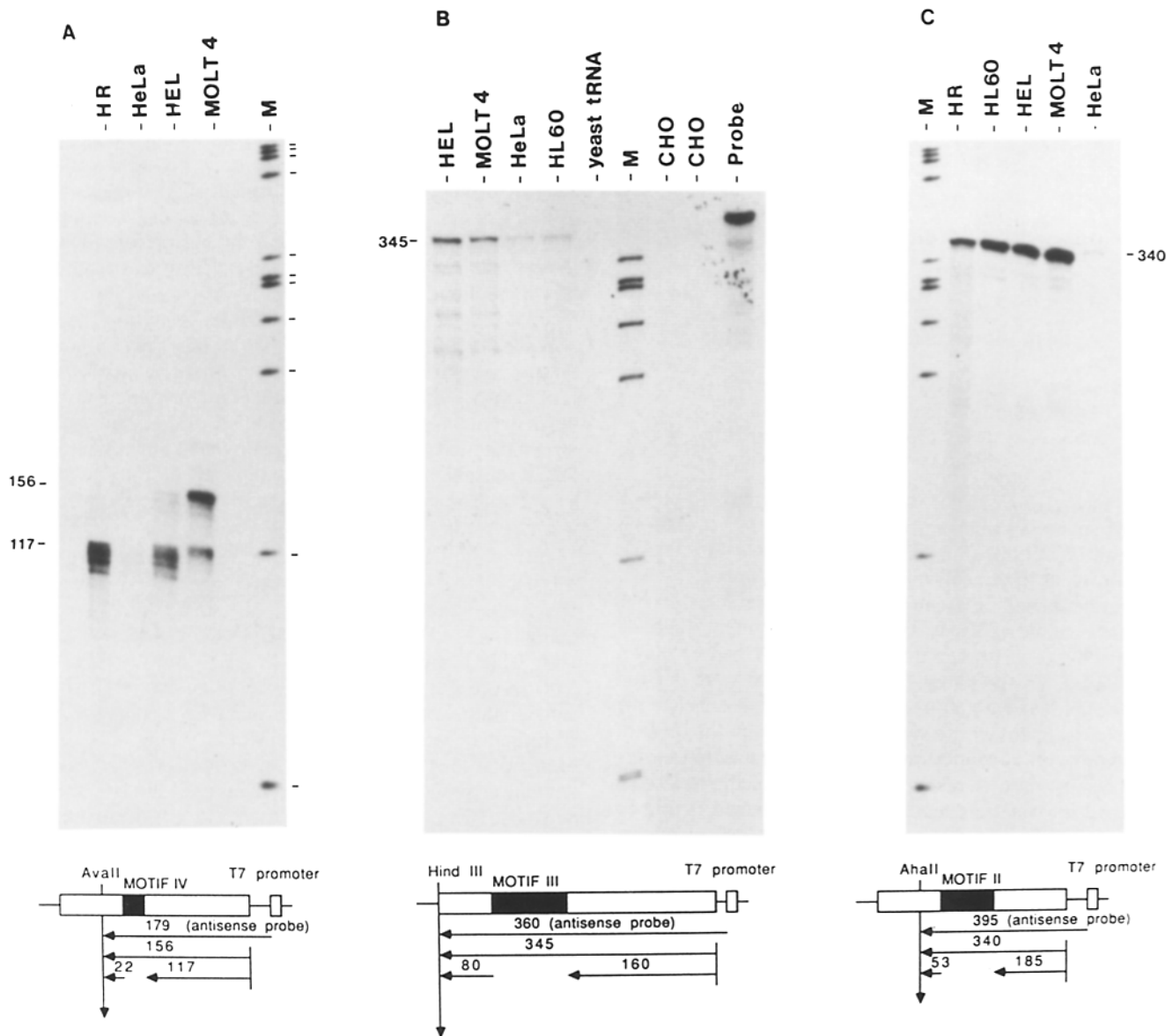
A

HR HeLa HEL MOLT 4 M

156 –
117 –

Avall MOTIF IV T7 promoter
179 (antisense probe)
156
22 117

B

HEL MOLT 4 HeLa HL60 yeast tRNA M CHO CHO Probe

345 –

Hind III MOTIF III T7 promoter
360 (antisense probe)
345
80 160

C

M HR HL60 HEL MOLT 4 HeLa

– 340

Ahall MOTIF II T7 promoter
395 (antisense probe)
340
53 185

Figure 4. Detection of erythroid and nonerythroid protein 4.1 mRNAs by RNase protection assay. Specific antisense probes were designed by constructing sequence-specific cDNA clones that contain specific motifs. The cDNA inserts of these clones were subcloned into a pGEM-4 or -7Z vector, which contains SP6 or T7 RNA polymerase promoters. Antisense RNAs were generated from the T7 promoter within the vector. Total RNAs (30 μg) from different tissues or cell lines were hybridized to the antisense probes that span Motif IV (*A*), Motif III (*B*), or Motif II (*C*). *HR*, human reticulocyte; *HL60*, human promyelocytic leukemia line; *CHO*, Chinese hamster ovary line; *HEL*, human erythroleukemia line; *MOLT 4*, human T cell leukemia line; *HeLa*, human epithelioid carcinoma line. Bars denote φX174 Hae III–digested DNA size markers (*M*). The positions of size markers are indicated in nucleotide.

peptides (30b, 16a, and 24a). Interestingly, higher molecular weight forms were also detected on the blot with anti-16a and -24a antibodies (Fig. 5, lanes *3* and *4*).

In Vitro Translation of Synthetic Full-Length mRNAs Encoding Protein 4.1

As noted earlier, the 80-kD protein 4.1 mRNA has two large ORFs (Fig. 3 *C*) that are not aligned in phase and are disrupted in the 5' untranslated region by several translation stop codons. Insertion of Motif IV and deletion of Motif V in the same mRNA species creates a new ORF continuous and in phase with ORF-2 (ORF-2 generates the 80-kD protein).

The new reading frame has a potential to encode a larger protein species. As shown in Fig. 5, immunochemical analysis demonstrated that cells contain high molecular weight proteins. To verify that the mRNA could be translated into a protein of predicted sequence, i.e., that the upstream initiator methionine functions and that the reading frame is indeed open, we constructed cDNAs containing and lacking the translatable configurations of Motifs IV and V. These cDNAs were then transcribed into synthetic mRNA, and translated in a cell-free system. Fig. 6 shows immunoprecipitation analysis of the translation products with antibodies against the 5' upstream regions. Fig. 7 compares the sizes of in vitro products to protein isolated from intact cells. As
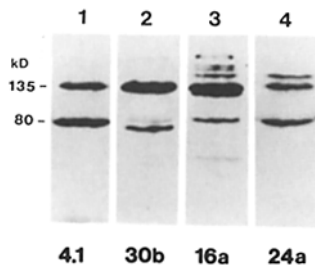
Figure 5. Immunoblot analysis of lymphoid protein 4.1 isoforms. Solubilized proteins from MOLT 4 cells were separated by SDS-PAGE and blotted onto nitrocellulose as described (Tang et al., 1988). Blotted proteins were probed with polyclonal antiintact erythroid 4.1 (lane *1*), anti-30b (lane *2*), -16a (lane *3*), and -24a (lane *4*) antibodies, and detected by ^{125}I-labeled Staphylococcus protein A (Correas et al., 1986*b*). The low molecular weight band has not yet been characterized.

shown in Fig. 6 (lane *8*), the full-length lymphoid cDNA (pLym) spanning all five sequence motifs and containing both translation initiation codons gave rise to a 135- and an 80-kD protein in the cell-free translation system.

To further identify these in vitro translated proteins, an immunoprecipitation assay was performed. A synthetic peptide mixture, 5'-1 (a mixture of three peptides a, b, and ab' shown in Fig. 3 *D*), was prepared. The 5'-1 peptides were derived from the predicted amino-terminal translated sequence of the larger 4.1 isoform, whose synthesis begins with the methionine within the 17-base Motif IV. Polyclonal antisera were raised against these peptides, purified, and characterized as described in Materials and Methods. As shown in Fig. 6 (lanes *1* and *2*), the 135- and 80-kD translated proteins reacted strongly with antiintact erythroid 4.1 antibody. How-

ever, only the 135-kD protein was detected when anti-5'-1 antibody was used as a probe (Fig. 6, lane *4*). Interestingly, the in vitro–translated proteins (135 and 80 kD) comigrate with proteins that were isolated from MOLT 4 cells (Fig. 7, lanes *3* and *4*).

A truncated cDNA (pEry 2) from which the initiator methionine in the 17-base motif had been excluded could be translated only into the 80-kD protein (Fig. 6, lanes *1* and *7*; Fig. 7, lane *2*). When both initiator codons were present, initiation of the upstream site was strongly favored (Fig. 6, lanes 2 and *8*; Fig. 7, lane *3*); however, the downstream initiator methionine is also a potentially strong initiation signal, since it is translated efficiently into the protein in the message lacking Motif IV (Fig. 6, lanes *1* and *7*; Fig. 7, lane 2).

It is interesting to note that the primary sequence of the lymphoid 4.1 clone (pLym) predicts a polypeptide with a calculated molecular mass of 100,534 D. However, the apparent size of the in vitro–translated protein (135 kD) seen on SDS-PAGE was much larger. This discrepancy may be due in part to aberrant migration in SDS-PAGE.
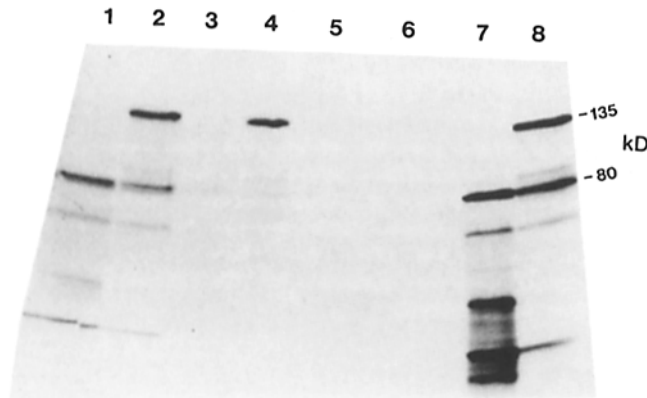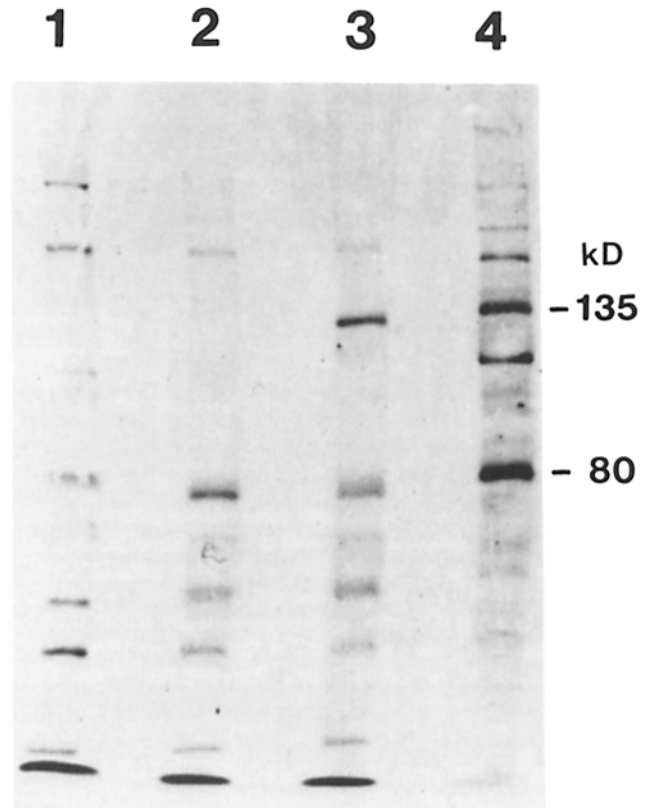

Figure 6. Analysis of synthetic protein 4.1 mRNAs translated in vitro. Two transcription 4.1 clones, a pLym clone that contains the first two translation initiation codons (AUG), and a truncated clone, pEry2, which removed Motifs IV and V, were constructed and subcloned into a pGEM-4 vector (see Materials and Methods). Sense-capped mRNA transcripts were translated in vitro by adding rabbit reticulocyte lysate and ^{35}S-methionine. The labeled translated proteins were either immunoprecipitated by antierythroid 4.1 antibody (lanes *1* and *2*), anti-5'-1 (lanes *3* and *4*), preimmune serum (lanes *5* and *6*), or direct analysis on NaDodSO4/PAGE (lanes *7* and *8*). Lanes *1*, *3*, *5*, and *7* are the proteins translated in vitro from pEry2 clone. Lanes 2, 4, 6, and 8 are translated proteins from pLym clone.

Figure 7. Immunoblot analysis of protein 4.1 isoforms expressed in vitro and in vivo. Plym and pEry2 clones were transcribed and translated in vitro as described in Materials and Methods. The translated proteins were resolved in SDS-PAGE and blotted onto nitrocellulose as described (Tang et al., 1988). Blotted proteins were probed with a mixture of anti-16a and -24a antibodies and detected with a goat anti-rabbit IgG alkaline phosphatase conjugated as described in Promega Biotec manual. (Lane *1*) rabbit reticulocyte lysate; (lane 2) translated proteins from pEry2 clone; (lane *3*) translated proteins from pLym clone; (lane *4*) total solubilized proteins from MOLT 4 cells.

Discussion

Previous studies have provided indirect evidence for the existence of multiple isoforms of protein 4.1 in human (Tang et al., 1988; Conboy et al., 1988) and avian species (Granger and Lazarides, 1984). These forms are immunologically cross-reactive, expressed in many tissues, and heterogeneous with respect to size and relative abundance of size classes. Moreover, localization of protein 4.1 by immunocytochemical methods suggested that these isoforms might be distributed in and around the nucleus (Tang, T. K., E. J. Benz, Jr., and V. T. Marchesi, manuscript in preparation; Leto et al., 1986*b*) and along stress fibers (Cohen et al., 1982), as well as under the cytoplasmic membrane. Our data established the existence of multiple protein 4.1 isoforms in erythroid and nonerythroid tissues, and provide a structural basis for many of these forms. Protein 4.1 appears to be a protein family arising from a multiplicity of messenger RNAs generated by alternative mRNA processing.

A novel result of our studies, not anticipated by previous work, was the existence of a high molecular weight class of protein 4.1 isoforms generated from the same gene by alternate splicing events that convert the 5' untranslated extremity of the mRNA into a translatable reading frame contiguous with the known 80-kD open reading frame of erythroid protein 4.1. As shown in Figs. 6 and 7, we have documented that this potentially usable ORF is in fact translated into protein both in cell-free translation extracts and in intact cells expressing the endogenous gene. The protein can be uniquely identified by an antipeptide antibody that we have generated on the basis of novel amino-terminal peptides translatable into protein only if the appropriate mRNA splicing events occur. The size of the protein isolated from whole cells agrees closely with the size of the primary translation product generated in the cell-free system. It is interesting that this high molecular weight protein is abundant in the nucleus as well as along submembranous sites at intercellular junctions (Tang, T. K., E. J. Benz, Jr., and V. T. Marchesi, manuscript in preparation). We postulate that protein 4.1 isoforms are important for assembly and function of skeletal structures within the nucleus, as well as the cytoplasm.

Protein 4.1 mRNAs capable of encoding both a 135- and 80-kD protein could be translated in vitro (Fig. 6), but we have not yet documented whether both translation initiation sites are used in vivo. It is interesting that the upstream initiator methionine is used almost exclusively when present, but that the downstream initiator works quite efficiently if it is the methionine encountered closest to the 5' end of the mRNA. This naturally occurring bifunctional mRNA should prove useful for studies of factors regulating the positioning of translation initiation.

Our sequence analysis of protein 4.1 cDNAs provides for as many as 32 mRNA species arising from a single mRNA precursor. Only 16 of these mRNAs would be translatable into protein, because only one combination of insertion and/or deletion of Motifs IV and V is compatible with translation of the larger isoform. Our studies of mRNA accumulation in various types of cells suggests that mRNAs capable of encoding the larger isoform comprise nearly two thirds of protein 4.1 mRNAs in MOLT 4 cells (Fig. 4 *A*). mRNAs containing Motif I are expressed predominantly in erythroid tissues (Tang et al., 1988), while mRNAs containing Motif II

(Fig. 4 *C*) and/or Motif III (Fig. 4 *B*) comprise over 90% of the mRNAs we have been able to identify. Recently, Conboy et al. (1988) have identified another motif (43 amino acids) just in front of Motif II in reticulocyte 4.1 sequence. However, this motif is present in all of our erythroid and nonerythroid cDNA clones. These results demonstrate that several protein 4.1 isoforms are actually produced in reasonable abundance in many cell types; however, considerably fewer forms may accumulate than are theoretically possible on the basis of potential alternate mRNA splicing events. In contrast, erythroid cells (human reticulocytes) contain very little protein 4.1 mRNA encoding the larger isoform (Fig. 4 *A*) but do produce isoforms expressing Motif I (Tang et al., 1988). Posttranslational modification of isoforms, such as proteolytic fragmentation, glycosylation (known to occur via cytoplasmic glycosylation enzymes [Holt et al., 1987]), or phosphorylation may need to be invoked to explain all of the forms encountered. It is entirely possible, of course, that tissues or cell types not yet screened by us may express one of the rare forms of protein 4.1 in greater abundance.

Protein 4.1 is a bifunctional protein, attaching to the cytoplasmic domain of transmembrane proteins (glycophorin and band 3) at its amino-terminal end, and to cytoskeletal structures (spectrin/actin complexes) by means of the 10-kD domain. Function appears to be regulated by phosphorylation. It is important to note that these functions, as well as the assignments of functions to specific domains, have been delineated only for erythroid forms of the protein 4.1 in systems using membrane cytoskeletal elements derived from erythroid cells. The function of isoforms expressed in nonerythroid tissues remains an open question. We have recently obtained evidence suggesting that retention of Motif I is important for binding of the protein 4.1 molecule to erythroid spectrin/actin complexes (our unpublished data). Therefore, it is entirely possible that protein 4.1 interacts with distinct components in nonerythroid tissues. This may account for its unexpected tissue distribution. The sequence motifs that we have identified could well contain functional cassettes that mediate novel functions of the protein. These motifs represent logical targets for future in vitro mutagenesis studies designed to delineate the structural basis for the unusual distribution of protein 4.1 isoforms in nonerythroid cells.

References

Aster, J. C., M. J. Welsh, G. J. Brewer, and H. Maisel. 1984. Identification of spectrin and protein 4.1-like proteins in mammalian lens. *Biochem. Biophys. Res. Commun.* 119:726–734.

Baines, A. J., and V. Bennett. 1985. Synapsin I is a spectrin-binding protein immunologically related to erythrocyte protein 4.1. *Nature (Lond.).* 315: 410–413.

Bennett, V. 1985. The membrane skeleton of human erythrocytes and its implications for more complex cells. *Annu. Rev. Biochem.* 54:273–304.

Cohen, C. M., S. F. Foley, and C. Korsgren. 1982. A protein immunologically related to erythrocyte band 4.1 is found on stress fibers of non-erythroid cells. *Nature (Lond.).* 299:648–650.

Conboy, J., Y. W. Kan, S. B. Shohet, and N. Mohandas. 1986*a*. Molecular cloning of protein 4.1, a major structural element of the human erythrocyte

membrane skeleton. *Proc. Natl. Acad. Sci. USA.* 83:9512–9516.

Conboy, J., N. Mohandas, G. Tchernia, and Y. W. Kan. 1986*b*. Molecular basis of hereditary elliptocytosis due to protein 4.1 deficiency. *N. Engl. J. Med.* 315:680–685.

Conboy, J. G., J. Chan, N. Mohandas, and Y. W. Kan. 1988. Multiple protein 4.1 isoforms produced by alternative splicing in human erythroid cells. *Proc. Natl. Acad. Sci. USA.* 85:9062–9065.

Correas, I., T. L. Leto, D. W. Speicher, and V. T. Marchesi. 1986*a*. Identification of the functional site of erythrocyte protein 4.1 involved in spectrin-actin association. *J. Biol. Chem.* 261:3310–3315.

Correas, I., D. W. Speicher, and V. T. Marchesi. 1986*b*. Structure of the spectrin-actin binding sites of erythrocyte protein 4.1. *J. Biol. Chem.* 261:13362–13366.

Davies, G. E., and C. M. Cohen. 1985. Platelets contain proteins immunologically related to red cell spectrin and protein 4.1. *Blood.* 65:52–59.

Davis, L. G., M. D. Dibner, and J. F. Battey. 1986. Basic Methods in Molecular Biology. Elsevier Science Publishing Co. Inc., New York. 302–305.

Goodman, S. R., L. A. Casoria, D. B. Coleman, and I. S. Zagon. 1984. Identification and isolation of brain 4.1. *Science (Wash. DC).* 224:1433–1436.

Granger, B. L., and E. Lazarides. 1984. Membrane skeletal protein 4.1 of avian erythrocytes is composed of multiple variants that exhibit tissue-specific expression. *Cell.* 37:595–607.

Holt, G. D., R. S. Haltiwanger, C.-R. Torres, and G. W. Hart. 1987. Erythrocytes contain cytoplasmic glycoproteins. *J. Biol. Chem.* 262:14847–14850.

Kozak, M. 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* 12:857–

872.

Leto, T. L., I. Correas, T. Tobe, R. A. Anderson, and W. C. Horne. 1986*a*. Structure and function of human erythrocyte cytoskeletal protein 4.1. *In* Membrane Skeletons and Cytoskeletal Membrane Associations. V. Bennett, C. M. Cohen, S. E. Lux, and J. Palek, editors. Alan R. Liss Inc., New York. 201–209.

Leto, T. L., B. M. Pratt, and J. A. Madri. 1986*b*. Mechanisms of cytoskeletal regulation: modulation of aortic endothelial cell protein band 4.1 by the extracellular matrix. *J. Cell. Physiol.* 127:423–431.

Marchesi, V. T. 1985. Stabilizing infrastructure of cell membranes. *Annu. Rev. Cell Biol.* 1:531–561.

Ngai, J., J. H. Stack, R. T. Moon, and E. Lazarides. 1987. Regulated expression of multiple chicken erythroid membrane skeletal protein 4.1 variants is governed by differential RNA processing and translational control. *Proc. Natl. Acad. Sci. USA.* 84:4432–4436.

Spiegel, J. E., D. S. Beardsley, F. S. Southwick, and S. E. Lux. 1984. An analogue of the erythroid membrane skeletal protein 4.1 in nonerythroid cells. *J. Cell Biol.* 99:886–893.

Tang, T. K., T. L. Leto, I. Correas, M. A. Alonso, V. T. Marchesi, and E. J. Benz, Jr. 1988. Selective expression of an erythroid-specific isoform of protein 4.1. *Proc. Natl. Acad. Sci. USA.* 85:3713–3717.

Tchernia, G., N. Mohandas, and S. B. Shohet. 1981. Deficiency of cytoskeletal membrane protein band 4.1 in homozygous hereditary elliptocytosis: implications for erythrocyte membrane stability. *J. Clin. Invest.* 68:454–460.

Zinn, K., D. Dimaio, and T. Maniatis. 1983. Identification of two distinct regulatory regions adjacent to the human β interferon gene. *Cell.* 34:865–879.