Human, Mouse, and Rat Calnexin cDNA Cloning: Identification of Potential Calcium Binding Motifs and Gene Localization to Human Chromosome 5‡

Larry W. Tjoelker,§ Christine E. Seyfried,§ Roger L. Eddy, Jr.,∥ Mary G. Byers,∥ Thomas B. Shows,∥ Jesus Calderon,⊥ Robert B. Schreiber,# and Patrick W. Gray*,§

ICOS Corporation, 22021 20th Avenue South East, Bothell, Washington 98021, Department of Human Genetics, Roswell Park Memorial Institute, 666 Elm Street, Buffalo, New York 14263, Department of Cell Biology, Centro de Investigacion, Mexico City, Mexico DF, and Department of Pathology, Washington University School of Medicine, St. Louis, Missouri 63110

Received August 18, 1993; Revised Manuscript Received November 17, 1993®

ABSTRACT: Calnexin is a 90-kDa integral membrane protein of the endoplasmic reticulum (ER). Calnexin binds Ca^{2+} and may function as a chaperone in the transition of proteins from the ER to the outer cellular membrane. We have purified human calnexin in association with the human interferon-γ receptor and cloned calnexin cDNA from placenta. Fragments of calnexin have been prepared as glutathione *S*-transferase fusion proteins and analyzed for their abilities to bind $^{45}Ca^{2+}$ and ruthenium red. A subdomain containing four internal repeats binds Ca^{2+} with the highest affinity. This sequence is highly conserved when compared to calreticulin (a luminal ER protein), an *Onchocerca* surface antigen, and yeast and plant calnexin homologues. Consequently, this sequence represents a conserved motif for the high-affinity binding of Ca^{2+}, which is clearly distinct from the "E-F hand" motif. An adjacent subdomain, also highly conserved and containing four internal repeats, fails to bind Ca^{2+}. The carboxyl-terminal, cytosolic domain is highly charged and binds Ca^{2+} with moderate affinity, presumably by electrostatic interactions. The calnexin amino-terminal domain (residues 1–253) also binds Ca^{2+}, in contrast to the amino-terminal domain of calreticulin, which is relatively less acidic. We have also determined the cDNA sequences of mouse and rat calnexins. Comparison of the known mammalian calnexin sequences reveals very high conservation of sequence identity (93–98%), suggesting that calnexin performs important cellular functions. The gene for human calnexin is located on the distal end of the long arm of human chromosome 5, at 5q35.

Calcium ions play a central role in the regulation of cellular metabolism, including signal-transduction events and the transport of proteins through the ER[1] (Sambrook, 1990). Ca^{2+} is sequestered in discreet compartments within resting cells, and the ER represents the major store of intracellular Ca^{2+}. Calnexin is a predominant integral membrane protein of the ER and was first identified by its ability to bind Ca^{2+} (Wada et al., 1991a). It has a long amino-terminal domain (460 amino acids) localized in the lumen of the ER, a single hydrophobic transmembrane domain, and a short, acidic cytosolic domain (91 amino acids). While the function of calnexin is not known with certainty, it is believed to serve vital cellular functions since its sequence is highly conserved. Dog and human calnexin share 94% amino acid sequence identity (Wada et al., 1991a; Galvin et al., 1992), and recently a calnexin homologue was isolated from the plant *Arabidopsis* and found to be 48% identical to the canine sequence (Huang et al., 1993). Calnexin has been shown to be associated with several cell surface proteins during translocation through the ER. It has been isolated as a complex with other ER resident proteins that are involved in the Ca^{2+}-dependent retention of proteins (Wada et al., 1991a). Further studies suggest that calnexin may function as a molecular chaperone. Calnexin has been isolated in association with major histocompatibility complex (MHC) class I molecules, the T-cell receptor, and membrane immunoglobulin (Degen & Williams, 1991; Hochstenbach et al., 1992; Ahluwalia et al., 1992). Calnexin may promote the proper assembly of protein complexes during transit through the ER. The retention and folding of such complexes is Ca^{2+} dependent (Sambrook, 1990). Calnexin has also been identified in nuclear membrane preparations (Gilchrist & Pierce, 1993), suggesting that it may also play a role in nuclear membrane trafficking or the regulation of nuclear Ca^{2+} transients.

Calnexin shares significant sequence similarity with calreticulin (Smith & Koch, 1989; Fliegel et al., 1989), the major Ca^{2+} binding protein found in the lumen of the ER. Both calreticulin and calnexin lack the "E-F hand" Ca^{2+} binding motif found in the calmodulin family of Ca^{2+}-regulated proteins (Kretsinger, 1980). However, these molecules contain unique subdomains that can be distinguished by charge, repeating motifs, and predicted secondary structure. Baksh and Michalak (1991) have shown that calreticulin contains a single high-affinity Ca^{2+} binding site, which is localized to the central 134 amino acids of the molecule. They have also determined that the carboxyl-terminal domain contains low-affinity, high-capacity Ca^{2+} binding sites. In order to better understand which regions of the calnexin molecule are responsible for Ca^{2+} binding, we have prepared fusion proteins with calnexin fragments and analyzed their ability to bind $^{45}Ca^{2+}$ and ruthenium red (a dye which stains Ca^{2+} binding proteins; Charuk et al., 1990). In addition, we have purified

‡ The nucleotide sequences reported in this paper have been submitted to the GenBank Data Bank with accession numbers L18887 (human calnexin cDNA), L18888 (mouse calnexin cDNA), and L18889 (rat calnexin cDNA).

* To whom correspondence should be addressed.
§ ICOS Corporation.
∥ Roswell Park Memorial Institute.
⊥ Centro de Investigacion.
Washington University School of Medicine.

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

[1] Abbreviations: bp, base pairs; ER, endoplasmic reticulum; GST, glutathione *S*-transferase; IFN-γ, interferon-γ; MHC, major histocompatibility complex; PCR, polymerase chain reaction; PVDF, poly(vinylidene difluoride); SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis.

0006-2960/94/0433-3229$04.50/0

human calnexin in association with the interferon-γ (IFN-γ) receptor, cloned its cDNA, and determined the chromosomal localization of its gene.

MATERIALS AND METHODS

Materials. Poly(vinylpyrrolidone), Ficoll, bovine serum albumin (fraction V), imidazole, ruthenium red, and glutathione–agarose were purchased from Sigma. $^{45}CaCl_2$ and ^{32}P-labeled nucleotides were obtained from DuPont-New England Nuclear. Restriction endonucleases and DNA modifying enzymes were purchased from Boehringer Mannheim. The plasmid pBluescript SK$^+$ and *Escherichia coli* strain XL-1Blue were from Stratagene. Sequenase (T7 DNA polymerase) utilized for sequencing was from U.S. Biochemical Corp.

Purification of Human Calnexin from Placenta. Human placental membranes were prepared for the isolation of IFN-γ receptor as previously described (Calderon et al., 1988). Membrane preparations were chromatographed on an IFN-γ receptor-specific monoclonal antibody resin (Sepharose containing covalently coupled GIR-301). The receptor preparation was further purified on wheat germ agglutinin–Sepharose. The IFN-γ receptor bound avidly to the lectin column, while a protein of apparent molecular mass 90 kDa eluted in the column flow-through fractions. This protein failed to react on Western blots with monoclonal antibodies specific for the IFN-γ receptor. The 90-kDa protein was electrophoresed on SDS–PAGE, transferred to poly(vinylidene difluoride) (PVDF) membrane, and subjected to microsequence analysis. Twenty residues were determined from the amino-terminal end by microsequencing techniques: H D G H D D D V I D I E D T L D D V I K.

Calnexin cDNA Cloning. The determined amino-terminal sequence was utilized for the design of a synthetic oligonucleotide probe: 5′-CATGATGGCCATGATGATGATGTGATTGACATTGAGGACACCCTGGATGATGTGATCAAG. In the design of this sequence, the third position of each codon was guessed, using codon choice tables as a guide (Wada et al., 1991b). The synthetic DNA was end-labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase and used to screen a placental cDNA library in λgt10 (Maniatis et al., 1978). Approximately 600 000 phage were screened on 12 nitrocellulose filters in a solution of 20% formamide, 0.75 M sodium chloride, 0.075 M sodium citrate, 0.05 M sodium phosphate (pH 6.5), 1% poly(vinylpyrrolidine), 1% Ficoll, 1% bovine serum albumin, and 10 μg/mL sonicated salmon sperm DNA. Following overnight hybridization at 42 °C, the filters were washed extensively in 0.03 M sodium chloride, 3 mM sodium citrate, and 0.1% SDS at 42 °C. Twelve duplicate hybridizing clones were chosen for plaque purification. DNA was prepared from each plaque, and three independent cDNA inserts were subcloned into pBluescript for sequencing by the dideoxy chain termination method (Smith, 1980). The encoded protein sequence differed from that obtained by amino-terminal sequencing by two residues. Homology comparisons of the encoded protein sequence with the canine calnexin sequence (Wada et al., 1991a) suggested that this cDNA encoded a human homologue of calnexin.

The polymerase chain reaction (PCR; Saiki et al., 1988) was used to isolate mouse and rat calnexin coding sequences. Template cDNA was prepared from the mouse LTK$^-$ cell line and from rat activated T-lymphocytes using reverse transcriptase (Copy Kit, Invitrogen). Synthetic oligonucleotide PCR primers were based on the sequence of the human calnexin gene and designed to amplify the entire coding region of the rodent calnexin genes by anchoring the 3′ termini of the sense (5′-AAGAAGAAAAGCGGCCGCCGGGCAGGCTAGAGATCATG) and antisense (5′-GTCGTCTAGATCACAGATCAAGCTCTAAGATTGTTTCA) primers at the translation initiation and termination codons, respectively. *Not*I and *Xba*I restriction sites were designed into the primers to facilitate cloning of the PCR products. PCR reaction mixtures contained 100 ng of template cDNA, 1 μg of each primer, 0.125 mM of each dNTP, 10 mM Tris-HCl, pH 8.4, 50 mM KCl, 1.5 mM $MgCl_2$, and 2.5 units of Taq polymerase. An initial template denaturation step of 94 °C for 4 min was followed by 30 cycles of amplification: 1 min at 94 °C, 2 min at 50 °C, and 4 min at 72 °C. PCR products were extracted with phenol and chloroform, digested with *Not*I and *Xba*I, purified by agarose gel electrophoresis, and ligated into pBluescript SK$^+$ (Stratagene) for nucleotide sequencing. Three distinct clones from each PCR amplification were sequenced to assure the absence of Taq polymerase errors.

Calcium Binding Studies. Fusion proteins of human calnexin domains were prepared with glutathione *S*-transferase (GST) (Smith & Johnson, 1988). GST fusion proteins are generally expressed well in *E. coli* and can be easily purified. The gene fusions were prepared with the plasmid pGEX-2T, which directs the expression of *Shistosoma japonicum* GST (Smith & Johnson, 1988). The carboxyl-terminal coding sequence contains *Bam*HI and *Eco*RI sites for cloning and encodes a thrombin cleavage site. Synthetic oligonucleotide primers were designed for PCR amplification of specific calnexin domains. All upstream primers contained an in-frame *Bam*HI site, while the downstream primers contained a stop codon and an *Eco*RI site. As diagrammed in Figure 2A, the primer sequences were as follows:

1-CGTGGATCCCATGATGGACATGATGATGA

254-GTTGGATCCCGTGAAATTGAGGACCCAG

320-CCTGGATCCGAAGACATGGATGGAGAA

482-CCTGGATCCTGCTGTTCTGGAAAGAAA

253-CGGAATTCTATGAAGGATTTACAGGAGG

334-GATGAATTCACTGAGGAGCCTCCCATTC

392-GGTGAATTCATTCCAGATCTTCAAAGAA

460-CGAGAATTCAGCGCTCTTCAGCTGCCTC

572-GGAGAATTCACTCTCTTCGTGGCTTTCTG

Each DNA segment was amplified with a cloned human calnexin cDNA template using the previously described PCR conditions. The PCR reaction mixture was then extracted with phenol and chloroform, ethanol precipitated, and digested with *Bam*HI and *Eco*RI. Each calnexin coding domain was purified by agarose gel electrophoresis, and a band of the predicted size was observed for each reaction. The DNA fragments were individually ligated into pGEX-2T vector cleaved with *Bam*HI and *Eco*RI. The ligation mixture was transformed into *E. coli* XL-1Blue and selected on ampicillin plates. Individual colonies were chosen for the preparation of plasmid DNA. Sequences of plasmid inserts were determined by dideoxy sequencing. Each fusion protein was prepared by growing 0.5-l cultures of transformed bacteria and inducing expression with isopropyl β-D-1-thiogalacto-

Table 1: Segregation of Calnexin with Human Chromosomes in *Eco*RI-Digested Human–Mouse Cell Hybrid DNA[a]

hybrid	DNA no.	calnexin	human chromosomes 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	translocations
ATR-13	660	+	+	+	+	+	+	+	+	+	+	+	–	+	+	+	+	+	+	+	+	–	–	+	t	5/X
DUA-3BSAGA	233	–	–	+	–	–	–	–	+	+	–	–	–	–	+	+	–	–	+	–	–	–	–	–	–	
JSR-2	389	–	–	–	+	+	–	–	+	–	–	–	–	–	+	+	–	–	–	–	–	–	–	–	+	
JSR-14	402	+	–	+	+	+	+	+	–	–	–	–	–	+	+	–	–	–	+	–	–	+	+	–	+	
JSR-17S	44	+	+	+	+	–	+	–	t	+	+	+	+	+	+	+	+	+	+	+	–	+	+	+	–	7/9
JWR-22H	653	–	t	t	+	+	–	+	–	+	–	+	+	+	+	+	+	–	+	+	–	+	+	–	–	2/1
JWR-26C	187	+	t	+	+	+	+	+	+	–	+	+	+	+	–	+	+	+	+	+	–	+	+	–	+	1/2
NSL-16	192	+	–	–	+	+	+	–	+	+	t	+	–	+	–	+	+	+	+	+	–	+	+	–	–	17/9
REW-7	16	+	+	+	+	+	+	+	+	+	–	+	+	+	+	+	+	–	+	+	+	+	+	+	+	
REW-8	195	–	–	–	–	+	–	–	–	–	+	+	–	–	–	–	–	–	+	–	–	+	+	+	+	
REX-11BSAgB	184	–	–	–	+	–	–	–	–	–	–	+	–	–	–	+	+	–	–	+	–	–	–	–	–	
REX-11BSHF	254	–	–	–	+	–	–	–	–	–	–	+	–	–	–	+	–	–	–	+	–	–	–	t	t	22/X
REX-57BSHB	350	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	+	–	–	–	+	t	t	22/X
SIR-8	673	+	+	+	+	+	+	–	+	+	+	+	+	+	+	+	+	+	+	+	+	–	+	+	+	
SIR-11	390	–	–	–	–	–	–	–	+	–	–	–	–	–	+	–	–	–	–	–	–	–	+	+	+	
TSL-1	643	–	–	+	+	+	–	–	–	–	–	+	+	–	+	–	–	+	+	+	–	+	+	–	–	
TSL-2	644	+	–	+	t	–	+	+	–	+	–	+	–	+	–	+	+	–	t	+	–	+	+	–	+	17/3
VTL-6	395	–	–	+	–	–	–	+	+	+	–	+	+	–	–	–	+	–	+	–	+	+	+	+	–	
VTL-7	419	–	–	–	–	–	–	–	t	–	–	–	+	–	–	+	–	–	+	–	+	+	+	+	–	7q-
VTL-8	387	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	+	–	+	–	–	+	+	+	–	
WIL-1	20	–	–	–	–	–	–	–	–	+	–	–	–	+	–	+	–	–	+	+	–	–	+	–	+	
WIL-2	12	–	–	–	–	–	–	–	–	+	–	–	–	+	–	–	+	–	+	–	–	–	+	–	+	
WIL-5	9	–	–	–	–	+	–	–	–	+	–	+	–	–	–	–	–	–	+	+	–	–	+	–	+	
WIL-6	425	+	–	+	–	+	+	+	+	+	–	+	+	–	–	+	–	–	+	–	+	+	+	–	+	
WIL-8X	424	+	–	–	+	+	+	–	+	+	–	+	+	+	–	+	–	–	+	+	+	+	+	–	+	
WIL-15	25	–	–	+	+	+	–	+	+	–	–	+	+	+	+	+	+	–	+	+	–	+	+	–	+	
XER-7	640	+	+	+	+	+	+	+	+	+	+	+	t	+	+	+	+	–	–	+	+	–	+	–	+	11/X
XOL-6	534	+	t	–	–	–	+	+	+	–	–	+	+	+	–	+	–	–	+	–	+	+	–	+	t	1/X
XOL-9	554	–	t	+	+	+	–	+	–	–	–	–	–	+	–	–	+	–	+	+	+	–	+	+	+	X/1
XTR-2	332	+	–	–	t	–	+	–	–	+	–	+	–	+	+	+	–	–	–	+	–	+	+	–	t	3/X
XTR-3BSAgB	57	–	–	–	t	–	–	–	–	–	+	t	–	+	–	–	–	–	–	–	–	+	+	–	t	3/X, 10q-
concordant no. of hybrids			21	21	19	20	31	22	21	22	21	21	20	24	18	21	19	22	15	20	22	20	15	15	15	
discordant no. of hybrids			6	9	9	11	0	9	8	9	9	9	10	7	13	10	12	9	15	11	9	11	16	14	10	
discordancy (%)			22	30	32	35	0	29	28	29	30	30	33	23	42	32	39	29	50	35	29	35	52	48	40	

[a] This data is compiled from 31 cell hybrids involving 13 unrelated human cell lines and four mouse cell lines (Shows et al., 1982, 1984). The hybrids were characterized by karyotypic analysis and by mapped enzyme markers (Shows et al., 1982; Shows, 1983). A "t" indicates a chromosome translocation for a particular chromosome, but no intact chromosome is present. The DNA probe for human calnexin was hybridized to Southern blots containing *Eco*RI-digested DNA from the human–mouse hybrids listed in the table. The scoring for the probe was determined by the presence (+) or absence (–) of human bands in the hybrids. Concordant hybrids have either retained or lost the human bands together with a specific human chromosome. Discordant hybrids have either retained the human bands, but not a specific chromosome, or the reverse. Percent discordance indicates the degree of discordant segregation for a marker and a chromosome.

pyranoside, followed by purification on glutathione–agarose as described (Smith & Johnson, 1988).

The GST–calnexin fusion proteins were analyzed for Ca^{2+} binding by the $^{45}Ca^{2+}$ overlay technique of Maruyama et al. (1984) and the ruthenium red overlay method of Charuk et al. (1990). Samples were reduced, resolved on 10% SDS–polyacrylamide gels, and transferred electrophoretically to nitrocellulose (100 V for 1 h). After transfer, the blot was washed for 1 h in 60 mM KCl, 5 mM $MgCl_2$, and 10 mM Tris-HCl, pH 7.5. The blot was then stained in the same solution containing 0.01 mM ruthenium red for 20 min. To reduce nonspecific staining, the blot was again washed in the above solution without ruthenium red for 20 min. Binding of $^{45}Ca^{2+}$ was observed by spotting GST–calnexin fusion proteins directly onto a PVDF membrane. The membrane was then washed two times in 10 mM imidazole, pH 6.8, 60 mM KCl, and 5 mM $MgCl_2$, and then $^{45}Ca^{2+}$ was added at a final concentration of 2 M for 20 min at room temperature. After being washed twice with 50% ethanol, the blot was air dried and exposed to X-ray film. Following $^{45}Ca^{2+}$ and ruthenium red analyses, blots were stained with 0.05% coomassie blue in 50% methanol and 7% acetic acid for 10 min. Blots were then destained in 50% methanol and 7% acetic acid.

Chromosomal Localization of the Human Calnexin Gene. Chromosome mapping was performed by somatic cell Southern blot hybridization and *in situ* hybridization. The human–murine somatic cell hybrids used in this study have been previously described (Shows et al., 1982; Shows, 1983). Hybrid cell lines were characterized for their chromosome contents by direct karyotyping and by assaying marker enzymes specific for each chromosome (Shows et al., 1982). The calnexin cDNA probe was hybridized to Southern blots containing *Eco*RI-digested DNA from the human–mouse hybrids listed in Table 1. Using conditions of high stringency, a human hybridization signal was observed at 7.4 kbp and a mouse signal was observed at 12.2 kbp. For more specific gene localization, *in situ* hybridization was performed on human metaphase chromosomes with a ^{3}H-labeled cDNA probe (Nakai et al., 1986; Zabel et al., 1983).

RESULTS

Our laboratories have had a long-standing interest in the characterization of IFN-γ and its receptor. In a previous study, the IFN-γ receptor was purified from human placental membranes by affinity chromatography over columns containing receptor-specific monoclonal antibodies and wheat germ agglutinin (Calderon et al., 1988). The IFN-γ receptor bound tightly to both matrices and has an apparent size of 90 kDa on SDS–PAGE. A unique protein was identified that also had an apparent size of 90 kDa and co-eluted with the IFN-γ receptor on the monoclonal antibody column. However, this protein did not bind to wheat germ agglutinin, and

```
                -20                 +1
Human   MEGKWLLCMLLVLGTAIVEAHDGHDDDVIDIEDDLDDVIEEVEDSKPDT-TAPPSSPKVTYKAPVPTGEVYFADSFDRGTLSGWILSKAKKDDTDDEIAKYDG   82
Mouse   ........L.......A..........A.....................SKSDAST.P.....................S.....................   83
Rat     ........L.......AIQ........M.....................SKSD.ST.P.....................S.....................   83
Dog     ..............T..Q..E.....M.....................SKPD.SA.T...........S.............................   83

Human   KWEVEEMKESKLPGDKGLVLMSRAKHHAISAKLNKPFLFDTKPLIVQYEVNFQNGIECGGAYVKLLSKTPELNLDQFHDKTPYTIMFGPDKCGEDYKLHFIFR  185
Mouse   ....D....T.....................................................A..S.............................  186
Rat     ....D....T.....................................................S.............................  186
Dog     ....D....T.....................................................................................  186

Human   HKNPKTGIYEEKHAKRPDADLKTYFTDKKTHLYTLILNPDNSFEILVDQSVVNSGNLLNDMTPPVNPSREIEDPEDRKPEDWDERPKIPDPEAVKPDDWDEDA  288
Mouse   .......V.....................................................................A..D...........  289
Rat     .......V.....................................................................A..D...........  289
Dog     .......V.....................I.....................Q.............D.......N...  289

Human   PAKIPDEEATKPEGWLDDEPEYVPDPDAEKPEDWDEDMDGEWEAPQIANPRCESAPGCGVWQRPVIDNPNYKGKWKPPMIDNPSYQGIWKPRKIPNPDFFEDL  391
Mouse   .S.....................I.....................K.............M.....................N...............  392
Rat     .S.....................I.....................K.............M.....................N...............  392
Dog     ...........D.....................K.............M.....................N...............  392

Human   EPFRMTPFSAIGLELWSMTSDIFFDNFIICADRRIVDDWANDGWGLKKAADGAAEPGVVGQMIEAAEERPWLWVVYILTVALPVFLVILFCCSGKKQTSGMEY  494
Mouse   ...K.....................SG...V.....................L..L.....................SNA...  495
Rat     .....................SG...V.....................L.....................SNA...  495
Dog     ...K.....................V.G...V.....................V.....................S.PV..  495

Human   KKTDAPQPDVKEEEEEKEEEKDKGDEEEEGEEKLEEKQKSDAEEDGGTVSQEEEDRKPKAEEDEILNRSPRNRKPRRE  572
Mouse   ...........-D..G.....N.R.-...E.............V.G..D...S.....................  571
Rat     ...........-D..G.....N...-...E.............G..D...S.....................  571
Dog     .....................................A....D.....................  573
```

FIGURE 1: Comparison of calnexin across four mammalian species. The amino-terminal end begins at residue +1; residues −1 to −20 represent the calnexin signal sequence. Repeat sequences 1 and 2 are delineated by heavy and light overlines, respectively. The predicted transmembrane region is underlined. Identity to the human sequence is indicated by periods. Dashes denote spaces introduced to optimize the alignment. The canine sequence is from Wada et al. (1991a).

it failed to react with IFN-γ receptor antibodies (results not shown). In addition, the sequence of the 20 amino-terminal residues was determined by Edman degradation and is clearly distinct from that of the IFN-γ receptor. This amino acid sequence was used to design a 60-base synthetic oligonucleotide probe, which was end-labeled with ^{32}P and used to screen a human placental cDNA library in λgt10. Twelve different hybridizing signals were plaque purified, and the DNA sequence of the longest insert was determined. This sequence was identified as the human homologue of the canine Ca^{2+} binding protein calnexin (Wada et al., 1991a). Galvin et al. (1992) have previously identified a partial cDNA clone for human calnexin. A full-length human sequence has recently been deposited in GenBank (B. Honore, H. Leffers, P. Madsen, and J. E. Celis, unpublished) which contains a single codon change compared to our sequence. As shown in Figure 1, the human and canine calnexin encoded sequences are 94% similar.

In order to extend the analysis of evolutionary conservation across mammalian species, mouse and rat calnexin cDNAs were also characterized. Synthetic oligonucleotides based on the human sequence were designed for PCR amplification of rodent cDNA. The upstream primer contained 5′ untranslated sequence and ended in the ATG initiation codon. The downstream primer contained 3′ untranslated sequence (non-coding strand) and the stop codon. These primers probably contain mismatches to the rodent cDNAs but nevertheless have enough homology to amplify specifically the calnexin sequences; anchoring the primers at the start and stop codons ensures that the 3′ ends of the primers will anneal and allow amplification of the entire coding region. PCR was performed on cDNA prepared from the mouse LTK$^-$ cell line and activated rat T-lymphocytes. Each reaction produced DNA of approximately 1750 bp, the expected size of the calnexin coding sequence. The PCR-generated cDNAs were cloned, and their predicted amino acid sequences are presented in Figure 1. The mouse and rat sequences exhibit the greatest similarity to each other (98% identical amino acids). Comparison of either rodent sequence with human or canine calnexin reveals 93-94% amino acid identity. Most of the differences in the comparisons are due to conservative amino acid substitutions. These results demonstrate that calnexin is a highly conserved sequence among mammalian species.

Ca^{2+} Binding Domains of Human Calnexin. Calnexin and the structurally similar calreticulin have previously been shown to bind Ca^{2+} (Wada et al., 1991a; Smith & Koch, 1989; Fliegel et al., 1989). Both molecules contain subdomains that can be defined by charge, repeating motifs, or hydrophobicity. The first half of calnexin's luminal portion (representing the amino-terminal end) is somewhat acidic and contains a predicted β-sheet structure. The second half of the luminal portion contains two sets of repeating structures, both of which are rich in proline and tryptophan. The cytoplasmic domain is very acidic with a predicted helical conformation. Calnexin and calreticulin lack the E-F hand motif believed to be responsible for binding Ca^{2+} in the calmodulin family of Ca^{2+} binding proteins. Therefore, it was of interest to determine which subdomains of the calnexin molecule mediate Ca^{2+} binding.

The various subdomains of calnexin were engineered as fusion proteins with glutathione *S*-transferase (GST). The design of these fusion proteins is presented in Figure 2A. The amino terminus of each fusion protein contains GST, which is highly expressed in *E. coli* and enables purification on glutathione–agarose (Smith & Johnson, 1988). All of the GST–calnexin fusion proteins were expressed as soluble proteins in *E. coli*. Each of the purified fusion proteins was tested for Ca^{2+} binding. The purified fusion proteins were electrophoresed on SDS–PAGE, electrotransferred to a nitrocellulose membrane, and then analyzed with a ruthenium red overlay procedure (Charuk et al., 1990). Ruthenium red has previously been shown to specifically bind to Ca^{2+} binding proteins. Most of the fusion proteins run aberrantly slowly on SDS–PAGE, presumably due to the acidic nature of the calnexin sequence; natural calnexin also exhibits reduced mobility on SDS–PAGE (Wada et al., 1991a). As shown in Figure 2B, GST alone fails to bind ruthenium red. The GST fusion protein containing the entire extraluminal domain (residues 1–460) binds ruthenium red, as does the cytosolic domain.

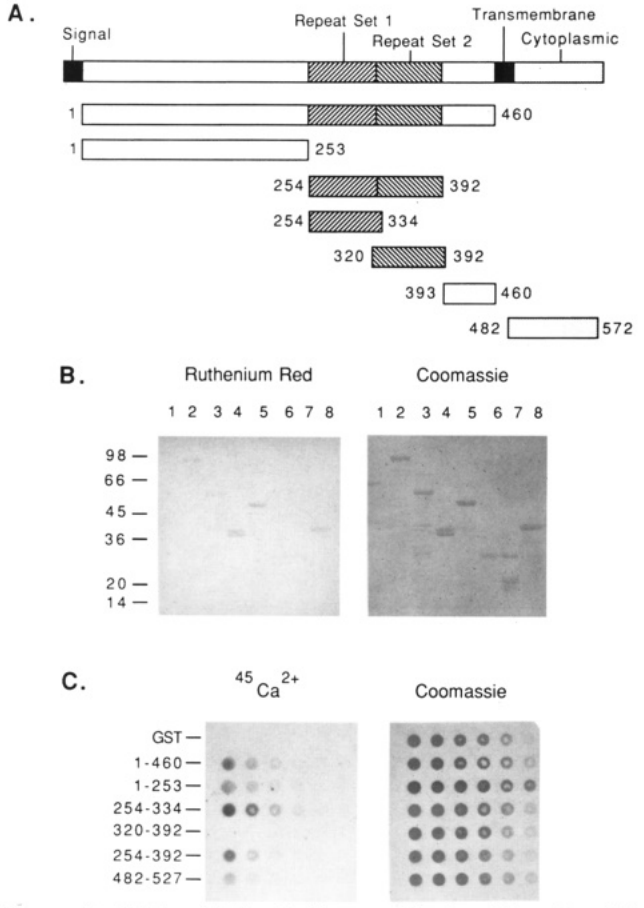

FIGURE 2: Calcium binding by human calnexin subdomains. (A) Schematic illustration of the individual regions of calnexin which were tested for the capacity to bind Ca^{2+}. Constructs were assembled for expression in *E. coli* as fusions with GST at the amino-terminal end and calnexin domains at the carboxyl-terminal end. Numbers identify the terminal residues of each subdomain as defined in Figure 1. (B) Nitrocellulose membrane containing purified fusion proteins with calnexin subdomains resolved by SDS–PAGE. Membrane was first treated with ruthenium red (left) and then stained with Coomasie blue (right). Lane 1 GST alone; lane 2, GST/1–460; lane 3, GST/1–253; lane 4, GST/254–334; lane 5, GST/320–392; lane 6, GST/254–392; lane 7, GST/393–460; lane 8, GST/482–572. (C) PVDF membranes containing calnexin domains were treated with $^{45}Ca^{2+}$ (left panel) followed by staining with Coomasie blue (right panel). From left to right, each spot contains approximately 25, 12.5, 6.3, 3.2, 1.5, or 0.75 μg of each GST fusion protein.

The fusion proteins with calnexin subdomains further localize the position of Ca^{2+} binding sites. The first half of the amino-terminal domain of calnexin (residues 1–253) appears to bind Ca^{2+}, perhaps because of the large number of acidic residues. The fusion proteins containing both sets of repeats (254–392) or just the first set of repeats (254–334) both bound ruthenium red effectively. Thus, this fusion protein containing only 81 residues of the 90 kDa calnexin binds Ca^{2+} as well as the fusion protein containing the entire extraluminal domain. This sequence therefore defines a unique Ca^{2+} binding motif, which is conserved in calreticulin (Smith & Koch, 1989; Fliegel et al., 1989), the *Onchocerca* surface antigen (Unnasch et al., 1988), and a yeast homologue of calnexin (C. DeVirgilio, M. Buerckert, T. Boller, and A. Wiemken, unpublished: GenBank X66470). The 254–334 fusion protein is isolated as a doublet, and both bands bind ruthenium red as shown in Figure 2B. GST can be released from both fusion proteins by thrombin cleavage (results not shown), suggesting that the smaller fusion protein contains 15 fewer amino acids at the carboxyl-terminal end. The second set of repeats do not bind ruthenium red, as shown by fusion protein 320–392.

In order to further compare the Ca^{2+} binding abilities of the fusion proteins, we prepared a dot blot of the fusion proteins and treated it with a $^{45}Ca^{2+}$ overlay procedure (Maruyama et al., 1984). Approximately 25 μg of each of six fusion proteins was spotted on PVDF membrane, and a series of 2-fold dilutions were made as shown in Figure 2C. The results are consistent with the SDS–PAGE transfer presented in Figure 2B and demonstrate in a more quantitative fashion that the first set of repeats avidly bind $^{45}Ca^{2+}$. The second set of repeats, however, fail to bind any $^{45}Ca^{2+}$. The amino-terminal sequence (1–253) appears to bind $^{45}Ca^{2+}$; in contrast, Baksh and Michalak (1991) observed no Ca^{2+} binding to the homologous region in calreticulin. This may be a reflection of the higher stringency required by their method of equilibrium dialysis or of structural differences in this region (see Discussion). The binding that we have observed is most likely due to weak electrostatic interactions due to the large number of acidic residues in the amino-terminal region.

Chromosomal Localization of the Human Calnexin Gene. Genes that are structurally or functionally related to each other are often encoded in the same region of the genome. We therefore mapped the chromosomal position of human calnexin to determine if any related genes are similarly located. This localization was performed by both Southern blot analysis of human–murine somatic cell hybrids (Shows, 1983) and *in situ* hybridization of metaphase chromosomes. The somatic cell hybrid data are presented in Table 1. DNA was isolated from human–mouse somatic cell hybrids, digested with *Eco*RI, and hybridized on Southern blots with the human calnexin cDNA probe. Under stringent hybridization conditions, the human gene hybridized as a 7.4-kbp band, while the mouse gene hybridized as a 12.2-kbp band. Hybridization of the calnexin gene consistently segregated with human chromosome 5. There were no discordant results in 31 different cell lines tested.

To localize the position of the calnexin gene more specifically, *in situ* hybridization was performed on human metaphase chromosomes (Zabel et al., 1983; Nakai et al., 1986). As shown in Figure 3, the *in situ* hybridization results confirm the somatic cell data since the majority of hybridization is to chromosome 5 (of 106 metaphase chromosomes analyzed, 156 grains were observed and 46 were found on chromosome 5). These results demonstrate that the calnexin gene is located on the distal end of human chromosome 5, at 5q35.

DISCUSSION

The data presented in this paper further characterize calnexin and its coding sequence. The four mammalian calnexin sequences that have been determined—canine (Wada et al., 1991a), human, rat, and mouse—are very highly conserved. They share 93–98% identity, suggesting that calnexin serves a key role in cellular functions. While this role has not been fully characterized, structural considerations and localization studies provide clues. Calnexin was first identified as a Ca^{2+} binding protein of the ER (Wada et al., 1991a). Calnexin has also been isolated in association with MHC class I molecules and immunoglobulin heavy chains (Galvin et al., 1992; Ahluwalia et al., 1992) and is thought to aid the assembly of these membrane complexes. Thus, calnexin has been proposed to be a molecular chaperone, aiding the transfer of cell surface molecules in transit from the ER to the outer cellular membrane.

Consistent with this hypothesis is our observed copurification of calnexin with the IFN-γ receptor. This association may

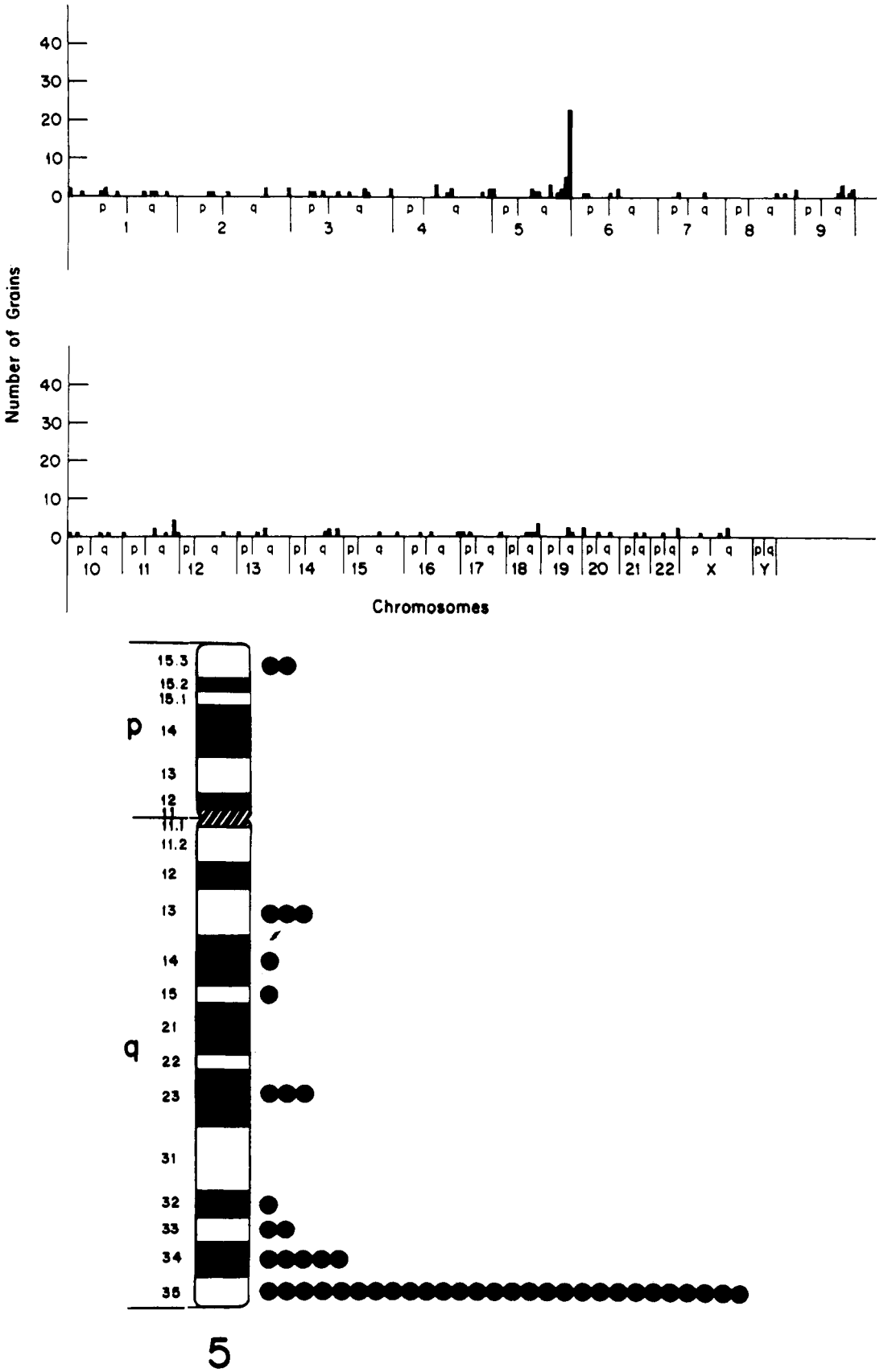

FIGURE 3: Results of *in situ* hybridization to human metaphase chromosomes with the human calnexin cDNA probe. Hybridization is presented as the number of grains observed on each chromosome. The chromosome 5 summary is presented below.

be fortuitous due to similar purification characteristics, but calnexin and the IFN-γ receptor share no structural similarity and exhibit distinct amino acid compositions, pI's, and predicted secondary structures. The two molecules are unlikely to copurify, particularly on an anti-IFN-γ receptor affinity resin, unless there is a significant association between the molecules. Calnexin may aid the formation of an IFN-γ receptor complex. IFN-γ signal transduction is known to

	A	B
	* * ** *	* * **
Consensus	PXXIPDPXAXKPEDWDE	GXWXPPXIXNPXYX
Human Calnexin	SRE.E..EDR.....	.E.EA.Q.A..RCE
	RPK....E.V..D....	.V.QR.V.D..N.K
	.AK...EE.T...G.LD	.K.K..M.D..S.Q
	.EYV...D.E.....	.I.K.RK.P..DFF
A. thaliana CNX1p	AKT....EDK.....	.M.EA.K.D..KCE
	RAK....N.V.....	.E.KR.MKR..A.K
	.ME.E.EE.E...G.LD	.K.SS.L.D..A.K
	.EEVD..E.T.....D	.I.K.RD.P..D.F
S. mansoni SmIrV1	.KE.D..EDK..S....	.E.VA.Q.N..KCA
	REK.V.TN.K..D....	.K.HR.I.P..K.K
	.AT.E.ES.V..SG.LD	.K.SA.M.P..N.K
	.EM....A.VP.K...R	.I.T.RK.P..H.F
Murine Calreticulin	.KK.K..D.A.....	.E.E..V.Q..E.K
	RAK.D..TDS.....K	.E.K.RQ.D..D.K
	.EH....D.K.....	.T.IH.E.D..E.S
O. volvulus λRal-1	.KK.K..D.K.....	.E.E..MVD..E.K
	REF.D.EDDK.....K	.E.K.KQKK..A.K
	.EH....D.K.....D	.K.IH.E.EI.D.T
S. cerevisiae CNX Hom.	.LM...VSVA..H...D	.E.I..M.K..LCT
	RIR....E.V.LS.R..	.QQI.GL.N.AK.K
	.EY.L..N.Q..SW.K.	.E.H..E.E..L.Y

FIGURE 4: Amino acid sequence alignment of the repeated motifs characteristic of human calnexin and related molecules. Column A depicts repeat sequence 1, which binds Ca^{2+} (calnexin residues 253–324). Column B represents repeat sequence 2, which does not bind Ca^{2+} (calnexin residues 328–388). Repeat sequences are listed in the order in which they appear from amino to carboxyl termini of the proteins. Consensus residues are defined as those present in more than 50% of the aligned sequences. Residues appearing in more than 90% of the sequences are indicated (*). Periods represent identity to the consensus residues.

require multiple components [Farrar et al., 1991; Gibbs et al., 1991; reviewed in Farrar and Schreiber (1993)], including a chromosome 21 gene product (Jung et al., 1987). This association may be initiated in the ER, similar to MHC and T-cell receptor complexes.

Calnexin clearly binds Ca^{2+} with high affinity and along with calreticulin and one other protein (SSR) represents the major Ca^{2+} binding protein of the ER (Wada et al., 1991a; Ou et al., 1992). In order to analyze the Ca^{2+} binding properties of calnexin more fully, we have prepared fusion proteins with subdomains of the molecule. The data demonstrate that the structures responsible for Ca^{2+} binding can be extracted from the intact calnexin molecule and fused to GST to produce chimeric molecules which bind Ca^{2+}. The first half of the luminal domain (residues 1–253) binds Ca^{2+}. This result is in contrast to the lack of Ca^{2+} binding observed for the homologous region of rabbit calreticulin by Baksh and Michalak (1991). This discrepancy may be due to the higher number and higher ratio of acidic amino acids found in the calnexin domain (49 acidic and 32 basic residues) compared with calreticulin (30 acidic and 23 basic residues). Adjacent to the amino-terminal domain are two sets of repeated structures. The first set, about 80 residues in length, binds Ca^{2+} with high affinity and is made up of four conserved repeats, shown in Figure 4. This sequence is distinct from the E-F hand Ca^{2+} binding sequence (Kretsinger, 1980) and therefore defines a unique Ca^{2+} binding structure. The same consensus sequence is found in calreticulin, repeated three times, and is the region of highest conservation when these molecules are compared. Baksh and Michalak (1991) have shown that this domain in calreticulin contains a single high-affinity Ca^{2+} binding site by binding studies with a 137 amino acid segment fused to GST. Our results are consistent with this observation and further localize the binding region to the first set of repeats. Similar repeats are found in an *Onchocerca* surface antigen (Unnasch et al., 1988), a *Shistosoma mansoni* antigen (Hawn et al., 1993), and a yeast calnexin homologue (C. DeVirgilio, M. Buerckert, T. Boller, and A. Wiemken, unpublished: GenBank X66470). This domain has a high concentration of acidic residues, but the cytoplasmic domain and the sequence amino terminal to the first set of repeats have a greater number of acidic residues. Consequently, the first repeat sequence must assume a conformation that allows for chelation of Ca^{2+} which is dependent upon structural as well as electrostatic interactions. Such a structure may be critical for the regulation of Ca^{2+} stores in the ER. Our results clearly demonstrate that the second set of repeats does not bind Ca^{2+}. This result is somewhat surprising, considering the similarities to the first repeat structure: both repeats are similar in size and amino acid composition (single tryptophan, multiple prolines, numerous acidic residues), and both are repeated four times. This second set of repeats is highly conserved, suggesting that it maintains an important structural role. Perhaps this sequence helps to mediate the Ca^{2+} binding function of calnexin by structural positioning of the Ca^{2+} binding repeats.

The cytoplasmic domain consists of 91 amino acids at the calnexin carboxyl terminus. This domain is very highly charged (37% acidic residues, 22% basic residues) and

effectively binds Ca^{2+}, as shown in Figures 2B,C. This domain exhibits structural similarity to calreticulin but is positioned on the opposite (cytoplasmic) side of the ER membrane. Our results are consistent with those of Baksh and Michalak (1991), who observed that the homologous calreticulin sequence is a low-affinity, high-capacity Ca^{2+} binding site when fused to GST. The cytoplasmic localization of this domain may serve to link calnexin with other cellular components. The cytoplasmic domain shares significant similarity (35%) to the midsized neurofilament protein, a cytoskeletal protein found in axons (Myers et al., 1987). Thus, calnexin may interact with other members of the cytoskeleton via a Ca^{2+}-mediated mechanism. Such structural interactions may be involved in the ability of calnexin to act as a molecular chaperone.

We have observed expression of the calnexin gene in all tissues and cell lines that we have investigated (results not shown), consistent with its key role in cellular functions. We have also demonstrated that the calnexin gene is encoded on the distal end of the long arm of human chromosome 5 (5q35). Another membrane-associated Ca^{2+} binding protein has been localized to this region, at 5q32–34. This protein is termed p68 and is a member of the annexin/lipocortin family (Davies et al., 1989). p68 (annexin VI) also lacks the E-F hand Ca^{2+} binding sequence, but exhibits no structural similarity to calnexin. Chromosome 5 also encodes a large number of cell surface proteins, including colony stimulating factor 1 receptor, γ-aminobutyric acid receptor 1, platelet-derived growth factor receptor, glucose transporter 6, and growth hormone receptor (O'Brien, 1990). However, genes for cell surface proteins known to be associated with calnexin are found on other chromosomes: genes for MHC class I molecules and IFN-γ receptor on chromosome 6; genes for T-cell receptor chains on chromosomes 7 and 14; and immunoglobulin genes on chromosomes 2, 14, and 22 (O'Brien, 1990). The only known sequence with significant homology to calnexin is that of calreticulin, which is encoded on human chromosome 19 (McCauliffe et al., 1990).

ACKNOWLEDGMENT

We thank William Henzel for protein sequencing analysis, Tom Camerato for human calnexin cDNA sequencing, Chris Eberhardt for figure preparation, and Carol Raport for critical review of the manuscript.

REFERENCES

Ahluwalia, N., Bergeron, J. J. M., Wada, I., Degen, E., & Williams, D. B. (1992) *J. Biol. Chem. 267*, 10914–10918.

Baksh, S., & Michalak, M. (1991) *J. Biol. Chem. 266*, 21458–21465.

Calderon, J., Sheehan, C. F., Chance, C., Thomas, M. L., & Schreiber, R. D. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4837–4841.

Charuk, J. H. M., Pirraglia, C. A., & Reithmeier, R. A. F. (1990) *Anal. Biochem. 188*, 123–131.

Davies, A. A., Moss, S. E., Crompton, M. R., Jones, T. A., Spurr, N. K., Sheer, D., Kozak, C., & Crumpton, M. J. (1989) *Hum. Genet. 82*, 234–238.

Degen, E., & Williams, D. B. (1991) *J. Cell Biol. 112*, 1099–1115.

Farrar, M. A., & Schreiber, R. D. (1993) *Annu. Rev. Immunol. 11*, 571–611.

Farrar, M. A., Fernandez-Luna, J., & Schreiber, R. D. (1991) *J. Biol. Chem. 266*, 19626–19635.

Fliegel, L., Burns, K., MacLennan, D. H., Reithmeier, R. A. F., & Michalak, M. (1989) *J. Biol. Chem. 264*, 21522–21528.

Galvin, K., Krishna, S., Ponchel, F., Frohlich, M., Cummings, D. E., Carlson, R., Wands, J. R., Isselbacher, K. J., Pillai, S., & Ozturk, M. (1992) *Proc. Natl. Acad. Sci. U.S.A., 89*, 8452–8456.

Gibbs, V. C., Williams, S. R., Gray, P. W., Schreiber, R. D., Pennica, D., Rice, G., & Goeddel, D. V. (1991) *Mol. Cell. Biol. 11*, 5860–5866.

Gilchrist, J. S. C., & Pierce, G. N. (1993) *J. Biol. Chem. 268*, 4291–4299.

Hawn, T. R., Tom, T. D., & Strand, M. (1993) *J. Biol. Chem. 268*, 7692–7698.

Hochstenbach, F., David, V., Watkins, S., & Brenner, M. B. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 4734–4738.

Huang, L., Franklin, A., & Hoffman, N. E. (1993) *J. Biol. Chem. 268*, 6560–6566.

Jung, V., Rashidbaigi, A., Jones, C., Tischfield, J. A., Shows, T. B., & Pestka, S. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4151–4155.

Kretsinger, R. H. (1980) *Ann. N.Y. Acad. Sci. 356*, 14–19.

Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K., & Efstratiadis, A. (1978) *Cell 15*, 687–701.

Maruyama, K., Mikawa, T., & Ebashi, S. (1984) *J. Biochem. (Tokyo) 95*, 511–519.

McCauliffe, D. P., Lux, F. A., Lieu, T.-S., Sanz, I., Hanke, J., Newkirk, M. M., Bachinski, L. L., Itoh, Y., Siciliano, M. J., Reichlin, M., Sontheimer, R. D., & Capra, J. D. (1990) *J. Clin. Invest. 85*, 1379–1390.

Myers, M. W., Lazzarini, R. A., Lee, V. M. Y., Schlaepfer, W. W., & Nelson, D. L. (1987) *EMBO J. 6*, 1617–1626.

Nakai, H., Byers, M. G., Shows, T. B., & Taggart, R. T. (1986) *Cell Genet. 43*, 215–217.

O'Brien, S. J. (1990) *Genetic Maps, Locus Maps of Complex Genomes, Book 5 Human Maps*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Ou, W.-J., Thomas, D. Y., Bell, A. W., & Bergeron, J. M. (1992) *J. Biol. Chem. 267*, 23789–23796.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., & Erlich, H. A. (1988) *Science 239*, 487–491.

Sambrook, J. F. (1990) *Cell 61*, 197–199.

Shows, T. B. (1983) *Isozymes: Curr. Top. Biol. Med. Res. 10*, 323–339.

Shows, T. B., Sakaguchi, A. Y., & Naylor, S. L. (1982) *Adv. Hum. Genet. 12*, 341–452.

Shows, T. B., Eddy, R., Haley, L., Byers, M., Henry, M., Fujita, T., Matsui, H., & Taniguchi, T. (1984) *Somatic. Cell Mol. Genet. 10*, 315–318.

Smith, A. J. H. (1980) *Methods Enzymol. 65*, 560–580.

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

Smith, M. J., & Koch, G. L. E. (1989) *EMBO J. 8*, 3581–3586.

Unnasch, T. R., Gallin, M. Y., Soboslay, P. T., Erttmann, K. D., & Greene, B. M. (1988) *J. Clin. Invest. 82*, 262–269.

Wada, I., Rindress, D., Cameron, P. H., Ou, W.-J., Doherty, J. J., Louvard, D., Bell, A. W., Dignard, D., Thomas, D. Y., & Bergeron, J. J. M. (1991a) *J. Biol. Chem. 266*, 19599–19610.

Wada, K., Wada, Y., Doi, H., Ishibashi, F., Gojobori, T., & Ikemura, T. (1991b) *Nucleic Acids Res. 19*, 1981–1985.

Zabel, B. U., Naylor, S. L., Sakaguchi, A. Y., Bell, G. I., & Shows, T. B. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 6932–6936.