Human Natural Resistance–associated Macrophage Protein: cDNA Cloning, Chromosomal Mapping, Genomic Organization, and Tissue-specific Expression

By Mathieu Cellier,*‡ Gregory Govoni,* Silvia Vidal,* Tony Kwan,* Normand Groulx,* Jing Liu,‡ Fabio Sanchez,‡ Emil Skamene,‡ Erwin Schurr,‡ and Philippe Gros*

*From the *Department of Biochemistry, McGill University, Montreal, Quebec, Canada, H3G 1Y6; and ‡Montreal General Hospital Research Institute, Montreal, Quebec, Canada, H3G 1A4*

Summary

Natural resistance to infection with unrelated intracellular parasites such as Mycobacteria, Salmonella, and Leishmania is controlled in the mouse by a single gene on chromosome 1, designated *Bcg*, *Ity*, or *Lsh*. A candidate gene for *Bcg*, designated natural resistance–associated macrophage protein (*Nramp*), has been isolated and shown to encode a novel macrophage-specific membrane protein, which is altered in susceptible animals. We have cloned and characterized cDNA clones corresponding to the human *NRAMP* gene. Nucleotide and predicted amino acid sequence analyses indicate that the human NRAMP polypeptide encodes a 550-amino acid residue membrane protein with 10–12 putative transmembrane domains, two N-linked glycosylation sites, and an evolutionary conserved consensus transport motif. Identification of genomic clones corresponding to human *NRAMP* indicates that the gene maps to chromosome 2q35 within a group of syntenic loci conserved with proximal mouse 1. The gene is composed of at least 15 exons, with several exons encoding discrete predicted structural domains of the protein. These studies have also identified an alternatively spliced exon encoded by an Alu element present within intron 4. Although this novel exon was found expressed in vivo, it would introduce a termination codon in the downstream exon V, resulting in a severely truncated protein. Northern blot analyses indicate that *NRAMP* mRNA expression is tightly controlled in a tissue-specific fashion, with the highest sites of expression being peripheral blood leukocytes, lungs, and spleen. Additional RNA expression studies in cultured cells identified the macrophage as a site of expression of human *NRAMP* and indicated that increased expression was correlated with an advanced state of differentiation of this lineage.

Tuberculosis remains one of the most important infectious diseases worldwide and is becoming even more threatening because of the recent emergence of multidrug-resistant and highly virulent strains of *Mycobacterium tuberculosis* (1). The mechanism of host defense against *M. tuberculosis* in general, and the mechanisms underlying long-term persistence and replication of tubercle bacilli inside mononuclear phagocytes in particular, remain unclear and need to be better understood. Epidemiological studies of tuberculosis have suggested that initial susceptibility to infection is under genetic control. In addition, racial differences in susceptibility to invasion by *M. tuberculosis*, survival of certain ethnic groups during large epidemics, and studies with mono- or dizygotic twins suggest that disease establishment and progression are influenced by genetic factors of the host (for a review see reference 2). Although the specific genes involved have not yet been identified, recent studies suggest that such genetic factors may influence bactericidal activity of host macrophages toward phagocytized tubercle bacilli (3). A genetic control has also been proposed for susceptibility to leprosy, another mycobacterial infection caused by *Mycobacterium leprae*. In this case, susceptibility appears to be under dual genetic control, with one major autosomal gene determining susceptibility per se, and one gene linked to the MHC determining the type and severity of disease ultimately developed (for reviews see references 4, 5).

Natural resistance to infection with mycobacteria, e.g., *M. lepraemurium* (6), *M. intracellulare* (7), and *M. bovis* (8), is controlled in mice by the expression of a single dominant gene on mouse chromosome 1 designated *Bcg*, which controls early replication of these bacteria in the reticuloendothelial system of the mouse (8, 9). The gene has been shown to also control resistance to other antigenically unrelated intracellular parasites such as *Salmonella typhimurium* (*Ity*; 10) and *Leishmania donovani* (*Lsh*; 11, 12). We have delineated the maximal genetic (13) and physical (14) intervals defining the *Bcg*

 J. Exp. Med. © The Rockefeller University Press • 0022-1007/94/11/1741/12 $2.00
Volume 180 November 1994 1741–1752

gene region on mouse chromosome 1, and have identified several candidate transcription units for *Bcg* (15). One of them was found to be expressed exclusively in reticuloendothelial organs and in mature tissue macrophages, the tissues and cell type known to phenotypically express the genetic difference at *Bcg* (9). This candidate gene, designated *Nramp*[1] for natural resistance–associated macrophage protein, was found to encode a novel macrophage-specific polypeptide with predicted features characteristic of an integral membrane protein, including a minimum of 10 putative trans-membrane (TM) domains and two predicted N-linked glycosylation signals (15). The Nramp protein also contains a conserved consensus transport motif known as the "binding protein–dependent transport system inner membrane component signature" previously identified in a series of bacterial periplasmic transport proteins (16, 17). *Nramp* also shares similarity through this motif and other predicted structural aspects with the nitrate/nitrite concentrator of *Aspergillus nidulans* encoded by the *crnA* gene (18), which raises the possibility that *Nramp* may be involved in the metabolism/transport of oxidation products of nitric oxide, a key effector of the cytocidal arsenal of activated macrophages (19, 20). Finally, nucleotide sequence analyses of the *Nramp* cDNA showed that in 27 inbred mouse strains of either Bcg^r and Bcg^s phenotypes, the susceptibility trait was associated with a nonconservative glycine-to-aspartic acid amino acid substitution within predicted TM2 of the protein (21).

The goals of the present study were to (a) clone and characterize a cDNA for the human *NRAMP* homolog and analyze the pattern of conservation between the human and mouse proteins with respect to the predicted structural domains of the two proteins in general and the region bearing the mutation found in susceptible (Bcg^s) inbred mice in particular; (b) elucidate the genomic organization of the gene to obtain sequence information required for analyzing the role of human *NRAMP* in resistance to infection in humans; (c) investigate the expression pattern of the human gene in normal organs and cells and its potential involvement in macrophage-specific effector functions in relation to its proposed role in the metabolism of oxidized nitrogen intermediates.

Materials and Methods

cDNA Cloning and Nucleotide Sequencing. 5×10^5 independent clones from a human adult spleen cDNA library in λ gt10 vector (HL1039a; Clontech Laboratories Inc., Palo Alto, CA) were screened with an EcoRI-SacII fragment of the murine *Nramp* cDNA as hybridization probe [position 1–1720 of published sequence (15)], under conditions of high stringency as described (22). The mouse cDNA probe was labeled to high specific activity by random priming ($1–2 \times 10^9$ cpm/μg) (23). Filters were washed under conditions of increasing stringency up to 0.5 × SSC, 0.1% SDS at 60°C for 30 min. For nucleotide sequencing (24), single-stranded DNA templates were generated from pBluescript KS+/− subclones by superinfection with the M13 helper phage MK105. Double-stranded DNA plasmid templates were also used for nucleotide sequencing. Oligonucleotide primers used for sequencing were either derived from the plasmid vector or from the cDNA sequence itself. The complete cDNA sequence of human *NRAMP* cDNA was compiled from the analysis of two overlapping cDNA clones, designated H41 and H42. The nucleotide sequence data of the human *NRAMP* cDNA are available from EMBL/GenBank/DDBJ under accession number L32185.

Isolation of Genomic DNA Clones and Determination of the Exon/Intron Gene Structure. 5×10^5 independent clones from a human genomic DNA library constructed in the BamHI site of cosmid vector pWE15 (22) were screened for the presence of homologous human *NRAMP* sequences. Four independent but overlapping clones were purified, and their detailed restriction enzyme map was established for several enzymes. Oligonucleotide primers derived from either the plasmid vector, the human *NRAMP* cDNA, or genomic introns were used to determine the nucleotide sequence of genomic exons. Individual exon/intron boundaries were detected by a loss of identity between the genomic and cDNA nucleotide sequences and also by the presence of consensus donor and acceptor splicing signals at the point of divergence. The size of introns was determined by partial nucleotide sequencing, restriction mapping, and PCR amplification by use of exon-derived oligonucleotide primers.

Cell Culture. The human cell lines used were obtained from American Type Culture Collection (Rockville, MD) and included the acute myelogenous leukemia KG1 (CCL246), the promyelocytic leukemia HL60 (CCL240), the histiocytic lymphoma U937 (CRL1593), and the monocytic cell line THP1 (TIB202). Cells were cultured in Dulbecco's modification of α minimal Eagle's medium with a high concentration of glucose (100 mM), 10% heat-inactivated FCS, 2 mM glutamine, 100 U/ml penicillin, and 100 μg/ml streptomycin (GIBCO BRL, Gaithersburg, MD). Cultures were maintained in a humidified atmosphere consisting of 95% air and 5% CO_2 at 37°C by passage of $1–2 \times 10^5$ cells/ml every other day. Cell viability was tested by trypan blue exclusion.

Amplification (RT-PCR) of Human NRAMP cDNAs. Total or Poly(A)$^+$ RNA was used as template and random hexamers as primers for the murine Moloney leukemia virus reverse transcriptase–mediated cDNA synthesis (GIBCO BRL), as previously described (25). Oligonucleotide primers HN1 (5′-CGTGGTGACAGGCAAGGA-3′, positions 417–435) and HN2 (5′-GACAAAGAGGTTGATGATA-3′, positions 988–970) were used to PCR amplify the alternatively spliced human *NRAMP* cDNAs. Two oligonucleotide primers of sequence 5′-ACCCCCACTGAAAAAGATGA-3′ and 5′-GTTCAAACTTCTGCTCCTGA-3′ were used for control amplification of a β_2-microglobulin cDNA fragment (26). Reaction conditions for PCR amplification of human *NRAMP* cDNA were 94°C (1 min), 50°C (1.5 min), and 72°C (2 min) for 30 cycles. Reaction products were visualized by agarose gel electrophoresis after ethidium bromide staining and analyzed by Southern blot with a 300-bp human *NRAMP* cDNA fragment generated by PCR amplification (positions 515–819) as hybridization probe.

Computer-assisted Analyses. Analyses of nucleotide and predicted amino acid sequences were performed on the VAX/VMS server and computer of the Université de Montréal by use of the UWGCG software package (University of Wisconsin, Madison, WI) (27) and with the DNA Strider package (28). Homology searches in the National Center for Biotechnology Information data bases were performed by use of the BLAST server (National Library of Medicine, National Institutes of Health, Rockville, MD) (29). Identification and analysis of the *Alu* sequences present in the *NRAMP* gene was performed using *Alu* program (30) of the PYTHIA server.

[1] *Abbreviations used in this paper:* NRAMP, natural resistance–associated macrophage protein; ORF, open reading frame; TM, trans-membrane.

RNA Blotting Experiments. Total RNA was extracted either from normal human spleen or cultured human cells as described (31). PMN from heparin-treated blood were purified by centrifugation on a density gradient (Histopaque™; Sigma Chemical Co., St. Louis, MO). The polyadenylated RNA fraction was selected by standard chromatography on oligo (dT)-cellulose (Pharmacia LKB Biotechnology, Piscataway, NJ) and analyzed by Northern blot after electrophoresis in formaldehyde-containing agarose gel (22). These membranes and human multiple tissues Northern filters I and II (Clontech) were prehybridized in a solution containing 10% dextran sulfate, 1 M NaCl, 1% SDS, and 200 μg/ml heat-denatured salmon sperm DNA for 2 h at 65°C. Hybridization for 16 h at 65°C was done in the same solution containing 2 × 10^6 cpm/ml ^{32}P-radiolabeled probe (human *NRAMP* PCR fragment, position 515–819; sp act 1–2 × 10^9 cpm/μg DNA). Final wash conditions were 0.5 × SSC, 0.5% SDS at 65°C for 30 min.

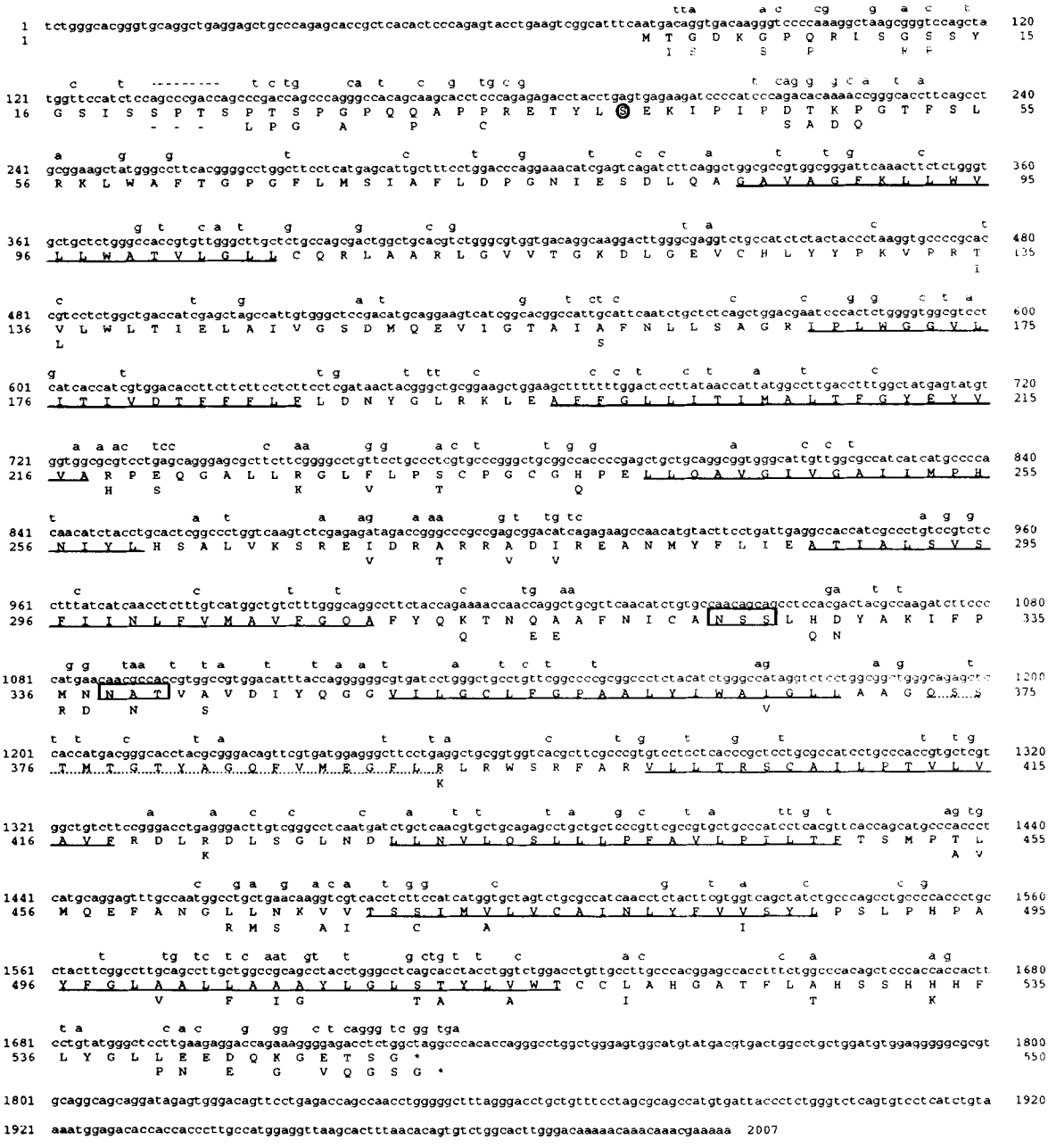
Results

Isolation of NRAMP cDNA Clones. A human spleen cDNA library was screened under high stringency conditions with a mouse *Nramp* cDNA probe yielding five positive clones for a frequency of ~0.001%. Restriction enzyme mapping and nucleotide sequence analyses showed that these clones were overlapping and derived from the same mRNA species, although none of them were full length. The nucleotide sequence of the assembled cDNA and the predicted amino acid sequence of the encoded protein are shown in Fig. 1 *A*. The first in-frame initiator methionine codon was found at position 77 of the sequence and was followed by a segment of 1,647 nucleotides forming a single open reading frame (ORF) capable of encoding a protein of 550 amino acid residues with a calculated molecular mass of ~60 kD. A TGA termination codon was found in the same reading frame at position 20 upstream from the putative initiator ATG, indicating that the segment upstream position 77 was untranslated. A TAG termination codon located immediately downstream from glycine 550 (nucleotide position 1727) was followed by an additional 276 nucleotides of putative 3′ untranslated sequence. No clear consensus polyadenylation signal (32) could be identified within this sequence.

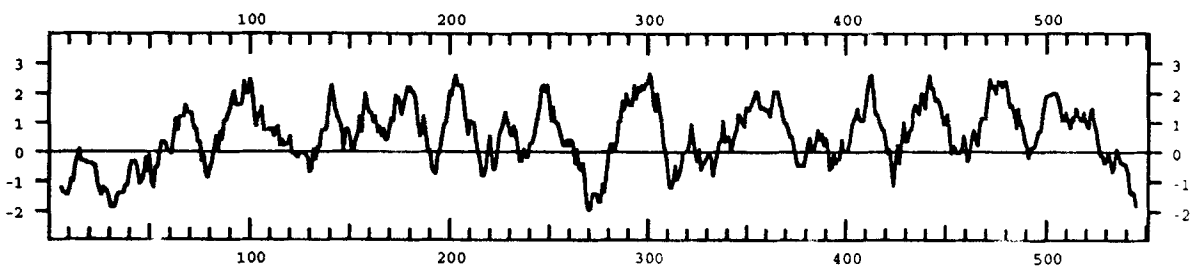
Analysis of the Predicted NRAMP Polypeptide. Analysis of the predicted amino acid sequence of the human NRAMP polypeptide revealed that is was most likely an integral membrane protein. Highly hydrophobic amino acids such as leucine (84, 15%), isoleucine (34, 6%), valine (37, 7%), alanine (54, 10%), phenylalanine (32, 6%), and the apolar residue glycine (47, 9%) accounted for more than half (53%) of the polypeptide composition, whereas charged residues such a arginine/lysine (35, 6%) or aspartic/glutamic acid (34, 6%) accounted for only 12%. Hydropathy analysis (Fig. 1 *B*) suggested that human NRAMP contained a minimum of 10 highly hydrophobic segments, which displayed minimal charge density. Although examination of the NH_2-terminal sequence of the protein did not reveal a classical cleavable signal sequence, this region contained a high proportion of hydrophylic or charged residues incompatible with membrane insertion but possibly acting as a hydrophylic cytoplasmic anchor. The human NRAMP protein showed within the segment separating putative TM domains 5 and 6, two predicted N-linked glycosylation sites at amino acid positions 224 (NXS) and 238 (NXT). The consensus sequence motif known as the "binding protein–dependent transport system inner membrane component signature" (17) was found in the segment separating predicted TM domains 6 and 7 (positions 373–392, dotted underline in Fig. 1 *A*). This motif, which has been evolutionarily conserved from prokaryotes to eukaryotes, is predicted to be localized in cytoplasmic loops of periplasmic transport proteins and possibly mediates interaction of the hydrophobic membrane anchors with peripheral ATP binding subunits to energize these transporters (16). The combined analysis of the hydropathy profile and the location of known structural/functional motifs suggested the following topology for human NRAMP. The NH_2 terminus would be cytoplasmic and followed by 10 consecutive TM domains forming five membrane loops, placing the COOH terminus on the cytoplasmic side. This arrangement would position the TM 5–6 loop as extracytoplasmic (predicted sites for N-linked glycosylation), whereas the TM 6–7 loop carrying the consensus transport motif would be intracytoplasmic.

Homology Between Mouse and Human NRAMP Proteins. Comparison of the human and mouse predicted NRAMP protein sequences revealed a remarkable degree of conservation between the two polypeptides, with 88% identical residues and 93% overall sequence similarity. Substitutions were not randomly distributed along the length of the protein but were clustered within discrete predicted structural domains (Fig. 2, *E* and *F*). The extreme NH_2- (residues 1–50, 66% identity) and COOH-terminal ends (residues 517–550, 74% identity) of the two proteins were the least conserved segments, and gaps of three residues (positions 20–22 of mouse sequence) and one residue (position 546 of human sequence) had to be introduced to optimize alignment in these segments. In addition, the predicted second (TM 3–4 segment, positions 218–240; 63% identity) and third extracytoplasmic loops (TM 5–6 segment, positions 310–349; 67% identity) showed significant sequence divergence. The fifth putative intracytoplasmic loop located between TM 8–9 (positions 450–468; 61% identity) and the last predicted TM 10 domain (positions 496–517; 67% identity) also contained a cluster of substitutions that were mainly conservative. The rest of the NH_2-terminal third up to and including TM 3 (residues 50–217) and the segment overlapping TM 6 to TM 8 (residues 340–462) were the most conserved portions (97% identical, 99% similar; Fig. 2, *E* and *F*). The two predicted phosphorylation sites by protein kinase C identified in the mouse sequence (15) were not conserved in the human protein; instead, a new one was found within the NH_2-terminal region (pst 54, Fig. 1 *A*). By contrast, both proteins have retained the two predicted N-linked glycosylation sites within the third predicted extracellular loop (TM 5-6 interval). As expected, the binding protein dependent transport system inner membrane component signature was highly conserved in human NRAMP, including a conservative arginine-to-lysine substitution at last position, 392, of the motif. The highly hydrophobic predicted membrane-associated domains of the proteins were strikingly conserved in the two species. Predicted

A



B



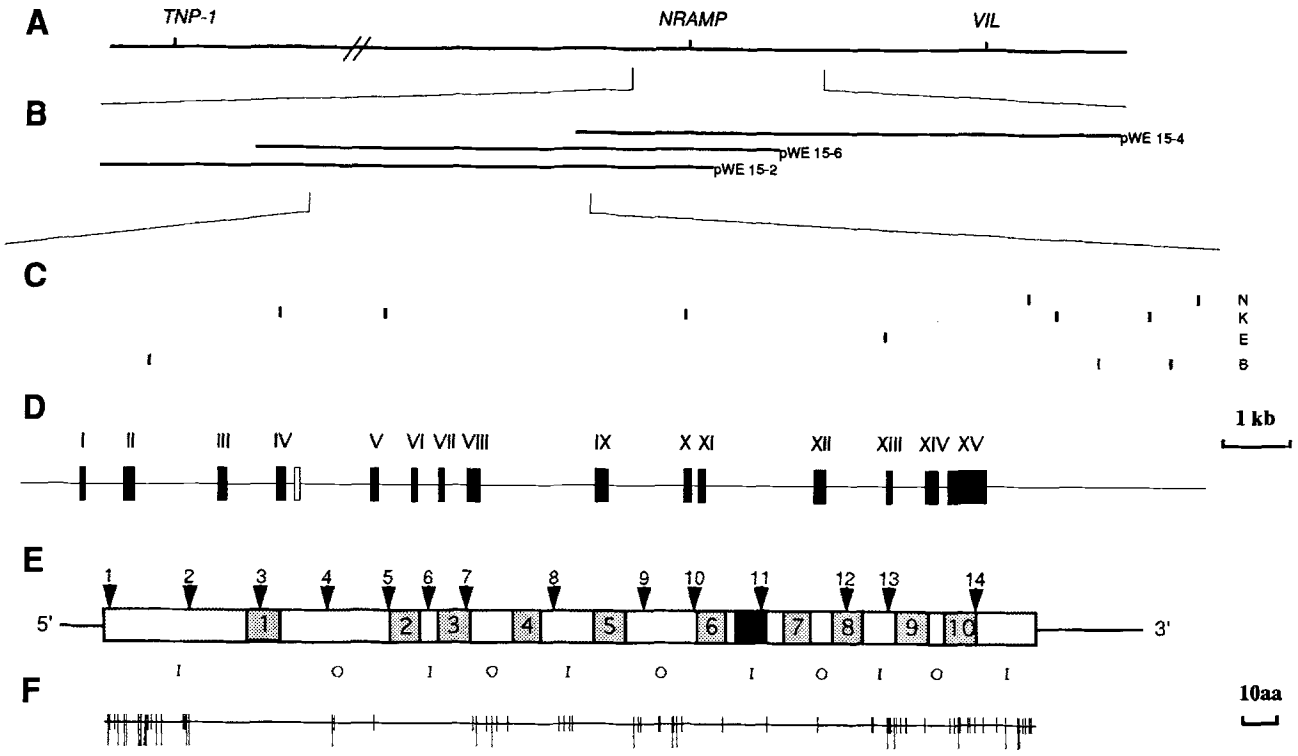


Figure 2. Genomic organization of the human *NRAMP* gene region. (*A*) Physical map of the *NRAMP* gene region on human chromosome 2q delineated by the *TNP-1* (2q35-36, proximal) and *VIL* (2q35, distal) genes. The maximal distance separating *NRAMP* and *VIL* is estimated at 220 kb. (*B*) Position and relationship of cosmid clones hybridizing to the *NRAMP* cDNA and used to characterize the gene. (*C*) The position of the recognition sites of restriction enzymes NotI (*N*), KpnI (*K*), EcoRI (*E*), and BamHI (*B*) within the cosmid clones are indicated. (*D*) Genomic organization and exon/intron structure of the human *NRAMP* gene. Exons are indicated by filled boxes numbered with roman numbers, whereas intervening and flanking sequences are represented by a thin line. The empty box between exons IV and V corresponds to the alternatively spliced Alu-derived exon (see text). The direction of transcription is from left to right. (*E*) Schematic representation of the human *NRAMP* mRNA and the putative encoded product. The thin lines represents the 5′ and 3′ untranslated sequences, and the boxed region indicates the mRNA coding region. The boxes numbered 1–10 correspond to the 10 putative transmembrane segments, and the solid box identifies the position of the binding-protein–dependent transport system inner membrane component signature (17). The proposed assignment of intracytoplasmic (*I*) and extracytoplasmic (*O*) loops is indicated. Arrowheads numbered from 1–14 identify intervening sequences in the precursor mRNA. The corresponding scales for the gene (1 kb) and the encoded polypeptide chain (10 amino acids) are indicated. (*F*) Schematic representation of the nonidentical amino acid residues at homologous positions of the mouse and human *NRAMP* polypeptide chains. Conservative substitutions are represented by short single bars, whereas nonconservative substitutions are shown as long double bars.

TM 1–8 were identical in sequence (except for a single isoleucine-to-valine conservative substitution in TM 6), including four charged amino acids in TM 1–3 and 7. This suggests that these TM domains and residues play key structural and/or functional roles in this protein.

Cloning of the Human NRAMP Gene. Genetic linkage and physical mapping studies have established that the mouse *Bcg/Nramp* gene maps on the proximal portion of chromosome 1, very close (125 kb) to the intestinal protein gene *villin* (14). This chromosomal region forms part of a large syntenic

Figure 1. Human *NRAMP* cDNA. (*A*) Nucleotide and deduced amino acid sequences of the human *NRAMP* cDNA. The complete nucleotide sequence was determined from the combined analysis of clones H41 and H42. Nucleotides are numbered positively in the 5′–3′ orientation from the first nucleotide identical between the H41 clone and the genomic sequence determined in this study. The deduced amino acid sequence corresponding to the ORF found in the nucleotide sequence is shown immediately below. Numbering starts at the first in-frame ATG and stops just before the star indicating the termination codon at the beginning of the 3′ untranslated sequence. Potential glycosylation sites are identified by boxed residues corresponding to the consensus sequence N-X-S/T. Predicted phosphorylation site for protein kinase C (S/T-X-R/K) is circled. The highly hydrophobic segments possibly corresponding to transmembrane domains are underlined. The protein motif binding-protein–dependent transport system inner membrane component signature was the only convincing functional signature identified in Bairoch's database (17) and is indicated by a dotted underline. A comparison of the nucleotide sequence and amino acid sequence of the corresponding ORF of the mouse *Nramp* cDNA is presented. Nucleotides and amino acid residues differing between the two sequences are shown immediately above and below the corresponding human sequence, respectively. (*B*) Hydropathy profile (hydrophobic, positive; hydrophilic, negative) obtained for the predicted polypeptide sequence deduced from the human *NRAMP* cDNA, using Kyte and Doolittle (42) hydropathy values and a window of 11 amino acids. These sequence data are available from EMBL/GenBank/DDBJ under accession number L32185.

Table 1. *Splice Junctions in the NRAMP Gene*

Exon/intron number	Exon length	Exon end*	Exon 3′ junction	Intron 5′ junction	Intron length‡	Intron 3′ junction	Exon 5′ junction	Intron type§
	bp				*bp*			
I	>73	83	ATG ACA G	gtgagtagtggc	585	gggcccccacag	GT GAC AAG	1
II	143	226	ACA AAA CCG	gtgggatgctgg	~1,100	acccccccaacag	GGC ACC TTC	0
III	123	349	GGA TTC AAA	gtaactaagtcg	~800	gcctcctcacag	CTT CTC TGG	0
IV	120	469	TAC CCT AAG	gtgagcttgggg	~1,200	tcctccccatag	GTG CCC CGC	0
V	107	576	GCT GGA CG	gtaccaccccag	~450	ctctggtttcag	A ATC CCA	2
VI	71	647	AAC TAC G	gtgggtgcacac	326	tctgctccgtag	GG CTG CGG	1
VII	68	715	GGC TAT GAG	gtaggaagccag	~300	tcctccctgcag	TAT GTG GTG	0
VIII	156	871	CTG GTC AAG	gtgagcagaggg	~1,500	tgatcttcgtag	TCT CGA GAG	0
IX	159	1,030	CAG GCT GCG	gtgagacacact	~1,200	ggtacccaacag	TTC AAC ATC	0
X	90	1,120	TAC CAG GGG	gtgagcgcgggt	87	ctcgtcctgcag	GGC GTG ATC	0
XI	120	1,240	GTG ATG GAG	gtagggcagggg	~1,500	cccccaccccag	GGC TTC CTG	0
XII	150	1,390	AGC CTG CTG	gtgagatacgcc	~900	tgctctccccag	CTC CCG TTC	0
XIII	74	1,464	AAT GGC CT	gtgagtaccccc	~400	cgtgtcccccag	G CTG AAC	2
XIV	154	1,618	ACC TAC CTG	gtacagtagggc	139	cccctcccccag	GTC TGG ACC	0
XV	>389	2,007						

Consensus Sequences:	N	G	/G	T	R	V	G	Y	Y	Y	Y	Y	Y	N	Y	A	G	/N
Frequency (%):	100	85.7	100	100	100	100	78.6	85	71.5	92.8	85.7	100	85.7	100	100	100	100	100

* Positions in the cDNA sequence.
‡ Length of the introns was determined by a combination of PCR and restriction enzyme mapping of genomic plasmid subclones.
§ Intron type is indicated by 0, 1, and 2 when a splice junction occurs between codons, after the first nucleotide of the codon, and after the second nucleotide of the codon, respectively.
N = A/C/G/T, R = A/G, Y = C/T/, V = A/G/C.

group conserved on the long arm of human chromosome 2 (2q32-37; reference 33), suggesting that the human *NRAMP* homolog may be located near 2q35, where the human *VIL* homolog has been previously mapped (34). Indeed, a specific *NRAMP* PCR amplification product could be obtained from several genomic YAC clones containing the *VIL* gene, by use of sequence-specific oligonucleotide primers derived from the human *NRAMP* cDNA (data not shown). Additional physical mapping of these YAC clones indicated that *NRAMP* and *VIL* were located on a genomic fragment of maximum size 220 kb (Fig. 2 *A* and data not shown). To elucidate the genomic exon/intron organization of the human *NRAMP* gene, the mouse *Nramp* cDNA was used to isolate overlapping cosmid clones (Fig. 2 *B*). Perfect concordance was obtained between the map of the cloned region and that of the corresponding human genomic DNA segment deduced by Southern blot analysis of total genomic DNA (data not shown). These hybridization experiments showed that the human *NRAMP* gene was present at a single copy per haploid genome spanning a 16-kb genomic region (Fig. 2 *C*). The gene was found to be composed of at least 15 exons (Fig. 2 *D*). Their length and position within the cDNA are illustrated in Fig. 2 *E* and are summarized in Table 1. All splice donor and acceptor sites conformed to the AG/GT rule and gave a consensus splicing sequence for the gene, which was strictly in accordance with the consensus proposed by Mount (35). Although exon I was defined as containing the first nucleotide of the cDNA sequence, the transcription initiation site of the gene has not yet been accurately mapped. Coding exons varied in size from 68 to 159 bp, whereas exon XV, which encodes COOH terminus of the protein and the 3′ untranslated of the mRNA, was the largest at 389 bp (Table 1). A juxtaposition of the intron's position onto the *NRAMP* cDNA and the predicted structural features of the protein is shown in Fig. 2 *E*. The first exon encoded 76 bp of untranslated sequence, the proposed initiator ATG codon, and four additional bases. Exon II encoded a large portion of the intracytoplasmic NH_2 terminus, whereas exon III encoded the rest of the intracytoplasmic domain and the beginning of TM 1. Likewise, exon IV encoded the rest of TM 1 and the beginning of the first extracytoplasmic loop, whereas exon V encoded the rest of the extracytoplasmic loop. Exons VI and VII were among the shortest in the gene, and each encoded a single TM domain (TM 2 and 3, respectively). Exon VIII encoded the second extracytoplasmic loop and TM 4, exon IX the third intracytoplasmic loop and TM 5, exon X

the third extracytoplasmic loop, and exon XI encoded most of the highly conserved transport motif together with TM 6. Finally, three of the remaining four TM domains were encoded by individual exons XII (TM 7) and XIV (TM 9–10). Three coding exons (II, XIV, and XV) were found to contain the majority of amino acid replacements between the mouse and human proteins (Fig. 2, *E* and *F*). Together, these results indicate a clear correlation between the organization of *NRAMP* gene exons and the predicted structural domains of the protein.

Identification of an Alternatively Spliced Exon Encoded by an Inverted Alu Sx Element. During the characterization of the human *NRAMP* cDNAs from the spleen library, one of the clones (clone H41) was found to contain an insertion of 74 bp at nucleotide position 470. This sequence had no homolog in the mouse *Nramp* cDNA, but its insertion corresponded precisely to the position of intron 4. This insertion apparently caused a frameshift in the mRNA, resulting in the appearance of a termination codon within exon V, 32 residues downstream of the insertion site. To determine if the identified insertion was a cloning artifact or reflected the presence of a splicing-competent exon within the boundaries of intron 4, the nucleotide sequence of intron 4 was determined (Fig. 3 *A*). Sequence comparison of the 74-bp insertion fragment present in the cDNA with that of intron 4 revealed 100% identity with an intronic segment flanked by consensus donor and acceptor splicing signals. Moreover, a homology search in GenBank showed that this additional exon mapped within the boundaries of an Alu element inserted in an inverted orientation relative to the direction of the gene transcription (Fig. 3 *A*). Comparison of the sense strand of this 300-bp element with Alu consensus sequences gave 87% identity (Fig. 3 *B*) and allowed us to establish that this Alu sequence belonged to the Sx Alu family (36). Fig. 3 *C* shows the results of BLASTP and BLASTN analyses against the NCBI data bases, using either the putative truncated protein sequence or the Alu nucleotide sequence as query sequences, respectively. These analyses indicated that other mRNAs contained similar Alu-derived exons capable of encoding a homologous peptide sequence in minor alternatively spliced isoforms. This suggested that a common mechanism of alternative splicing mediated by the Alu Sx sequence described here for the *NRAMP* gene could also occur in other genes (37–39). To determine if the alternative splicing event detected by cDNA and genomic cloning occurred in vivo and to quantitate the respective expression levels of the two mRNAs, we tested cellular mRNAs for the presence of the two splice forms by RT-PCR by use of *NRAMP*-specific oligonucleotide primers. Amplification products were analyzed by agarose gel electrophoresis and quantitated by Southern blotting (Fig. 4 *C*). Finally, the authenticity of the amplification products was verified by nucleotide sequencing (data not shown). Results showed that both splice forms were indeed present in all cell lines and tissues tested, at an approximate ratio of 1:5, in favor of the mRNA species lacking the Alu-derived exon (Fig. 4 *C*).

Tissue-specific Expression of NRAMP mRNA. The pattern of tissue-specific expression of human *NRAMP* mRNA was investigated by Northern blot analysis under conditions of high stringency. *NRAMP* mRNA expression was found to be highly restricted to certain tissues; in these tissues, two major hybridizing mRNA species of ~2 and 4 kb in size were coexpressed. The size of the shorter species at 2 kb was consistent with the 2,007-bp size of the reconstructed *NRAMP* cDNA (Fig. 1 *A*), suggesting that the cDNA reported here is near or full length. The origin of the larger 4-kb mRNA remains unclear, but could result from the use of alternative polyadenylation signals in the gene. The site of highest *NRAMP* mRNA expression was found to be PBL. This very high expression was not unique to the individual used to produce the blot, because a comparable signal was obtained with RNA from eight additional and unrelated healthy volunteers (data not shown). A high expression level was observed in lungs and a moderate to low level detected in spleen and liver, respectively. Normalization of hybridization signals with actin (Fig. 4 *B*) and GAPDH (data not shown) control probes revealed that *NRAMP* expression in spleen was much higher than in liver. The high levels of *NRAMP* expression detected in PBL, lung, and spleen, together with the absence of expression in other tissues, suggested that expression of this gene may be restricted to cells of the reticuloendothelial system such as the macrophage/monocyte lineage. That possibility was further investigated in vitro by use of RT-PCR analysis of RNA from a series of human cell lines showing characteristics of this lineage. Cell lines KG-1, HL 60, U 937, and THP1 represent immature precursors of the myelomonocytic lineage blocked by immortalization/transformation in a proliferative state at different stages of differentiation towards phagocytic leukocytes (40). KG-1 cells are the closest to normal bone marrow precursors, whereas THP1 are closest to the monocytic cell type. Results shown in Fig. 4 *C* indicate that all cell types analyzed expressed *NRAMP* mRNA. Interestingly, an apparent correlation between the level of expression and the commitment of the cell towards the phagocytic type was noted.

Discussion

Southern blot analysis of total human genomic DNA with the mouse *Nramp* cDNA probe under low stringency conditions identified two sets of hybridizing fragments, a set of strongly hybridizing ones and a set of a weakly cross-hybridizing ones, suggesting the presence of at least two sequence-related *NRAMP* genes in humans (Vidal, S., and P. Gros, unpublished observations). Therefore, high stringency conditions were used to isolate a human *NRAMP* homolog. Sequence analysis of the positive clones identified a very high degree of nucleotide sequence conservation with the mouse counterpart (Fig. 1 *A*), and mapping studies indicated that the corresponding human gene mapped on human 2q35 proximal to the *VIL* gene on a chromosomal segment homologous to the portion of mouse 1 carrying *Bcg/Nramp* (Fig. 2 *A*) (33). Additional mapping studies excluded human chromosome 2 as the site for the second gene identified by cross-hybridization in Southern blotting (data not shown). Together, these results indicated that the cDNA and gene

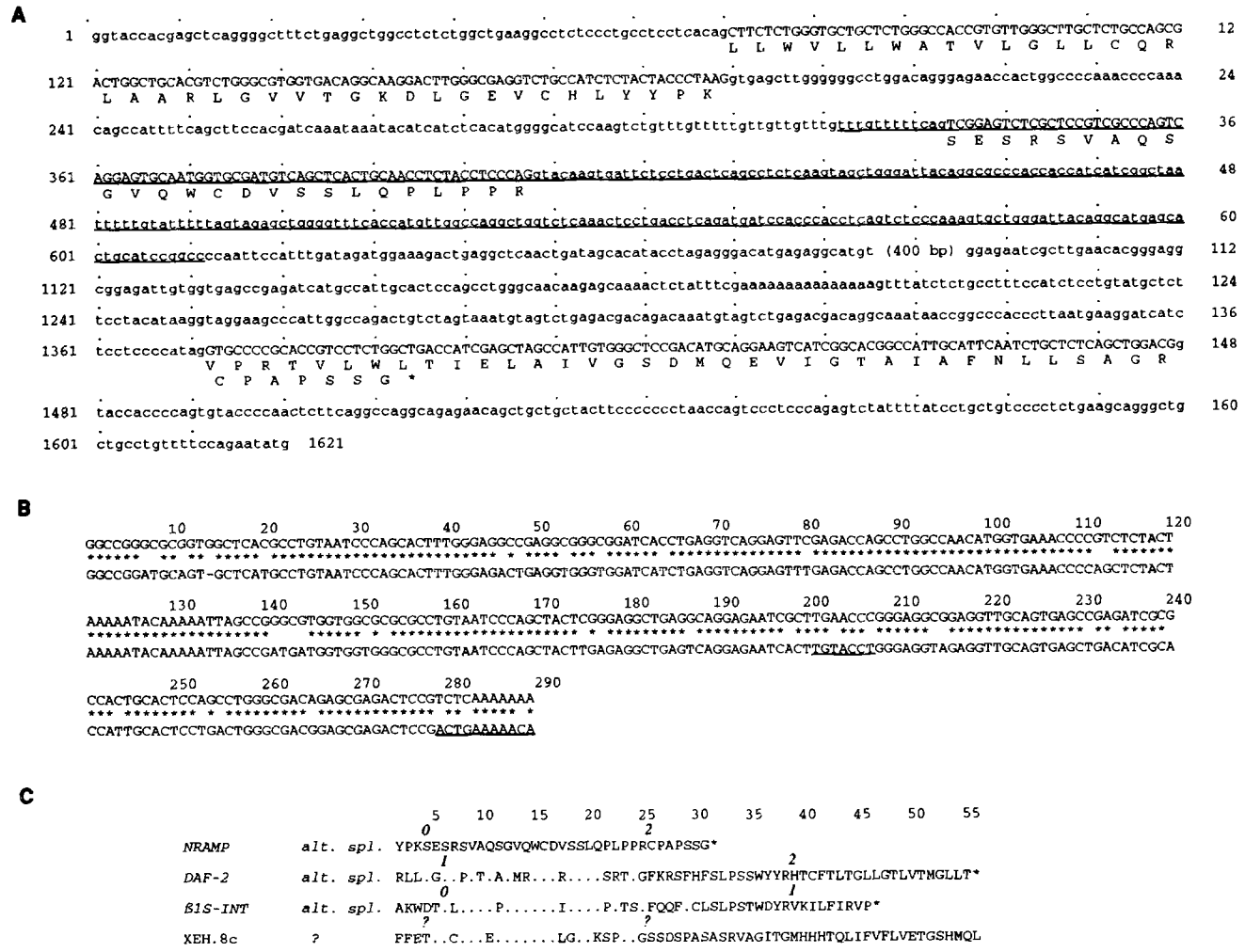


Figure 3. Identification of an alternatively spliced exon encoded by an inverted Alu Sx sequence in intron 4 of the *NRAMP*. (*A*) Nucleotide sequence of a portion of the *NRAMP* gene encompassing part of intron 3, exon IV (positions 71–190), parts of intron 4, exon V (1373–1379), and part of intron 5. The inverted Alu sequence identified within intron 4 is underlined. Exons (including the alternatively spliced exon within intron 4) are in capital letters in the nucleotide sequence, and the encoded amino acid sequences are indicated below. Alternative splicing of the Alu-encoded exon introduces a frameshift resulting in the appearance of a termination codon in the following exon V. (*B*) Nucleotide sequence alignment of the sense strand of the inverted Alu sequence found in intron 4 of *NRAMP* (*lower line*) with the sense strand of the consensus Alu sequence (*upper line*) gave 87% identical positions, indicated by stars. Analysis of the divergent positions pattern indicates that this Alu belongs to the Sx family (36). (*C*) Polypeptide sequences from other human genes sharing similarity with the predicted sequence encoded by the alternatively spliced Alu-encoded exon. These include alternatively spliced exons (*alt. spl.*) from the *DAF* gene (complement decay accelerating factor; 37), the *β1s-INT* gene (*β*-integrin; 38) and a sequence from the XEH. 8c-expressed sequence derived from a candidate gene for X-linked retinopathy (39). Only nonidentical residues are shown. Numbers in italics above each polypeptide sequence indicate the type of intervening sequences and position of the splice site for each particular Alu-encoded sequence. These sequence data are available from EMBL/GenBank/DDBJ under accession number L36019.

identified here correspond to the human homolog of mouse *Nramp*. Comparison of the nucleotide and predicted amino acid sequences of the mouse and human *NRAMP* cDNAs revealed one important difference at the 5′ end of the mRNA. In the previously reported mouse *Nramp* cDNA sequence (15), the first in-frame initiator methionine was followed by an ORF of 1,452 nucleotides, as opposed to the 1,647 nucleotides ORF detected in human *NRAMP*. At the 5′ end, nucleotide sequence conservation between the two species extended 51 nucleotides upstream of the proposed murine ATG, and then disappeared abruptly. Translation of this 51-nucleotide putative 5′ UT segment showed that the corresponding amino acid sequence would be identical to its human counterpart (Fig. 1 *A*). In addition, the abrupt loss of nucleotide sequence similarity between the human and mouse sequences in this region coincided with the position of intron 2 in the human gene (Table 1). Subsequent S1 mapping experiments and genomic DNA sequencing verified that the 5′ end of the published mouse *Nramp* cDNA sequence contains an unspliced intron. Additional mouse *Nramp* cDNA clones were therefore constructed, and their analysis revealed an additional 64 amino acid residues upstream of the previously suggested ATG (Govoni, G., and P. Gros, unpublished results; sequence data available from EMBL/GenBank/DDBJ under corrected ac-

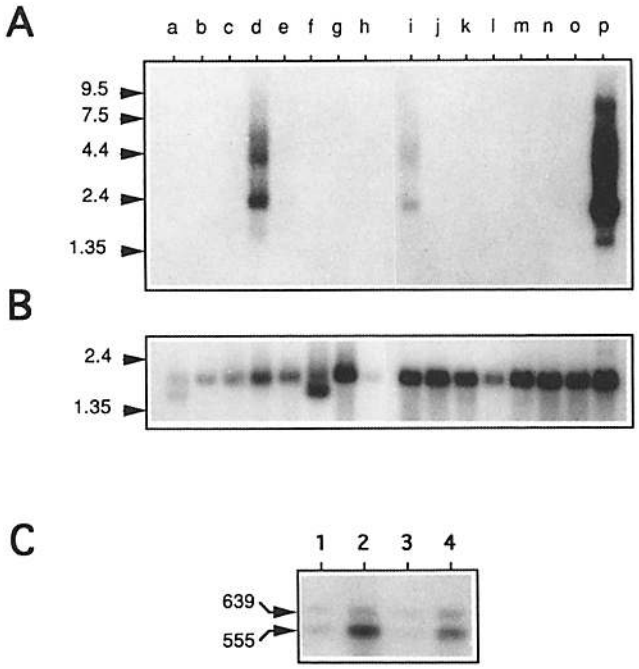


Figure 4. Tissue-specific expression of the human *NRAMP* gene in normal organs and cell lines. (*A*) Northern blot analysis of *NRAMP* transcripts expressed in Poly(A)$^+$ mRNA from different normal human tissues, including heart (*a*), brain (*b*), placenta (*c*), lung (*d*), liver (*e*), skeletal muscle (*f*), kidney (*g*), pancreas (*h*), spleen (*i*), thymus (*j*), prostate (*k*), testis (*l*), ovary (*m*), small intestine (*n*), colon (*o*), and peripheral blood leukocytes (*p*). (*B*) Rehybridization of the blot shown in (*A*) with a control chicken β-actin cDNA probe. Position of size markers (in kilobases) are shown (*arrowheads*). (*C*) RT-PCR analysis of the human *NRAMP* mRNA expression in human cell lines representing different steps along the macrophage differentiation pathway: U 937 (*1*), THP 1 (*2*), KG-1 (*3*), and HL 60 (*4*). RT-PCR and Southern blotting were as described in Materials and Methods. The two hybridizing cDNA species, corresponding to either the normal (555 bp) or the alternatively spliced Alu-derived exon (639 bp) are indicated by arrows.

cession number L13732). This novel additional sequence was homologous to that predicted for human NRAMP and was used in our analysis (Figs. 1 *A* and 2 *F*). During the course of these studies, portions of the 5′ end sequence of the mouse cDNA and exon 2 of the human gene were independently reported (41). These partial sequences were found identical to those reported here.

Analysis of the full-length predicted amino acid sequence of the NRAMP polypeptide, including the presence of landmark functional and structural features, has led us to propose a possible topology in which the NH_2- and COOH-terminal segments of the protein would be intracytoplasmic and separated by 10 putative TM domains. The predicted 10-TM domain configuration has interesting consequences on the distribution of charged residues with respect to the membrane. For positively charged residues (Arg/Lys), 21 would be intracytoplasmic, 12 would be extracytoplasmic, and 2 would map within predicted TM domains; for negatively charged residues, the corresponding distribution would be 20, 12, and 2, respectively, resulting in an almost perfect charge balance on either side and also within the membrane (net charge of +1). Careful scrutiny of the hydropathy profile (42) and hydrophobic moment analysis (43) of the sequence for the presence of hydrophobic or amphipathic α helices identified two additional putative TM domains in the NH_2-terminal segment of NRAMP. The first one would be delineated by residues 58–78 between the initiator ATG and the identified TM 1, whereas the second one would map at positions 137–158, between the identified TM 1 and TM 2. In addition, these putative additional TM domains map in a region of the molecule that is highly conserved in the mouse homolog. However, the latter additional TM domain would be thermodynamically disfavored, requiring the stable insertion of three negatively charged residues in the membrane. Therefore, NRAMP would encode a minimum of 10 and a maximum of 12 putative TM domains.

Sequence comparison of the human and mouse proteins revealed a remarkably high degree of conservation between the two species (88% identity, 93% similarity). Whereas conservation of the predicted membrane topology, N-linked glycosylation signals, and "transport motif" was expected, the sequences of 10 out of the 12 possible TM domains were precisely conserved, including the presence of several charged residues in these domains. This extremely high degree of conservation is somewhat surprising since conservative substitution of hydrophobic amino acids are usually tolerated in TM domains without alteration or loss of function between homologs of the same proteins (44). This is of particular interest since the only amino acid sequence variation detected in Nramp between inbred mouse strains of opposite genotypes at *Bcg* is a single nonconservative glycine-to-aspartic acid substitution within predicted TM 2 identified in innately susceptible (Bcg^s) strains (15, 21). The observation that this TM domain in general and glycine in particular are precisely conserved in human and mouse (this study) but also in the homologous region of the rat and chicken gene (21) strongly argues that this domain is important for function and that nonconservative substitutions in this region are likely to affect the function of this putative transporter.

Although the putative NRAMP substrate or mechanism of transport remain to be identified, we have detected in molecular database searches a few sequences locally similar to NRAMP, in particular, in the region of the consensus transport motif. These short "expressed sequence tags" (45) are issued from plant genome research projects. The rice (*O. sativa*; sequence data available from EMBL/GenBank/DDBJ under accession number D15268) and Arabidopsis (*A. thaliana*; EMBL/GenBank/DDBJ accession number Z30530) protein sequences exhibited 46% identity (79% similarity) and 51% identity (73% similarity), respectively (Cellier, M., and P. Gros, unpublished results). Given the very low probability for these TBLASTN hits to occur by chance ($p = 5.0\ e^{-17}$ and $p = 1.7\ e^{-09}$, respectively), these similarities in primary amino acid sequences are likely to reflect structural and/or functional homology between NRAMP and the corresponding plant proteins (29). The intriguing link between NRAMP and these plant proteins is under investigation with

respect to the "nitrate transporter" hypothesis for NRAMP function (15).

The juxtaposition of intron positions on the primary amino acid and predicted structural features of the NRAMP polypeptide revealed that the genomic organization of *NRAMP* had a highly modular structure. This is in agreement with previous observations suggesting that introns often mark turns or edges in protein secondary structure, including integral membrane proteins (46). Characterization of the intron/exon structure of the *NRAMP* gene revealed the presence of an unusual alternatively spliced exon contained within the normal intron 4. Nucleotide sequence analysis of intron 4 revealed that this new exon was defined by consensus-splicing signals mapping within the boundaries of an Alu element inserted in a reverse orientation with respect to gene transcription. This Alu element was found to belong to the ancient Sx family of Alu sequences (50% of common Alu) (36). The type of splicing event detected here and restricted to Alu elements inserted in the orientation opposite to transcription has been proposed to be an evolutionarily important mechanism for the generation of protein sequence diversity (49). Such insertion/splicing events have been shown to involve Alu elements belonging to the Sx and J subfamilies, which inserted in the genome at least 40 millions years ago (48), and cryptic splice sites could have been activated within the Alu element as a result of mutations subsequent to insertion (47). Although both alternatively spliced forms were expressed in vivo (Fig. 4 *C*), the physiological consequences of the alternatively spliced mRNA variant remain unclear as addition of the novel exon would result in a severely truncated protein. The predicted polypeptide encoded by the novel exon in *NRAMP* was similar to several other polypeptides surveyed (Fig. 3 *C*). Interestingly, in the case of the human complement decay accelerating factor (37) and integrin β1s genes (38), Alu-mediated alternative splicing events such as that observed here have been shown to give rise to distinct protein isoforms with different subcellular localization and biological activities.

Northern blot analysis revealed *NRAMP* expression in reticulo-endothelial organs and suggested that, by analogy with the mouse, the macrophage would be an important site of *NRAMP* mRNA expression in humans. In support of this proposal was the observation that of four committed progenitors of the monocytic lineage blocked at different stages of differentiation, all expressed *NRAMP* (Fig. 4 *C*). On the other hand, the relative abundance of *NRAMP* mRNA in PBL compared with spleen could not be explained solely on the basis of different numbers of the same cell type (either mature macrophages or other cell type of the monocytic lineage) present at these two sites. In addition, the high levels of expression detected in human lungs is in sharp contrast to the absence of expression of the mouse homolog in the same tissue (15). Several possibilities can be put forward to explain these findings. The first is that macrophages at different stages of differentiation or activation may be present in different amounts at these sites and may themselves express widely different amounts of specific mRNA (49). This is in agreement with our finding that *NRAMP* expression is increased concomitantly with the state of differentiation of the model cell types used in our analysis of the monocytic pathway (Fig. 4 *C*). The second possibility is that more than one cell type expresses *NRAMP*, and that cells of the granulocytic and/or lymphocytic pathways may also express the gene. The absence of *NRAMP* expression in adult thymus (Fig. 4 *A*) argues against a strong level of expression in lymphocytes. The PBL isolation procedure used here depleted most of granulocytes; however we cannot exclude the possibility that the remaining fraction of these phagocytic cells contributed to the signal detected in our analysis. A third alternative is that the human lung RNA sample analyzed here may have originated from a patient with an unusually high proportion of monocytes or other *NRAMP*-positive cells brought about by specific infection, inflammation, or other conditions. Such cells may not be present in normal lung tissue from mice raised in captivity under semisterile conditions. Finally, the levels of the mouse and human *NRAMP* mRNAs may be regulated very differently at the transcriptional and posttranscriptional levels during maturation and/or activation of the macrophage in the two species. This situation is reminiscent of the gene encoding the inducible nitric oxide synthase expression in the macrophage, where strikingly different mechanisms of induction have been proposed for the two genes (50–52). The cloning of genomic sequences reported here for human *NRAMP*, and the identification of cell lines showing differentiation-dependent expression, should help define the *cis*-acting sequences and *trans*-acting cellular factors controlling *NRAMP* gene expression.

The identification of a human *NRAMP* homolog, its chromosomal localization, the characterization of the nucleotide sequence of its cDNA, and the elucidation of the exon/intron structure of the gene reported here, together provide the necessary tools to test the possible role of this gene in differential resistance/susceptibility of humans to infection with intracellular pathogens such as Mycobacteria, Salmonella, and Leishmania. This proposal can now be formally tested by segregation analysis or direct DNA sequencing of the *NRAMP* gene from large informative familial pedigrees from endemic areas of such diseases.

We thank Drs. M. Dubow (McGill University) for preparing stocks of helper phage MK105, J. Price for providing frozen human spleen samples, D. Labuda (Université de Montréal) and P. Chartrand (Canadian Red Cross) for helpful discussions and suggestions during this work and sharing unpublished results, and G. Privé (University of California at Los Angeles) for computing hydrophobic moments.

This work was supported by U.S. Public Health Service grant 1-RO1-AI-35237 to P. Gros and E. Skamene. E. Schurr is supported by a scholarship from the Medical Research Council of Canada. P. Gros is supported by a EWR Steacie Memorial Fellowship from the Natural Sciences and Engineering Council of Canada and is an International Research Scholar from the Howard Hughes Medical Institute.

Address correspondence to Dr. Philippe Gros, Department of Biochemistry, McGill University, 3655 Drummond Street, Montreal, Quebec, Canada H3G 1Y6.

Received for publication 31 May 1994 and in revised form 25 July 1994.

References

1. Bloom, B.R. 1992. Tuberculosis: back to a frightening future. *Nature (Lond.).* 358:538.
2. Stead, W.W. 1992. Genetics and resistance to tuberculosis. Could resistance be enhanced by genetic engineering? *Ann. Intern. Med.* 116:937.
3. Crowle, A.J., and N. Elkins. 1990. Relative permissiveness of macrophages from black and white people for virulent tubercle bacilli. *Infect. Immunol.* 58:632.
4. Schurr, E., D. Malo, D. Radzioch, E. Buschman, K. Morgan, P. Gros, and E. Skamene. 1991. Genetic control of innate resistance to mycobacterial infections. *Immunol. Today.* 12:A42.
5. Schurr, E., K. Morgan, P. Gros, and E. Skamene. 1991. Genetics of leprosy. *Am. J. Trop. Med. Hyg.* 44:4.
6. Skamene, E., P. Gros, A. Forget, P.J. Patel, and M. Nesbitt. 1984. Regulation of resistance to leprosy by chromosome 1 locus in the mouse. *Immunogenetics.* 19:117.
7. Goto, Y., E. Buschman, and E. Skamene. 1989. Regulation of host resistance to *Mycobacterium intracellulare in vivo* and *in vitro* by the *Bcg* gene. *Immunogenetics.* 30:218.
8. Gros, P., E. Skamene, and A. Forget. 1981. Genetic control of natural resistance to *Mycobacterium bovis* BCG. *J. Immunol.* 127:2417.
9. Gros, P., E. Skamene, and A. Forget. 1983. Cellular mechanisms of genetically-controlled host resistance to *Mycobacterium bovis* BCG. *J. Immunol.* 131:1966.
10. Plant, J.E., and A. Glynn. 1976. Genetics of resistance to infection with *Salmonella typhimurium* in mice. *J. Infect. Dis.* 133:72.
11. Bradley, D.J. 1977. Genetic control of *Leishmania* populations within the host. II. Genetic control of acute susceptibility of mice to *L. donovani* infection. *Clin. Exp. Immunol.* 30:130.
12. Skamene, E., P. Gros, A. Forget, P.A.L. Kongshavn, C. St.-Charles, and B.A. Taylor. 1982. Genetic regulation of resistance to intracellular pathogenes. *Nature (Lond.).* 297:506.
13. Malo, D., S.M. Vidal, J. Hu, E. Skamene, and P. Gros. 1993. High resolution linkage map in the vicinity of the host resistance locus *Bcg*. *Genomics.* 16:655.
14. Malo, D., S. Vidal, J.H. Lieman, D.C. Ward, and P. Gros. 1993. Physical delineation of the minimal chromosomal segment corresponding the murine host resistance locus *Bcg*. *Genomics.* 17:667.
15. Vidal, S.M., D. Malo, K. Vogan, E. Skamene, and P. Gros. 1993. Natural resistance to infection with intracellular parasites: isolation of a candidate for *Bcg*. *Cell.* 73:469.
16. Kerppola, R.E., and G.F.-L. Ames. 1992. Topology of the hydrophobic membrane-bound components of the histidine periplasmic permease. Comparison with other members of the family. *J. Biol. Chem.* 267:2329.
17. Bairoch, A. 1991. PROSITE: a dictionary of sites and patterns in proteins. *Nucleic Acids Res.* 19:2241.
18. Unkles, S.E., K.L. Hawker, C. Grieve, E.I. Campbell, P. Montague, and J.R. Kinghorn. 1991. *crnA* encodes a nitrate transporter in *Aspergillus nidulans*. *Proc. Natl. Acad. Sci. USA.* 88:204.
19. Stuehr, D.J., and C.F. Nathan. 1989. Nitric oxide. A macrophage product responsible for cytostasis and respiratory inhibition in tumor target cells. *J. Exp. Med.* 169:1543.
20. Nathan, C.F., and J.B. Hibbs. 1991. Role of nitric oxide synthesis in macrophage antimicrobial activity. *Curr. Opin. Immunol.* 3:65.
21. Malo, D., K. Vogan, S. Vidal, J. Hu, M. Cellier, E. Schurr, A. Fuks, K. Morgan, and P. Gros. 1994. Haplotype mapping and sequence analysis of the mouse *Nramp* gene predict susceptibility to infection with intracellular parasites. *Genomics.* In press.
22. Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular Cloning: A Laboratory Manual. 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
23. Feinberg, A.P., and B. Vogelstein. 1984. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132:6.
24. Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain termination inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463.
25. Epstein, D.J., M. Vekemans, and P. Gros. 1991. *splotch* Sp^{2H}, a mutation affecting development of the mouse neural tube, shows a deletion within the paired homeodomain of Pax-3. *Cell.* 67:767.
26. Noonan, K.E., and I.B. Roninson. 1988. mRNA phenotyping by enzymatic amplification of randomly primed cDNA. *Nucleic Acids Res.* 16:10366.
27. Devereux, J. 1991. The GCG Sequence Analysis Software Package. Version 7.0. Genetics Computer Group, Inc., University Research Park, Madison, WI.
28. Mangalan, H. 1993. Striding the turf of the gang of four. *Trends Genet.* 18:187.
29. Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403.
30. Milosavljevic, A., and J. Jurka. 1993. Discovery by minimal length encoding: a case study in molecular evolution. *Machine Learning.* 12:1.
31. Chirgwin, J.M., A.A. Przybyla, R.J. MacDonald, and W.J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry.* 18:5294.
32. Sheets, M.D., S.C. Ogg, and M.P. Wickens. 1990. Point mutations in AAUAAA and the polyA addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Res.* 13:1905.
33. Schurr, E., E. Skamene, K. Morgan, M.-L. Chu, and P. Gros. 1990. Mapping of *Col3a1* and *Col6a3* to proximal murine chro-

mosome 1 identifies conserved linkage of structural protein genes between murine chromosome 1 and human chromosome 2q. *Genomics.* 8:477.
34. Rousseau-Merck, M.F., D. Simon-Chazottes, M. Arpin, E. Pringeault, D. Louvard, and R. Berger. 1988. Localization of the villin gene on human chromosome 2q35-37 and on mouse chromosome 1. *Hum. Genet.* 78:130.
35. Mount, S.M. 1982. A catalog of splice junction sequences. *Nucleic Acids Res.* 10:459.
36. Jurka, J., and A. Milosavljevic. 1991. Reconstruction and analysis of human Alu genes. *J. Mol. Evol.* 32:105.
37. Caras, I.W., M.A. Davitz, L. Rhee, G. Weddell, D.W. Martin, and V. Nussenzweig. 1987. Cloning of decay-accelerating factor suggests novel use of splicing to generate two proteins. *Nature (Lond.).* 325:545.
38. Languino, L.R., and E. Ruoslahti. 1992. An alternative form of the integrin β_1 subunit with a variant cytoplasmic domain. *J. Biol. Chem.* 267:7116.
39. Wong, P., I.M. MacDonald, R. Sood, R. Pilon, and M. Tenniswood. 1993. Identification and partial characterization of a candidate gene for X-linked retinopathies using a lateral approach. *Genomics.* 15:467.
40. Lubbert, M., F. Herrmann, and H.P. Koeffler. 1991. Expression and regulation of myeloid-specific genes in normal and leukemic myeloid cells. *Blood.* 77:909.
41. Barton, H.C., J.K. White, T.I.A. Roach, and J.M. Blackwell. 1994. NH_2-terminal sequence of macrophage-expressed natural resistance–associated macrophage protein *Nramp* encodes a proline/serine-rich putative *Src* homology 3-binding domain. *J. Exp. Med.* 179:1683.
42. Kyte, J., and R.F. Doolittle. 1982. A simple method for displaying the hydropathy of a protein. *J. Mol. Biol.* 157:105.
43. Eisenberg, D., E. Schwarz, M. Komaromy, and R. Wall. 1984. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* 179:125.
44. Devault, A., and P. Gros. 1990. The mouse *mdr* gene family: two of the three members confer multidrug resistance but with preferential drug specificities. *Mol. Cell. Biol.* 10:52.
45. Boguski, M.S., T.M.J. Lowe, and C.M. Tolstshev. 1993. dbEST-database for "expressed sequence tags." *Nature Genetics.* 4:332.
46. Nathans, J., and D.S. Hogness. 1983. Isolation, sequence analysis, and intron-exon arrangement of the gene encoding bovine rhodopsin. *Cell.* 34:807.
47. Makalowski, W., G.A. Mitchell, and D. Labuda. 1994. Alu-sequences in the coding regions of mRNA: a source of protein variability. *Trends Genet.* In press.
48. Britten, R.J., W.F. Baron, D.B. Stout, and E.H. Davidson. 1988. Sources and evolution of human *Alu* repeated sequences. *Proc. Natl. Acad. Sci. USA.* 85:4770.
49. Rutherford, M.S., A. Witsell, and L.B. Schook. 1993. Mechanisms generating functionally heterogeneous macrophages: chaos revisited. *J. Leukoc. Biol.* 53:602.
50. Schneemann, M., G. Shoedon, S. Hofer, N. Blau, L. Guerero, and A. Schaffner. 1993. Nitric oxide synthase is not a constituent of the antimicrobial armature of human mononuclear phagocytes. *J. Infect. Dis.* 167:1358.
51. Kolb, J.P., N. Paul-Eugene, C. Damais, K. Yamaoka, J.C. Drapier, and B. Dugas. 1994. Interleukin-4 stimulates cGMP production by IFN-γ–activated human monocytes. Involvement of the nitric oxide synthase pathway. *J. Biol. Chem.* 269:9811.
52. Lowenstein, C.J., E.W. Alley, P. Raval, A.M. Snowman, S. Snyder, S.W. Russell, and W.J. Murphy. 1993. Macrophage nitric oxide synthase gene: two upstream regions mediate induction by interferon γ and lipopolysaccharide. *Proc. Natl. Acad. Sci. USA.* 90:9730.