**Mycobacterium bovis** Bacillus Calmette-Guerin and Its Cell Wall Complex Induce a Novel Lysosomal Membrane Protein, SIMPLE, That Bridges the Missing Link between Lipopolysaccharide and p53-inducible Gene, **LITAF**(PIG7), and Estrogen-inducible Gene, **EET-1**

Received for publication, December 26, 2000, and in revised form, March 5, 2001
Published, JBC Papers in Press, March 26, 2001, DOI 10.1074/jbc.M011660200

Yasuhiro Moriaki‡§, Nasim A. Begum‡§, Mika Kobayashi‡§, Misako Matsumoto‡, Kumao Toyoshima‡, and Tsukasa Seya‡§

From the ‡Department of Immunology, Osaka Medical Center for Cancer and Cardiovascular Diseases, Higashinari-ku, Osaka 557 and the §Department of Molecular Immunology, Nara Institute of Science and Technology, Ikoma, Nara 631-0101, Japan

**LITAF** and PIG7 encode an identical protein, and they have recently been reported as lipopolysaccharide and p53-inducible genes, respectively. By using the differential display approach, we identified a **Mycobacterium bovis** BCG cell wall skeleton (BCG-CWS)-inducible gene fragment from human monocytes, showing no homology to any reported gene. Full-length cloning of this fragment reveals the following. 1) The differential display product represents the incomplete 3′-untranslated region of LITAF/PIG7. 2) The coding region of the transcript differs from LITAF/PIG7 due to an absence of a single guanine residue, resulting in a potential translational frameshift. 3) The newly coded protein turns out to be 86% identical and 90% similar to an estrogen-inducible rat gene, **EET-1**. Repeated analysis, expressed sequence tag search, comparison with homologues, and genome sequence analysis confirmed the absence of the single guanine residue. One interesting feature of this protein is that it possesses the RING domain signature and is predicted to be localized in the nucleus. However, detailed analysis together with experimental evidence suggests it is neither a RING family member nor a nuclear protein. Comparison of a total collection of 18 proteins from various species indicates that proteins of this family are small in size and mainly conserved at the C-terminal domain with a unique motif. We characterize this novel protein as an unglycosylated small integral membrane protein of the lysosome/late endosome (SIMPLE) whose expression is elicited in monocytes by live and heat-killed BCG, BCG cell wall complex, lipopolysaccharide, and tumor necrosis factor-α. To our knowledge this is the first report of pathogen-associated molecular pattern (PAMP)-induced differential expression of a lysosomal membrane protein presumably involved in apoptosis.

* This work was supported in part by Organization for Pharmaceutical Safety and Research, grant-in-aid from the Ministry of Education, Science, and Culture, and the Ministry of Public Welfare of Japan. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide (sequence(s) reported in this paper has been submitted to the DDBJ/GenBank™/EBI Data Bank with accession number(s) AB034747.

† Supported by Japan Science and Technology Corp. fellowship.

¶ To whom correspondence should be addressed: Dept. of Immunology, Osaka Medical Center for Cancer and Cardiovascular Diseases, Higashinari-ku, Osaka 557 Japan. Tel/Fax: 81 6 6973 1209; E-mail: tseya@mail.mc.pref.osaka.jp.

This paper is available on line at http://www.jbc.org

© 2001 by The American Society for Biochemistry and Molecular Biology, Inc.

Printed in U.S.A.
TNF-α promoter and is reported to be involved in TNF-α expression during LPS induction (10, 12). The same gene was found to be induced several-fold during p53-mediated apoptosis (11, 13) and was described as PIG7. Our findings raise the possibility that the gene we found to be induced by BCG-CWS and named SIMPLE, which is also a human homologue of rat EET-1 (14), could actually be the LITAF/PIG7. Thus, the induction of this gene in various contexts is intriguing and may not be simply a coincidence; here we describe the detailed analysis of the SIMPLE transcript, including its expression, localization, genomic organization, and its assignment under a new family.

MATERIALS AND METHODS

Preparation and Treatment of Monocytes with BCG-CWS and BCG—Peripheral blood mononuclear cells were isolated by standard density centrifugation with Ficoll-Paque (Amersham Pharmacia Biotech) from 400 ml of citrate phosphate dextrose-supplemented human blood. CD14+ monocytes were separated from peripheral blood mononuclear cells by anti-CD14-coated microbeads and a magnetic cell sorting column (Miltenyi Biotec GmbH). Cells were cultured overnight in 10-cm dishes in the presence of RPMI supplemented with 10% FBS and 2% human AB serum. The next day cells were treated for 8 h with 15 μg/ml of total BCG-CWS prepared in emulsion buffer (PBS containing 1% Drakol, 6VR and 1% Tween 80). Cells in the control plates were treated with 15 μl/ml emulsion buffer for the same time. For heat-killed and live BCG treatment M. bovis BCG Tokyo strain has been used at a concentration of 1 bacillus/monocyte for 5 h. In some experiments immature dendritic cells (iDC) have been used to stimulate with BCG-CWS, and the preparation of iDC was essentially the same as described previously (9).

Differential Display RT-PCR—Total RNAs from BCG-CWS-stimulated and unstimulated human monocytes were isolated using TRIZOL (Life Technologies, Inc.) reagent according to the manufacturer’s instructions. Two micrograms of total RNA from control and from stimulated cells were reverse-transcribed using Superscript RT (Life Technologies, Inc.) by 12 types of T12MN (M-adapter). cDNA was amplified by arbitrary primers (AP) sets selected from RNAmap Kit I and II of 1 bacillus/monocyte for 8 h. In some experiments immature dendritic and unstimulated human monocytes were isolated using TRIZOL and the cytosolic fraction were analyzed directly by immunoblotting. THP-1 cells or RK13 cells expressing SIMPLE were solubilized in 20 ml Triton X-100 non-fat milk and treated with rabbit anti-SIMPLE polyclonal antibody in a 1:50 dilution. After washing, the blots were incubated with horseshad peroxidase-conjugated goat anti-rabbit IgG (Bio-Rad) and developed with ECL (Amersham Pharmacia Biotech). Subcellular Fractionation—THP-1 cells were harvested by centrifugation at 1000 × g and then washed 3 times with PBS. Cell pellets were resuspended in 1 ml PBS supplemented with 10% FBS and 2% non-fat milk and treated with rabbit anti-SIMPLE polyclonal antibody and then washed 3 times with PBS. Cell pellets were resuspended in 50 μl of 0.5% Triton X-114 in PBS for 1 h on ice, and the nuclei and cell debris were removed by centrifugation at 1800 × g for 10 min at 4 °C. The proteins were then partitioned into the detergent and aqueous phases according to the method of Bordier (35). The separated detergent and aqueous phases were analyzed directly by immunoblotting. THP-1 cells or RK13 cells expressing SIMPLE were solubilized in 20 ml Triton X-100 non-fat milk and treated with rabbit anti-SIMPLE polyclonal antibody and then washed 3 times with PBS. After partitioning the supernatant was incubated with 100 microunits of neuraminidase (Sigma) for 1 h at 37 °C followed by 4.5 milliunits of 3°C. Colocalized green (fluorescein isothiocyanate) fluorescence and red (rhodamine B-conjugated goat anti-mouse IgG (Organon Teknika Corp.) was carefully analyzed using a Laser-Scanning Confocal Microscope C1 (Nikon, Japan). The merged images were acquired by confocal laser scanning microscope (Olympus FLUOVIEW). Colocalized green (fluorescein isothiocyanate) fluorescence and red (rhodamine-B) fluorescence appeared yellow in the merged images.

RESULTS

BCG-CWS-induced Differential Display RT-PCR Product, GCAP2, Defines the 3′ End of LITAF/PIG7—Differential display RT-PCR has been performed between unstimulated and BCG-CWS-stimulated monocyte RNA. A combination of pairs

- BCG, LPS, TNF-α, p53 and Estrogen-inducible Gene
- Constructs and Transfection—The coding region of SIMPLE (234–719 bp) was cloned into two mammalian expression vectors pEGFP-C1 (CLONTECH) and pcDNAs (Invitrogen), respectively. The primer combinations used for cloning into XhoI and HindIII sites of pEGFP-C1 were 5′-ctc gag eca cca tga tgt eeg cag cct a-3′ and 5′-aag ctt cta cca aag cct tga ggt gc-3′. Primers containing BamHI and EcoRI sites, 5′-gga tct cta cca aag ctt cta cca aag cct tga ggt gc-3′, respectively, were used for cloning into pcDNA3. All DNA constructs were checked by sequencing and transfected into various mammalian cell lines (HeLa, COS-7, and RK13) with Lipofectant reagent (Life Technologies, Inc.).

Immunoblotting—For detection of SIMPLE protein, SIMPLE-transfected RK13 cells, human monocytes, HeLa, THP-1 and DLD-1 cells were lysed in cell lysis buffer containing 1% Nonidet P-40, 10 μl ETA, 140 μl NaCl, 20 μl Tris-HCl, pH 7.4, 1.0 μl phenylmethylsulfonyl fluoride, and 5 μl iodoacetamide. The cell lysates were solubilized either in reducing or non-reducing sample buffer and resolved in a 12.5% SDS-PAGE and transferred onto polyvinylidene difluoride membranes (Millipore, Bedford, MA). The blots were blocked with 10% non-fat milk and treated with rabbit anti-SIMPLE polyclonal antibody at a dilution of 1:5000. After washing, the blasts were incubated with horseradish peroxidase-conjugated goat anti-rabbit IgG (Bio-Rad) and developed with ECL (Amersham Pharmacia Biotech).

Subcellular Fractionation—THP-1 cells were harvested by centrifugation at 1000 × g and then washed 3 times with PBS. Cell pellets were resuspended in 1 ml PBS supplemented with 10% FBS and 2% non-fat milk and treated with rabbit anti-SIMPLE polyclonal antibody in a 1:50 dilution. After centrifugation at 1800 × g for 10 min at 4 °C. The proteins were then partitioned into the detergent and aqueous phases according to the method of Bordier (35). The separated detergent and aqueous phases were analyzed directly by immunoblotting. THP-1 cells or RK13 cells expressing SIMPLE were solubilized in 20 ml Triton X-100 non-fat milk and treated with rabbit anti-SIMPLE polyclonal antibody and then washed 3 times with PBS. After partitioning the supernatant was incubated with 100 microunits of neuraminidase (Sigma) for 1 h at 37 °C followed by 4.5 milliunits of O-glycanase (endo-α-N-acetylgalactosaminidase; Genzyme, Cambridge, MA) treatment for 16 h at 37 °C. The pellets were fixed for 1 h with 0.5% paraformaldehyde in PBS and were permeabilized with 0.5% saponin, 1% BSA/PBS for 30 min, washed 4 times with PBS. After soaking in 1% BSA/PBS, the cells were treated for 1 h at room temperature with rabbit anti-SIMPLE polyclonal antibody or pre-immune rabbit anti-serum in 1% BSA/PBS at a dilution of 1:500. The cells were washed with 1% BSA/PBS and treated for 30 min with fluorescein isothiocyanate-conjugated goat anti-rabbit IgG (1:100) (Organon Teknika Corp., West Chester, PA) diluted in 1% BSA/PBS. Cells were washed 4 times with 1% BSA/PBS and incubated for 1 h with mouse anti-LAMP-1 antibody (5 μg/ml) in 1% BSA/PBS, washed again, and incubated for 30 min with rhodamine B-conjugated goat anti-mouse IgG (Organon Teknika Corp.) (1:100). Visualization of acid pH compartments was performed by staining cells with 75 nM Lysotracker Red DND-99 (Molecular Probes, Leiden, The Netherlands) for 1 h at 37 °C. Single- or double-stained cells were examined with a confocal laser scanning microscope (Olympus FLUOVIEW). Colocalized green (fluorescein isothiocyanate) fluorescence and red (Lysotracker or Rhodamin-B) fluorescence appeared yellow in the merged images.
of primers, T12GC (3' anchor) and AP2 (5' arbitrary), produced a distinct band designated as GCAP2 in BCG-CWS induced RNA (Fig. 1A). Amplification and sequencing of this band identified a 373-bp fragment having identity, with a few acceptable mismatches, to the paired primers at the ends (Fig. 2, boxed sequence). This fragment did not show any homology with any known genes in the data base; however, we found several human ESTs. Furthermore, the fragment when used as a probe showed a distinct signal on Northern blots, suggesting the presence of poly(A) tail. We then employed the in silico cloning strategy (16, 17) to generate an extended virtual contig with a group of EST clones that were directly, or through another overlapping EST, linked to the GCAP2 sequence. This procedure identified a human EST (AW022014) that bridged between the GCAP2 fragment and ESTs denoted by the LITAF/PIG7-end sequence. Sets of EST contigs for this zone are also currently available under a TIGR transcript (THC116999, October, 2000) at NCBI. In order to verify this finding, RT-PCRs with 5' primers from LITAF and 3' primers from the GCAP2 fragment and 5' rapid amplification of cDNA ends from GCAP2 were conducted. Various combinations of RT-PCR primer pairs yielded the expected size bands that were further verified by sequencing and confirmed the results. A representative result of RT-PCR is shown in Fig. 1B. The newly identified 3' region exhibited a polyadenylation signal (Fig. 2) located 18 bp upstream to the beginning of the poly(A) tail shown. No such polyadenylation signal has been detected in the published LITAF sequence or in the deposited PIG7 sequence. The poly(A) tail consisting of 19 A residues described for LITAF does not seem to be the true poly(A) tail, rather the sequence corresponds to a stretch of 19 As (1754–1777 bp) interrupted by a G in the SIMPLE cDNA (Fig. 2). The presence of the same A stretch in both cDNA and in genomic sequences (exon 4, Fig. 3C) further emphasizes that the long A-tailing starting at 1755 bp in the LITAF sequence is an intra-cDNA region, whereas the poly(A) tail shown for SIMPLE was not derived from the genome. Thus, the GCAP2 fragment defines the 3' end of the LITAF/PIG7 transcript, which in turn leads us to conclude that LITAF/PIG7 is a differentially expressed gene in BCG-CWS-treated human monocytes.

Identification and Verification of a Translational Frameshift in LITAF/PIG7—During sequencing of several RT-PCRs and 5' rapid amplification of cDNA ends products from human monocyte RNA, we consistently noticed a single G was missing from a stretch of 5 consecutive guanine residues present in the coding region of LITAF (608–612 bp; GenBank/EMBL/DDBJ accession number NM_004862 or U77396/PIG7 (454–458 bp; GenBank/EMBL/DDBJ accession number AF010312). Absence of this G residue created a translational frameshift (Fig. 3B) yielding a protein of 161 amino acids, whereas LITAF/PIG7 encoded a protein of 228 amino acids. This raised a potential question whether there were two types of transcripts producing two different proteins. In order to verify this further, first we used Pfu polymerase to amplify the coding region from various RNA sources that included monocytes from different donors’ blood, THP-1 monocytic cells (from which LITAF has been isolated), and DLD-1 colon carcinoma cells (from which PIG7 has been cloned). After cloning the PCR products in each case, we sequenced the critical region for 10–12 independent clones and detected the stretch of 4 Gs (Fig. 3A, top) instead of 5 Gs in all of them. Second, we looked for ESTs specific for that region because many tissues express the gene, and if the zone is polymorphic that could be represented by some of the ESTs. We found many ESTs (Fig. 3A) with 4 Gs but could not detect any human or murine EST showing 5 Gs for that particular region, neither in the EST database nor in the pooled EST set for LITAF/PIG7 under UniGene collection (Hs.76507, October, 2000). Third, we wanted to verify the region from the genomic sequence. For this purpose, we utilized the information from the unfinished genome sequencing project through the HTG database at NCBI. We partially succeeded, but a few contigs remain yet to be aligned, identifying the genomic boundaries (Fig. 3C) for the entire SIMPLE transcript from a chromosome 16 clone, RP11-547D14 (accession number AC007616.3). The astonishing observation was that the region of dispute, 4 Gs or 5 Gs, falls into a potential splice junction (Fig. 3, B and C). According to the splice junction donor acceptor rule (AG/ GT), it is clearly in favor of the presence of 4 Gs in the cDNA. Finally, we focused on the deduced amino acid sequence of the transcript. If the protein with the changed C-terminal amino acid truly represented a conserved protein, then we might find other homologues and orthologues in the database. Due to the advantage of the small size of SIMPLE, we easily could find full-length or near full-length proteins from various species described below (Fig. 9). The C-terminal domain was found to be the most conserved region in a number of proteins from different species, indicating the evolutionarily conserved feature of this domain. On the other hand, we failed to generate any such data for the C-terminal domain specific for LITAF/PIG7. Based on the above four lines of evidence, we conclude that the 161-amino acid sequence coding the SIMPLE gene is the natural and the most abundant transcript. For clarity, from this point, the 161-amino acid-coded 2368-bp transcript will be referred as SIMPLE. For comparison the first base number of SIMPLE cDNA (Fig. 2) has been kept the same as LITAF.

SIMPLE Is a Widely but Variably Expressed Transcript—Multiple tissue Northern blots from CLONTECH were independently hybridized with full-length cDNA, the coding region, and with the GCAP2 fragment. Three types of hybridizations successively detected a single transcript of the same size on multiple blots. The result obtained by hybridizing with the coding region of SIMPLE is presented in Fig. 4. Most of the
human tissues except testis expressed SIMPLE abundantly, and this was also consistent with the fact that there was a vast collection of human ESTs (UniGene Hs.76507, October, 2000) from various organs and a single EST (AA62566) of testis origin. In respect to the RNA size marker provided on the multiple tissue Northern blot, the message size fits reasonably with the full-length SIMPLE (2.4 kb). The message size appeared to be the same in a number of human cell lines, and in 12 paired tumor-normal colon carcinoma samples (data not shown), suggesting there is no aberrant or variant transcript that could be detectable by size difference on Northern blot. The tissue distribution pattern also suggests SIMPLE probably has a more generalized function rather than playing a unique role for the sole benefit of the monocyte/macrophage lineage. However, there is a great variability in the relative expression level among the tissues tested; peripheral blood lymphocyte showed the highest level of expression: little in brain and, compared with ovary, little in testis. Inter-library comparative profile of the SAGE-tag, TGAATACTAC (Fig. 2), indicates that the expression could be regulated by a hormone in the human reproductive tissues: as evident in LNCaP with DHT versus MCF7-estradiol 3 h. The SAGE data further demonstrate that the transcript might be differentially expressed in certain malignancies, such as prostate, breast, and ovary.

Expression of SIMPLE Is Induced by BCG-CWS and BCG—Often differential display fragments gives rise to false positives, so it was essential to reverify the differential expression in response to stimulation. In different batches of monocytes and iDC, BCG-CWS enhanced the expression of SIMPLE about 2-fold at 8 h of induction (Fig. 5A), the same incubation time for the differential display study. The expression has also been checked in response to heat-killed and live M. bovis BCG; both are found to be potent inducers of this gene (Fig. 5B). M. bovis BCG doubling time in monocytes was 20 h, suggesting the induction was replication-independent. However, a detailed study is required to confirm whether the expression was phagocytosis-dependent or the expression undergoes alteration with bacilli growth.

SIMPLE Versus TNF-α Induction in Human and Murine Monocytes—BCG-CWS and BCG are both potent inducers of TNF-α, and LITAF has been described as a novel regulator of LNCaP without 5α-dihydrotestosterone or MCF7 3 h versus MCF7-estradiol 3 h. The SAGE data further demonstrate that the transcript could be differentially expressed in certain malignancies, such as prostate, breast, and ovary.
TNF-α expression during LPS stimulation. We wanted to see if there was any correlation of SIMPLE and TNF-α expression in our experimental condition. Human monocytes and murine RAW cells were stimulated by LPS and BCG, and at different time points RNA was prepared. The blots were first probed with species-specific SIMPLE cDNA and then stripped and reprobed with species-specific TNF-α cDNA (Fig. 6). In human monocytes BCG-CWS as well as LPS both enhanced the SIMPLE expression, but the induction was faster in the case of LPS as evident from the peak expression levels, 2 versus 4 h by LPS and BCG-CWS, respectively (Fig. 6A). The induction of TNF-α expression was prior to the SIMPLE induction, suggesting TNF-α itself could be an inducer. This point was further verified by stimulating monocytes with recombinant TNF-α (Fig. 6B); however, the induction was not as robust as found in LPS treatment. In the case of a murine macrophage cell line, RAW, expression of SIMPLE gradually increased with a peak level at 24 h by BCG-CWS, whereas induction by LPS was observed at
earliest time points (Fig. 6C). Despite the degradation of RNA in the control lane of LPS, it was not difficult to conclude that the expression of TNF-α was prior or concomitant to the SIMPLE induction as observed in human monocytes. Again, rapid induction by LPS and gradual induction by BCG-CWS for SIMPLE message has been reflected in RAW cells, suggesting PAMPs derived from mycobacteria and Gram-negative bacteria differentially modulate the expression of SIMPLE. Interestingly, RAW cells showed a second transcript of ~1.5 kb and that was also altered due to the induction. We also tested THP-1 cells under similar conditions, but we could not detect significant differences in SIMPLE expression between stimulated and unstimulated cells. THP-1 cells itself had a good basal expression level that we found to be reduced upon phorbol 12-myristate 13-acetate treatment and cannot be elicited further by LPS/BCG-CWS treatment (data not shown).

Sequence Analysis of SIMPLE—Next we focused on structural analysis of the deduced amino acid sequence of SIMPLE to characterize its possible function. One easily noticeable feature of the deduced amino acid sequence of SIMPLE is its proline-rich N terminus and the cysteine-rich C terminus (Fig. 2). The total proline content is 15% which is about 3 times higher than the proline content typically found in eukaryotic proteins. The majority (87%) of the prolines are concentrated in the N-terminal half; the proline content in this region exceeds 22% and is completely devoid of any cysteine. Similarly the C-terminal half of 68 amino acids lacks proline abundance and possesses 11 cysteines. Unlike many proline-rich proteins, it does not have many glutamines; instead, it is often punctuated by serine-threonine and to some extent by tyrosine, and the majority of these residues are present in the N-terminal half. Proline-rich proteins with repetitive or non-repetitive motifs are found to be involved in various functions (18, 19). However, the overall pattern of proline in combination with serine, threonine, and especially with tyrosine, SIMPLE shows similarity with tyrosine-hydroxyproline-rich extensin family of plant cell wall proteins elicited during infection and wounding (20).

The most striking feature of SIMPLE is the C-terminal domain; motif search and profile scan indicate that the amino acid residues 96–152 of SIMPLE are similar to C3H4 type zinc RING finger (21, 22). As shown, cysteines (Fig. 9A, light green vertical bars) and possible loop1 and loop2 regions of SIMPLE can be aligned with a group of known RING finger proteins (Fig. 9, A and B). The alignment fits well with the consensus of RING and allied zinc finger motifs (23) such as FYVE and FYVE related fingers (24, 25). The protein does not have any known nuclear localization signal sequence, but the k = 9/23 in k-NN prediction (PSORT) suggests a 52% probability for nuclear localization. The presence of proline-cysteine-rich domains are the characteristics of several RING proteins including many transcription factors (26–28), and they are found to be involved in diverse cellular functions through RNA or DNA binding, protein-protein interaction, or both (23, 29–31).

Apparently SIMPLE has all the features to qualify as a RING protein and to be involved in transcriptional regulation. But the hydropathy profile (Fig. 8A) suggests SIMPLE has a potential TM domain, and the region lies within the predicted RING structure (Fig. 9A). The sequence has been analyzed through several transmembrane prediction programs available at the Expasy site and is detected as an integral membrane protein with the same TM domain. The SOSUI-predicted TM-spanning region of 23 amino acid (Fig. 2) residues is long enough to conform stable integration in the membrane. Another intriguing feature of this C-terminal domain is the presence of a di-leucine motif and a XXXΦ motif (Fig. 2), where Χ is any amino acid and Φ is a bulky hydrophobic amino acid. These motifs are known to interact with a family of adapter protein complexes during intracellular sorting of the transmembrane proteins and membrane receptors and also to function in lysosomal/endosomal and trans-Golgi network targeting (32, 33). The above analysis prompted us to examine whether SIMPLE

Fig. 5. A, confirmation of the BCG-CWS-mediated induction of SIMPLE in human monocytes and iDC cells. Monocytes or iDCs of various batches were stimulated by BCG-CWS or emulsion buffer (control), and after 8 h total RNA was harvested for Northern analysis. B, monocytes were treated with either heat-killed or live M. bovis BCG at a concentration of 1 bacillus/monocyte, and after 8 h post-infection cells were washed, and RNA was prepared for Northern hybridization. Hybridization signals were quantitated by NIH image program, and normalized values were plotted.
is a member of the nuclear or cytoplasmic RING family proteins or a completely new type of membrane protein localized in the cell surface or intracellular vesicular compartment.

**Subcellular Localization of SIMPLE**—We first took the routine approach of constructing GFP fusion proteins and prepared N-terminally GFP-tagged versions of SIMPLE as it lacks the signal sequence. Expression in 3 different cell lines, COS-7, NIH3T3, and HeLa, produced comparable results. A representative example of expression in COS-7 cells is shown in Fig. 7A; N-terminally GFP-fused SIMPLE was concentrated in a paranuclear position with a number of vesicles surrounding the nucleus. Expression of the GFP-only vector showed diffuse cytoplasmic and nuclear fluorescence (data not shown). A paranuclear localization position is mainly suggestive of Golgi, however, and the trans-Golgi network, newly synthesized transport vesicles, and lysosomes are also present around this location, and overexpression makes this particularly difficult to ascertain. Therefore, a polyclonal antibody was generated against bacterially expressed His<sub>-</sub>SIMPLE and used, after checking the specificity, for immunostaining in HeLa cells. Results of confocal analysis are shown in Fig. 7B; SIMPLE is distributed with LAMP-1 and LysoTracker-stained vesicular compartments, indicating SIMPLE is associated with perinuclear lysosomes and late endosomes. Cells with or without permeabilization were also analyzed by fluorescence-activated cell sorter; no definitive staining was observed in the plasma membrane or cells stained with preimmune serum, whereas permeabilized cells showed a strong fluorescence shift (data not shown).

In parallel, subcellular fractionation was carried out to check whether the immunostaining pattern corresponds. As evident from the Western blot of various subcellular fractions, the lysosome-enriched 8,000 × g pellet (Fig. 7C, lane 2, top) contained SIMPLE; the same fraction showed the strongest signal for the presence of LAMP-1-positive vesicles (Fig. 7C, lane 2, below). Absence of any cell surface expression further suggests that SIMPLE is a lysosomal/late endosomal residence protein rather than a recycling receptor. This work clearly demonstrated that SIMPLE is not a nuclear or cytosolic protein even though most of the RING proteins are found to be lysosomal/late endosomal residence proteins.

**SIMPLE Is an Integral Membrane Protein of the Lysosome**—RING proteins like EEA1 are localized in early endosomes and utilize their RING domain for peripheral membrane anchoring (34). To see whether SIMPLE is an integral membrane protein of the lysosome, we have analyzed the partition of the protein during phase separation in a solution of Triton X-114 according to Bordier (35). In this method integral membrane proteins with an amphiphilic nature are recovered in the detergent phase, whereas peripheral and cytosolic proteins remain exclusively in aqueous phase. After two rounds of Triton X-114 extraction, the protein was completely recovered in the detergent phase (Fig. 8A), demonstrating that SIMPLE represented an integral membrane protein of the lysosome. Simultaneously, a duplicate blot also has been tested with a control 7-TM integral membrane protein to monitor the efficiency of extraction (data not shown).

Lysosomal membrane proteins are in general heavily N-glycosylated to be protected from unwanted degradation inside the lumen. SIMPLE does not have any N-glycosylation sites.
although it has several O-glycosylation sites as predicted by the NetOglyc 2.0 program (Fig. 8B). In order to verify its glycosylation status, we used THP-1 cells as a natural source of the protein and SIMPLE cDNA-transfected rabbit RK13 cells for expressed protein. Expressed SIMPLE or THP-1-derived protein remained unaffected by O-glycosidase (Fig. 8B; THP-1 data not provided). Under similar conditions the deglycosylation status was observed for a known O-glycosylated protein (CD46), suggesting that SIMPLE is an unglycosylated protein, which is consistent with the fact that the O-glycosylation sites are poorly defined and not necessarily used (36). The protein is approximately 24 kDa (in both reduced and non-reduced condition) as detected in THP-1, human monocytes, HeLa, DLD-1, and in RK13-transfected cells. However, the unglycosylated molecular size (24 kDa) of SIMPLE is slightly higher than its unmodified calculated mass of 17 kDa; the slower migration could be attributed to the phosphorylation status or due to the proline richness of the protein as noted for other proteins such as Zyxin, Krupple, and TESK-1 (37, 38).

Assignment of SIMPLE into a New Family—As mentioned above, the C-terminal domain of SIMPLE resembles the RING structure but interrupted by a TM region. That is an unusual structural feature unable to satisfy RING family characteristics, yet unique by itself. We wanted to know whether proteins containing this structural feature compose a new family, and we searched data bases to collect SIMPLE homologues and related proteins (see Fig. 9). The homology among human (SIMPLE), rat (EET-1; accession number AF53184), and murine (TBX1; accession number AF17100) is high (90–91%). A considerable degree of homology is present with Zebrafish (53%; EST AW184464) and chicken (72%; EST AI979890), and the most conserved region appeared to be the C-terminal domain 70–75 residues long (N-terminal alignment for these proteins has not been shown). Based on this region of SIMPLE, we performed TBLASTN and BLASTP searches against several data bases and 2 rounds of PSI-BLAST iteration. Several hits appeared in these query modes, and simple visual inspection could identify that they have a pattern. An alignment of 18 sequences from all the species (2 human, 2 rodent, 1 fish, 1 avian, 5 insects, 8 nematode, and 1 from plant) is provided in Fig. 9C. Transmembrane prediction analysis was done for each protein, and strong TM regions were underlined whenever detected. It is now more convincing that this domain has a consensus (shown below the alignment profile) to be clearly distinguishable from the RING-like domain, yet borrowing the first and last pair of cysteines conserved among the zinc finger family. Between the cysteine dyad there is a long variable region that often harbors the membrane-spanning region. The TM region is preceded and followed by two unique consensus sequence signatures, CPXCYX_T and #X,#X,HXCYX_C, respectively (Fig. 9C). The majority of the proteins in this family seem to be small in size (around 160 aa); however, we can see the domain and the motifs in proteins of larger sizes such as C16orf5 (261 aa) and DmCG13515 (283 aa). The domain is not necessarily restricted to the C terminus of all proteins as in the case of C16orf5 (39) and CeT26805 (see TL/CL), suggestive of a module domain that could be utilized by a variety of proteins at different locations but serving a common function. In this connection Caenorhabditis elegans protein CeT26805 of 386 amino acids can be mentioned. The N-terminal region residues 31–98
of this protein show the SIMPLE-like domain signature, and the rest of the sequence (105–386 residues) is 70% similar to the WD repeat region of human \( \text{bCOP} \). CeT26805 is a hypothetical protein, yet it is a good example where the SIMPLE domain has been fused to the WD domain generating a new protein. Based on our analysis above we propose to designate this family, the domain, and the motif by the name of SIMPLE.

A plant protein has been shown as an example, just below the consensus, to show that the hypothetical protein follows the consensus pattern and can be considered as a member of this new family (Fig. 9C).

**DISCUSSION**

The identification of SIMPLE, which is similar to \( \text{LITAF} \) or \( \text{PIG7} \) transcripts but with a different coding potential, was an accidental finding. We confirmed in various ways the presence of SIMPLE as a single transcript and protein. Our main findings are as follows: 1) identity of SIMPLE with \( \text{LITAF}/\text{PIG7} \) at the nucleotide level but not at the level of coded protein; 2) perfect agreement of all exon-intron junctions in the SIMPLE transcript but not in \( \text{LITAF}/\text{PIG7} \); 3) lack of evidence for the presence of 5 Gs in the coding region sequences from monocytes, THP-1, and DLD-1 cells; and 4) an abrupt change (Fig. 9A) in the amino acid sequence in \( \text{LITAF}/\text{PIG7} \), compared with SIMPLE, EET-1, and TBX1, supports a frameshift in the LITAF due to the misincorporation of an additional G residue. However, the region of dispute corresponds to a splice junction; aberrant splicing or allelic polymorphism of 4 Gs versus 5 Gs may still create a \( \text{LITAF}/\text{PIG7} \)-coded protein, and it could act as a dominant negative form against the natural version, SIMPLE.

**Similarity with RING Family**—The next question is whether SIMPLE can be considered as a variant of the zinc RING proteins because it has a similar sequence motif. We have not examined the zinc binding potential of this protein due to the difficulties of purifying its native form, and in addition, the predicted RING region has been found to be disrupted by a single potential TM domain. Our experimental evidence, including the phase separation, supports that the protein is tightly fastened in the intracellular membrane compartment, suggesting the predicted TM domain within the RING is the anchoring region. The TM domain signature also corresponds to the BLOCK pattern of the 5th TM domain of the \( \text{Srg} \) family 7-transmembrane receptor (40) of \( \text{C. elegans} \), which further supports the integral nature of the domain. The presence of this TM domain dampens the possibility of considering this protein as a RING protein or even to be a divergent type. However, it is apparent from the alignment that the C-terminal domains of SIMPLE family genes are nicely bracketed by a pair...
FIG. 9. A, ClustalW alignment of SIMPLE with LITAF/PIG7, EET-1 (rat homologue), and TBX1 (murine homologue). Identical residues are in red with an asterisk; strongly similar residues are in green with a colon, and weakly similar residues are in blue with a single dot. From residue 127 of LITAF/PIG7 that corresponds to the region of 5 Gs in the coding sequence, there is an abrupt change (sequence in black italics) in amino acid sequence.

B, cysteine residues in the C-terminal domain of SIMPLE, EET-1, and TBX1 correspond to the conserved positions (numbered in magenta text, 1–8) of cysteines and histidine in the RING domain. The aligned RING proteins were collected from Refs. 24 and 64. Below these sequences the consensus of RING family and related zinc finger domains (23, 25) are also shown in magenta. The number of amino acid residues between the conserved cysteines and histidine is omitted for the alignment purpose and is shown as dashed lines.

C, alignment of the C-terminal domain of human, mouse, rat, chicken, zebrafish, Drosophila melanogaster, and Drosophila virilis. The consensus sequence and the corresponding domain motifs are shown.
of CXCC motifs (Fig. 9C, green bars), resembling the first and last pair of cysteines in zinc finger proteins. This feature, together with the fairly well spaced additional cysteines within the bracket, can easily be mistaken as a RING-like contour. We conclude from the alignment profile that the proteins under this family share a domain that has similarity in organization with the RING domain; however, the presence of the TM domain limits further comparison.

Comparison with Major Lysosomal Membrane Proteins—SIMPLE is a new lysosomal membrane protein, a motif search showed that the N-terminal 14 amino acids (residues 10–23) of SIMPLE have homology with the LAMP block. We compared SIMPLE with major integral membrane proteins of lysosomes (41), LAMPs, LIMPgs, and also with Endolyn (42, 43). ClustalW alignment showed patches of sequence similarity with those groups mainly due to the proline, serine, and threonine (Pro > Ser = Thr) richness of the N-terminal domain of SIMPLE. The partially matched regions correspond to the mucin-like domains (Ser-Thr-rich) of Endolyn, CD168, and DC-LAMP and the hinge regions (Pro-Ser/Ser-Thr-rich) of LAMPs. There is little architectural similarity between SIMPLE and classic bipartite or semi-bipartite patterns of the extracellular domain of LAMP or the Endolyn family (41, 43). Mucin-like domains and hinge regions of the above families of proteins are heavily N- and O-glycosylated (41, 44), whereas SIMPLE is completely devoid of glycosylation. The C-terminal domain also does not show any similarity except for the presence of dileucine (LL) and YXXΦ motifs, one of which is invariably present in the cytoplasmic domain of the above-mentioned families. In the case of the LAMP family, YXXΦ is preceded by a G and is known to be critical for direct delivery from trans-Golgi network to lysosome (45). SIMPLE, EET-1, and TBX1 all possess the YXXΦ motif with a GT prefix, although the most conflicting and contradictory aspect is the predicted type II orientation of SIMPLE. SIMPLE lacks the typical signal sequence and contains only one potential hydrophobic stretch that presumably works as a stop transfer signal. According to the charge difference rule (46) of TM topology the Δ(C-N) value is −2.5, indicating that the dileucine and YXXΦ motifs will be in the luminal side. The charge difference is not the sole factor involved in membrane orientation; type II proteins may have a type III configuration (47, 48), and in that case LL/YXXΦ will be retained in the cytoplasmic tail, which remains to be experimentally verified for SIMPLE.

Induction of SIMPLE by Microbial Components—BCG cell wall components and LPS are potent effectors exerting maturation and survival signals for monocytes and dendritic cells, and these pathways are intimately linked to TLR2- and TLR4-mediated signaling (49). TLR2 has been characterized recently as a novel death receptor (50) that implies an analogous mechanism to the TNF receptor family; the signaling events of these receptors are bifurcated downstream, and they can modulate life and death upon ligand activation during infection. The genes regulated through these pathways leading to the final cellular response are mainly unknown. The putative promoter region of SIMPLE contains AP1- and p53-like binding motifs; if those are functional this could explain how SIMPLE could be induced through TLR/TNF receptor and p53-mediated pathways.

Relevance of SIMPLE Induction by M. bovis—During microbial infection one of the most predominant innate responses exerted by an immunocompetent host is the induction of apoptosis of the infected cells thereby minimizing the spread and restricting the infection. The similarity of SIMPLE with PIG7, which is more than 10-fold increased in a p53-mediated apoptotic environment, and its localization in lysosomes, which play role in the process of cell self destruction, suggest SIMPLE could be involved in host cell apoptosis. In vitro studies have shown that apoptosis is responsible for intracellular killing of mycobacteria (51), and down-regulation of anti-apoptotic genes has been observed due to BCG or heat-killed Mycobacterium tuberculosis (52). Since promoting apoptosis is not beneficial for the growth of the bacilli, the components of programmed cell death are also impaired. It is evident from recent work that both live and heat-killed M. bovis BCG were capable of increasing the viability of monocytes through up-regulation of an anti-apoptotic gene A1 (53). Hence, it is not unlikely that many genes in apoptotic and anti-apoptotic pathways could be altered during mycobacterial infection. Several lines of evidence suggest that avirulent strains of Mycobacterium are most active in eliciting the apoptotic response, whereas the virulent strain bypasses this, and cells remain less apoptotic (54, 55). If the expression of SIMPLE is potentially connected to elicit the host cell apoptosis, the expression could be affected differentially by virulent and avirulent strains.

At this point, there is no direct evidence that PIG7 or SIMPLE is involved in apoptosis because their mechanism of action is unknown. However, SIMPLE as a lysosomal membrane protein and being proline-rich may raise an interesting possibility in the view of lysosomal and ubiquitin-mediated intracellular protein degradation pathways (56–58). Lysosomal membrane protein LAPT5M (59) that is specifically expressed in hematopoietic cells possesses a proline-rich carboxyl-terminal domain, and the domain has been found to interact with precursors of ubiquitin, leading to the concept that LAPT5M mediates degradation of ubiquitinated protein in the lysosome. Another lysosomal membrane protein LAMP2/ Lgp96 (60) has been found to be a receptor for selective uptake of proteins into the lysosome and subsequent degradation. Recently, a unique sequence motif has been detected in the cytosolic tail of the LAMP2a isoform, which is required for the binding of substrate protein and is proposed to be important for chaperone-mediated autophagy by the receptor (61). Structural analysis of the proline-rich domain and the identification of interacting proteins for SIMPLE may provide important clues about the function of this gene. It is also intriguing that the rat homologue, EET-1, was rapidly induced by estrogen treatment in the rat uterus. Programmed cell death is an essential feature of normal ovarian and uterine cycles (62, 63), and increased lysosomal activity is known during endometrial/luteal degeneration. Studying rodent reproductive tissue may reveal the functional aspect of SIMPLE as it lacks estrogen-responsive elements in the 3′-untranslated region, and its regulation by estrogen in human has yet to be defined.
In summary, identification of SIMPLE revealed that LITAF/PIG7 could encode the same protein as EET-1 provided a G residue was absent from a specific region of the coding sequence. We confirmed the absence of the G residue in an identical transcript. SIMPLE belongs to a new family of proteins defining its novel role in programmed cell death. There is little information regarding the role of lysosomal membrane proteins in apoptosis and their alteration during infection. Characterization of SIMPLE as a novel member of lysosomal membrane proteins in apoptosis and their alteration during infection is of interest.

Acknowledgments—We thank Dr. Graham F. Barnard, University of Massachusetts Medical Center, for providing the colon carcinoma biopsies and for critical discussions. We also thank Dr. J. M. Chakrabartty, University of Illinois, and Dr. I. Azuma, Hakodate National College of Technology for helpful discussions. We are grateful to Dr. A. Hayashi, Dr. H. Koyama, and Dr. M. Tatsuta (Osaka Medical College) for valuable discussions. We also gratefully acknowledge the Note Added in Proof—Type II topology of SIMPLE has been confirmed. The preparation of the photographic reprints by K. Okugawa. Dr. Nasim A. Begum conducted the main part of this work.

REFERENCES

1. Hayashi, A. (1994) Proc. Jpn. Acad. 70, 205–209
2. Hayashi, A., Doi, O., Azuma, I., and Toyoshima, K. (1998) Proc. Jpn. Acad. 74, 50–55
3. Medzhitov, R., and Janeway, C. A., Jr. (1998) Curr. Opin. Immunol. 10, 12–15
4. Aderem, A., and Ulevitch, R. J. (2000) Nature 407, 762–767
5. Tsuji, S., Matsumoto, M., Takesaki, O., Akira, S., Azuma, I., Hayashi, A., Toyoshima, K., and Saya, T. (2000) Infect. Immun. 68, 4882–4889
6. Saya, T., Begum, N. A., Nomura, M., Tsuji, S., Matsumoto, M., Hayashi, A., Azuma, I., and Toyoshima, K. (2000) Adv. Exp. Med. Biol. 465, 229–237
7. Hayashi, A., Tsuji, S., Nomura, M., Amano, K., Azuma, I., Hayashi, A., Matsumoto, M., Saya, T., and Toyoshima, K. (1999) Biochim. Biophys. Res. Commun. 256, 325–329
8. Hoffmann, J. A., Kafatos, F. C., Janeway, C. A., and Ezekowitz, R. A. (1999) Science 284, 1313–1316
9. Janeway, C. A., Jr., and Medzhitov, R. (1999) Curr. Biol. 9, R873–R882
10. Miyaki, F., Takashiba, S., Lebo, R., and Amano, K. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 4518–4523
11. Poyarkov, K., Xia, Y., Zweier, J. L., Kinzler, K. W., and Vogelstein, B. (1999) Nature 399, 300–305
12. Takahashi, N., Van Dyke, T. E., Shapira, L., and Amano, K. (1995) Infect. Immun. 63, 1519–1534
13. Zhu, J., Jiang, J., Zhou, W., Zhu, K., and Chen, X. (1999) Oncogene 18, 2149–2155
14. Everett, L. M., Li, A., Devaraju, G., Caprelli-Grant, A., and Biggs, R. M. (1997) Endocrinology 138, 3836–3841
15. Begum, N. A., Mori, M., Matsumata, T., Takenaka, K., Sugimachi, K., and Barnard, G. F. (1995) Hepatology 22, 1447–1455
16. Capone, M. C., German, D. M., Ching, E. P., and Zlotnik, A. (1996) J. Immunol. 157, 969–973
17. Mao, M., Fu, G., Wu, J. S., Zhang, Q. H., Zhou, J., Kan, L. X., Huang, Q. H., He, K. L., Gu, B. W., Han, Z. G., Shen, Y., Gu, J., Yu, P. Y., Xu, S. H., Wang, Y. X., Chen, S. J., and Chen, Z. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 8175–8180
18. Kay, B. K., Williamson, M. P., and Sudol, M. (2000) FASEB J. 14, 231–241
19. Williamson, M. P. (1994) Biochem. J. 297, 249–260
20. Merkouropoulos, G., Barnett, D. C., and Shirsat, A. H. (1998) Planta 206, 212–219
21. Koken, M. H., Saib, A., and de The, H. (1995) C. R. Acad. Sci. 315, 733–739
22. Borden, K. L., and Freemon, P. S. (1996) Curr. Opin. Struct. Biol. 6, 395–401
23. Saurin, A. J., Borden, K. L., Reddy, M. N., and Freemont, P. S. (1996) Trends Biochem. Sci. 21, 208–214
24. Mackay, J. P., and Crossley, M. (1998) Trends Biochem. Sci. 23, 1–4
25. Stenmark, H., and Aasland, R. (1999) J. Cell Sci. 112, 4175–4183
26. Kim, M., Tenuka, T., Suzuki, Y., Sugano, S., Hirai, M., and Yamamoto, T. (1999) Gene (Amst.) 239, 145–154
27. Le, X. F., Yang, P., and Chang, K. S. (1996) J. Biol. Chem. 271, 130–135
28. Azano, H., Ishida, A., Hasegawa, M., Ota, T., Yoshida, M. C., Taniguchi, M., and Kanno, M. (1993) DNA Seq. 3, 369–377
29. Freemont, P. S. (2000) Curr. Biol. 10, R84–R87
30. Lai, Z., Freedman, D. A., Levine, A. J., and McLendon, G. L. (1998) Biochemistry 37, 17055–17051
31. Hu, G., and Fearon, E. R. (1999) Mol. Cell. Biol. 19, 724–732
32. Simmen, T., Schmidt, A., Hunziker, W., and Beermann, F. (1999) J. Cell Sci. 112, 45–53
33. Bonifacino, J. S., and Dell’Angelica, E. C. (1999) J. Cell Biol. 145, 923–926
34. Kutateladze, T. G., Ogburn, K. D., Weiss, D. S., and Godowski, P. (1998) Science 280, 1044–1047