Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 10094–10098, November 1991
Immunology

Two putative subunits of a peptide pump encoded in the human major histocompatibility complex class II region

SEIAMAK BAHRAM, DANIELE ARNOLD, MAUREEN BRESNAHAN, JACK L. STROMINGER, AND THOMAS SPIES

Division of Tumor Virology, Dana–Farber Cancer Institute, Harvard Medical School, 44 Binney Street, Boston, MA 02115

Contributed by Jack L. Strominger, August 7, 1991

ABSTRACT **The class II region of the human major histocompatibility complex (MHC) may encode several genes controlling the processing of endogenous antigen and the presentation of peptide epitopes by MHC class I molecules to cytotoxic T lymphocytes. A previously described peptide supply factor (PSF1) is a member of the multidrug-resistance family of transporters and may pump cytosolic peptides into the membrane-bound compartment where class I molecules assemble. A second transporter gene, *PSF2*, was identified 10 kilobases (kb) from *PSF1*, near the class II *DOB* gene. The complete sequences of PSF1 and PSF2 were determined from cDNA clones. The translation products are closely related in sequence and predicted secondary structure. Both contain a highly conserved ATP-binding fold and share 25% homology in a hydrophobic domain with a tentative number of eight membrane-spanning segments. Based on the principle dimeric organization of these two domains in other transporters, PSF1 and PSF2 may function as complementary subunits, independently as homodimers, or both. Taken together with previous genetic evidence, the coregulation of *PSF1* and *PSF2* by γ interferon and the to-some-degree coordinate transcription of these genes suggest a common role in peptide-loading of class I molecules, although a distinct function of *PSF2* cannot be ruled out.**

Cytotoxic T lymphocytes recognize antigenic peptides bound to major histocompatibility complex (MHC) class I molecules, which consist of polymorphic HLA-A, -B, or -C chains noncovalently associated with β_2-microglobulin (1, 2). Peptides are produced by cytosolic degradation of the endogenous protein pool that may include viral antigen (3, 4). By selective binding, they induce a stable conformation of nascent class I chains and pairing with β_2-microglobulin in the endoplasmic reticulum (ER) or in the pre-Golgi intermediate compartment (5–7). In the assembled complex, peptides of eight or nine amino acids are firmly integrated in a groove formed by two polymorphic α-helices above a β-pleated sheet of class I chains (8–10). Assembly and surface expression of class I molecules are impaired in lymphoblastoid cell line (LCL) mutants incapable of presenting endogenous viral antigen (11–14). This phenotype is caused by lack of suitable peptides in the ER (5, 13) and has been genetically linked in one mutant to a peptide supply factor gene (*PSF1*; formerly *PSF*)* in the MHC class II region (15). Structural and functional evidence suggest that PSF1 is a pump mediating entry of yet undefined precursor peptides into the ER (15, 16), consistent with the requirement of a signal sequence-independent mechanism of peptide transport across lipid membranes (17).

PSF1 is identical to RING4 (18) and is homologous to a family of polytopic integral membrane proteins functioning in ATP-dependent membrane translocation of a wide range of substrates that include peptides. Members of this family are the multidrug-resistance (MDR) P-glycoproteins (PGY) (19–21); the cystic fibrosis transmembrane conductance regulator (CFTR) (22); a peroxisomal membrane protein (PMP70) (23); the yeast *STE6* gene product, which exports the 12-amino acid a-type mating pheromone (24, 25); and the bacterial CyaB, HlyB, and LktB proteins, which mediate the secretion of large polypeptide toxins (26). More distantly related are a number of bacterial multicomponent dedicated import systems that are associated with periplasmic substrate-binding proteins, such as the Opp oligopeptide permease (27, 28). All of these transporters share a conserved cytoplasmic ATP-binding fold. This domain is thought to couple energy to a hydrophobic domain in which five to eight transmembrane segments form part of a transport channel. Although the hydrophobic domain shows little sequence homology among different transporters, the tertiary structures may be quite similar. In most eukaryotic transporters, these two domains are tandemly duplicated in a single transcriptional unit. A principle dimeric organization may be common to all transporters, except that the two ATP-binding domains and the two membrane-anchored domains are encoded by a variable number of one to four genes (29). PSF1 contains a single copy of each domain and thus may function as a homodimer or by pairing with a complementary subunit.

Direct evidence for peptide transport has not yet been obtained, but gene transfer of PSF1 restores normal surface expression of class I molecules in a mutant LCL (721.134) with a transcriptionally inactive single *PSF1* gene (16). Moreover, the demonstrated role of *PSF1* in peptide loading and assembly of class I molecules is corroborated by the coordinate transcriptional regulation with class I genes by interferons (18). However, in another mutant LCL (721.174), gene transfer of *PSF1* fails to reconstitute normal expression of class I molecules (16). Presumably, this is related to the homozygous deletion of *PSF1* together with several closely linked genes (15), some of which may be required for proteolytic processing or transport of peptides, or both.

A second transporter gene inside the LCL 721.174 deletion, *PSF2*,† was identified 10 kilobases (kb) from *PSF1*, near the 5′ end of the class II *DOB* gene. The ATP-binding domains are highly conserved in PSF1 and PSF2, and the transmembrane domains share an average 25% homology. Evidently, *PSF1* and *PSF2* may encode subunits associated in a functional peptide pump, as the duplicated transmembrane domains in MDR proteins and STE6 are similarly diverse. This is supported by the coregulation of *PSF1* and *PSF2* at least by IFN-γ and by the somewhat coordinate transcription of these genes in various cell lines, although it remains an open possibility that the two transporter proteins function independently.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: MHC, major histocompatibility complex; PSF, peptide supply factor; ER, endoplasmic reticulum; IFN, interferon; LCL, lymphoblastoid cell line; MDR, multidrug resistance; CFTR, cystic fibrosis transmembrane conductance regulator.

*The current designation of this gene is D6S114E (52).

†The sequence for *PSF2* reported in this paper has been deposited in the GenBank data base (accession no. M74447).

MATERIALS AND METHODS

DNA Sequence Analysis. Restriction fragments from cDNA clones were subcloned into phage M13, and sequences were obtained from both strands by using the dideoxynucleotide chain-termination procedure (31), adenosine 5′-[α-(^{35}S)thio]-triphosphate, and T7 polymerase (Sequenase; United States Biochemical). Computer sequence analysis and homology searches of GenBank used the programs EUGENE (Baylor College of Medicine, Houston) and BLAST (32) of the Molecular Biology Computer Research Resource (MBCRR) at Dana–Farber Cancer Institute, respectively. Alignment of amino acid sequences and hydropathy plotting were done as described by Feng and Doolittle (33) and Kyte and Doolittle (34), respectively.

RNA Blot Hybridization. Total-cell RNA preparations were obtained by the guanidinium isothiocyanate method and centrifugation in cushions of 5.7 M cesium chloride (31). RNA samples (20 μg) were fractionated by electrophoresis in agarose/formaldehyde gels before blot transfer (31). Hybridization was as cited (15). Recombinant IFN-γ (provided by Biogen) was used at 250 units per ml of HeLa cell culture in a 24-hr induction.

RESULTS AND DISCUSSION

PSF1 (also designated *Y3*) is one among five genes (*Y1–Y5*) identified in the MHC class II region between *DOB* and *DNA* by isolation of cDNA clones with cosmid probes (15). A partial PSF1 cDNA sequence corresponds to RING4 (human) (18), Ham-1 (mouse) (35), and MTP1 (rat) cDNAs (36). The full-length PSF1 sequence was derived from a cDNA clone (Y3-1) shown to direct synthesis of functional protein in a transfection assay (16). This sequence is 2667 base pairs (bp) long, includes 72 bp upstream of a translation initiation codon, and terminates 2 bp before the polyadenylylation site in RING4 mRNA. The 5′ end of the PSF1 cDNA thus lacks 127 bp of *RING4* genomic DNA sequence proposed to extend the open reading frame to an upstream ATG and to encode a putative leader peptide (18). However, the aligned PSF1 and RING4 decoded amino acid sequences are identical, suggesting that the 81-kilodalton (kDa) gene product may be invariant, or at least not highly polymorphic, as the PSF1 and RING4 cDNA clones have been isolated from different MHC-heterozygous cell lines.

Low-stringency blot hybridization of cosmid DNA restriction digests with a PSF1 cDNA probe revealed cross-hybridization with the previously mapped *Y1* gene (15), now referred to as *PSF2* (data not shown). These genes are separated by a 10-kb interval containing an unrelated gene, *Y2* (ref. 15; unpublished data) and are in the same transcriptional orientation as *DOB* 10–15 kb downstream of *PSF2* (Fig. 1). The location of *PSF2* corresponds to the mouse *Ham-2* and rat *MTP2* genes, which by partial cDNA sequencing are related to *Ham-1* and *MTP1*, respectively (35, 36). The PSF2 cDNA sequence was determined from two incomplete overlapping clones, Y1-7 and Y1-12. The total length of 2553 bp is close to the estimated 2.7-kb length of PSF2 mRNA (Fig. 2) (15). The proposed initiator methionine at nucleotide position 97 complies sufficiently with a consensus for eukaryotic translation initiation sites (38) and is preceded by an inframe termination codon. A single long open-reading frame encodes a 77-kDa translation product of 703 amino acids (Fig. 2). The canonical AATAAA polyadenylylation signal and a poly(A) tail are missing at the 3′ end of the Y1-12 cDNA. Reverse transcription and amplification of PSF2 mRNA by the polymerase chain reaction (PCR) with oligo(dT) together with a primer corresponding to nucleotides 1805–1824 yielded a DNA fragment of about 750 bp (data not shown); thus, PSF2 mRNA could have an uncommon polyadenylylation signal, such as the TATAAA motif seven nucleotides from the end of the truncated sequence, or the Y1-12 cDNA could be short of the actual polyadenylylation signal by a few nucleotides.

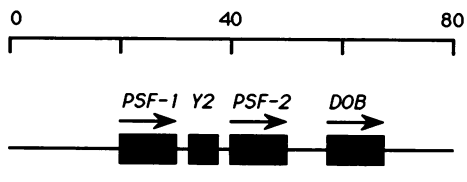

FIG. 1. Arrangement of the *PSF1* and *PSF2* transporter genes. Arrows above the genes show the transcriptional direction. The upper line gives the scale in kb. The diagram is adapted from ref. 15.

```
1     GGACCAGAGCCGGTAGCGAGGTTGGGAGAGACGGAGCGGACCTCAGCGCTGAAGCAGAAG
61    TCCCCGGAGCTGCGGTCTCCCCGCCGCGGCTGAGCCATGCGGCTCCCTGACCTGAGACCC
                                                  M  R  L  P  D  L  R  P   8
121   TGGACCTCCCTGCTGCTGGTGGACGCGGCTTTACTGTGGCTGCTTCAGGGCCCTCTGGGG
       W  T  S  L  L  L  V  D  A  A  L  L  W  L  L  Q  G  P  L  G      28
181   ACTTTGCTTCCTCAAGGGCTGCCAGGACTATGGCTGGAGGGGACCCTGCGGCTGGGAGGG
       T  L  L  P  Q  G  L  P  G  L  W  L  E  G  T  L  R  L  G  G      48
241   CTGTGGGGGCTGCTAAAGCTAAGAGGGCTGCTGGGATTTGTGGGGACACTGCTGCTCCCG
       L  W  G  L  L  K  L  R  G  L  L  G  F  V  G  T  L  L  L  P      68
301   CTCTGTCTGGCCACCCCCCTGACTGTCTCCCTGAGAGCCCTGGTCGCGGGGGCCTCACGT
       L  C  L  A  T  P  L  T  V  S  L  R  A  L  V  A  G  A  S  R      88
361   GCTCCCCCAGCCAGAGTCGCTTCAGCCCCTTGGAGCTGGCTGCTGGTGGGGTACGGGGCT
       A  P  P  A  R  V  A  S  A  P  W  S  N  I  L  V  G  Y  G  A     108
421   GCGGGGCTCAGCTGGTCACTGTGGGCTGTTCTGAGCCCTCCTGGAGCCCAGGAGAAGGAG
       A  G  L  S  W  S  L  W  A  V  L  S  P  P  G  A  Q  E  K  E     128
481   CAGGACCAGGTGAACAACAAAGTCTTGATGTGGAGGCTGCTGAAGCTCTCCAGGCCGGAC
       Q  D  Q  V  N  N  K  V  L  M  W  R  L  L  K  L  S  R  P  D     148
541   CTGCCTCTCCTCGTTGCCGCCTTCTTCTTCCTTGTCCTTGCTGTTTTGGGTGAGACATTA
       L  P  L  L  V  A  A  F  F  F  L  V  L  A  V  L  G  E  T  L     168
601   ATCCCTCACTATTCTGGTCGTGTGATTGACATCCTGGGAGGTGATTTTGACCCCCATGCC
       I  P  H  Y  S  G  R  V  I  D  I  L  G  G  D  F  D  P  H  A     188
661   TTTGCCAGTGCCATCTTCTTCATGTGCCTCTTCTCCTTTGGCAGCTCACTGTCTGCAGGC
       F  A  S  A  I  F  F  M  C  L  F  S  F  G  S  S  L  S  A  G     208
721   TGCCGAGGAGGCTGCTTCACCTACACCATGTCTCGAATCAACTTGCGGATCCGGGAGCAG
       C  R  G  G  C  F  T  Y  T  M  S  R  I  N  L  R  I  R  E  Q     228
781   CTTTTCTCCTCCCTGCTGCGCCAGGACCTCGGTTTCTTCCAGGAGACTAAGACAGGGGAG
       L  F  S  S  L  L  R  Q  D  L  G  F  F  Q  E  T  K  T  G  E     248
841   CTGAACTCACGGCTGAGCTCGGATACCACCCTGATGAGTAACTGGCTTCCTTTAAATGCC
       L  N  S  R  L  S  S  D  T  T  L  M  S  N  W  L  P  L  N  A     268
901   AATGTGCTCTTGCGAAGCCTGGTGAAAGTGGTGGGGCTGTATGGCTTCATGCTCAGCATA
       N  V  L  L  R  S  L  V  K  V  V  G  L  Y  G  F  M  L  S  I     288
961   TCGCCTCGACTCACCCTCCTTTCTCTGCTGCACATGCCCTTCACAATAGCAGCGGAGAAG
       S  P  R  L  T  L  L  S  L  L  H  M  P  F  T  I  A  A  E  K     308
1021  GTGTACAACACCCGCCATCAGGAAGTGCTTCGGGAGATCCAGGATGCAGTGGCCAGGGCG
       V  Y  N  T  R  H  Q  E  V  L  R  E  I  Q  D  A  V  A  R  A     328
1081  GGGCAGGTGGTGCGGGAAGCCGTTGGAGGGCTGCAGACCGTTCGCAGTTTTGGGGCCGAG
       G  Q  V  V  R  E  A  V  G  G  L  Q  T  V  R  S  F  G  A  E     348
1141  GAGCATGAAGTCTGTCGCTATAAAGAGGCCCTTGAACAATGTCGGCAGCTGTATTGGCGG
       E  H  E  V  C  R  Y  K  E  A  L  E  Q  C  R  Q  L  Y  W  R     368
1201  AGAGACCTGGAACGCGCCTTGTACCTGCTCATAAGGAGGGTGCTGCACTTGGGGGTGCAG
       R  D  L  E  R  A  L  Y  L  L  I  R  R  V  L  H  L  G  V  Q     388
1261  ATGCTGATGCTGAGCTGTGGGCTGCAGCAGATGCAGGATGGGGAGCTCACCCAGGGCAGC
       M  L  M  L  S  C  G  L  Q  Q  M  Q  D  G  E  L  T  Q  G  S     408
1321  CTGCTTTCCTTTATGATCTACCAGGAGAGCGTGGGGAGCTATGTGCAGACCCTGGTATAC
       L  L  S  F  M  I  Y  Q  E  S  V  G  S  Y  V  Q  T  L  V  Y     428
1381  ATATATGGGGATATGCTCAGCAACGTGGGAGCTGCAGAGAAGGTTTTCTCCTACATGGAC
       I  Y  G  D  M  L  S  N  V  G  A  A  E  K  V  F  S  Y  M  D     448
1441  CGACAGCCAAATCTGCCTTCACCTGGCACGCTTGCCCCCACCACTCTGCAGGGGGTTGTG
       R  Q  P  N  L  P  S  P  G  T  L  A  P  T  T  L  Q  G  V  V     468
1501  AAATTCCAAGACGTCTCCTTTGCATATCCCAATCGCCCTGACAGGCCTGTGCTCAAGGGG
       K  F  Q  D  V  S  F  A  Y  P  N  R  P  D  R  P  V  L  K  G     488
1561  CTGACGTTTACCCTACGTCCTGGTGAGGTGACGGCGCTGGTGGGACCCAATGGGTCTGGG
       L  T  F  T  L  R  P  G  E  V  T  A  L  V  G  P  N  G  S  G     508
1621  AAGAGCACAGTGGCTGCCCTGCTGCAGAATCTGTACCAGCCCACAGGGGGACAGGTGCTG
       K  S  T  V  A  A  L  L  Q  N  L  Y  Q  P  T  G  G  Q  V  L     528
1681  CTGGATGAAAAGCCCATCTCACAGTATGAACACTGCTACCTGCACAGCCAGGTGGTTTCA
       L  D  E  K  P  I  S  Q  Y  E  H  C  Y  L  H  S  Q  V  V  S     548
1741  GTTGGGCAGGAGCCTGTGCTGTTCTCCGGTTCTGTGAGGAACAACATTGCTTATGGGCTG
       V  G  Q  E  P  V  L  F  S  G  S  V  R  N  N  I  A  Y  G  L     568
1801  CAGAGCTGCGAAGATGATAAGGTGATGGCGGCTGCCCAGGCTGCCCACGCAGATGACTTC
       Q  S  C  E  D  D  K  V  M  A  A  A  Q  A  A  H  A  D  D  F     588
1861  ATCCAGGAAATGGAGCATGGAATATACACAGATGTAGGGGAGAAGGGAAGCCAGCTGGCT
       I  Q  E  M  E  H  G  I  Y  T  D  V  G  E  K  G  S  Q  L  A     608
1921  GCGGGACAGAAACAACGTCTGGCCATTGCCCGGGCCCTTGTACGAGACCCGCGGGTCCTC
       A  G  Q  K  Q  R  L  A  I  A  R  A  L  V  R  D  P  R  V  L     628
1981  ATCCTGGATGAGGCTACTAGTGCCCTAGATGTGCAGTGCGAGCAGGCCCTGCAGGACTGG
       I  L  D  E  A  T  S  A  L  D  V  Q  C  E  Q  A  L  Q  D  W     648
2041  AATTCCCGTGGGGATCGCACAGTGCTGGTGATTGCTCACAGGCTGCAGGCAGTTCAGCGC
       N  S  R  G  D  R  T  V  L  V  I  A  H  R  L  Q  A  V  Q  R     668
2101  GCCCACCAGATCCTGGTGCTCCAGGAGGGCAAGCTGCAGAAGCTTGCCCAGCTCCAGGAG
       A  H  Q  I  L  V  L  Q  E  G  K  L  Q  K  L  A  Q  L  Q  E     688
2161  GGACAGGACCTCTATTCCCGCCTGGTTCAGCAGCGGCTGATGGACTGAGGCCCCAGGGAT
       G  Q  D  L  Y  S  R  L  V  Q  Q  R  L  M  D  *                 703
2221  ACTGGGCCCTCTTCTCAGGGGCGTCTCCAGGACCCAGAGCTGTTCCTGCTTTGAGTTTCC
2281  CTAGAGCTGTGCGGCCAGATAGCTGTTCCTGAGTTGCAAGCACGATGGAGATTTGGACAC
2341  TGTGTGCTTTTGGTGGGGTGGAGAGGTGGGGTGGGGTGGGGTGTGGTGGGGTGGGAGGCT
2401  GTCTGTGTCCAGGAAACTTAATTCCCTGGTGACTAGAGCTTTGCCCTGGTGATGAGGAGTA
2461  TTTTGTGGCATAATACATATATTTTAAAATATTTTCCTTCTTACGTGAACTGTATACATT
2521  CATATAGAAAATTTAGACAATATAAAAAAGTAC
```

FIG. 2. Nucleotide sequence of PSF2 derived from cDNA clones. Decoded amino acids in single-letter code are shown below codons. Numbers on the left and on the right correspond to nucleotide and amino acid positions, respectively. Transmembrane segments of 21 amino acids predicted by the Eisenberg algorithm (37) are underlined. Boxed residues 483–683 correspond to the hydrophilic ATP-binding fold containing the underlined Walker A and B consensus sequences.

The PSF2 translation product is similar to PSF1 in sequence, size, and domain structure and is equally related to the MDR family of transporters (Fig. 3). Both lack a potential leader peptide, in accord with all known eukaryotic transporters, which may use an internal hydrophobic segment for cotranslational ER membrane insertion. In PSF2, the N-terminal hydrophobic domain (482 amino acids) is followed by the highly conserved hydrophilic ATP-binding fold (200 amino acids) (Fig. 2) containing the universal motif A (also called Walker A) [Gly-$(Xaa)_4$-Gly-Lys-(Ser or Thr)] derived by Walker in ATP-binding proteins (27, 39). The more degenerate motif B (Walker B) [(Arg or Lys)-$(Xaa)_3$-Gly-$(Xaa)_3$-Leu-(hydrophobic)$_4$-Asp] is not readily discernible and was placed with reference to previous sequence alignments (Fig. 3) (29, 35). The ATP-binding domain shows the highest degree of homology shared by PSF2 and PSF1, with 56 identical positions in a segment of 60 amino acids including the Walker A motif and with 25 matches in a sequence of 35 amino acids containing the Walker B motif (Fig. 3). These two regions correspond to the sequences most conserved among all eukaryotic and bacterial transporters (for a recent compilation of sequences, see ref. 35). The function of the ATP-binding fold in hydrolysis of ATP is supported by direct biochemical evidence and mutational analysis in several systems (29, 40, 41). Deleterious mutations in CFTR are clustered in the first ATP-binding fold (42). Moreover, a tertiary structure model of a consensus transporter ATP-binding fold is similar to the structure of adenylate kinase, determined by crystallography, except for the presence of two loops between Walker motifs A and B (29). The corresponding conserved sequences in PSF1 and PSF2 are amino acid residues 541–567 and 603–617 (PSF1 coordinates in Fig. 3). It has been suggested that these two loops might be specifically involved in the transport mechanism by coupling ATP-dependent conformation changes to the membrane-anchored hydrophobic domain(s) (29).

Within the N-terminal hydrophobic domain, PSF1 and PSF2 have 25% amino acid residues in common, indicating a much greater sequence diversity than in the ATP-binding fold (Fig. 3). Amino acid substitutions are randomly distributed. Except for several gaps at the beginning and a single-residue gap at position 160 of the PSF2 sequence (PSF1 coordinate), all of the PSF1 and PSF2 amino acids are aligned, suggesting a high degree of secondary structure conservation. Moreover, the PSF1 and PSF2 hydropathy profiles resemble each other more closely than those of other transporter proteins (Fig. 4). It is evident, therefore, that *PSF1* and *PSF2* evolved by gene duplication and subsequent diversification. Accordingly, the total amino acid sequence homology between PSF1 and PSF2 (33%) is lower than that between PSF1 and MTP1 (70%) and Ham-1 (57%) (Fig. 3). By comparison, the human MDR proteins PGY1 and PGY3, which function independently with as yet undefined natural substrate specificities, share 80% identical amino acids in their corresponding hydrophobic domains (44). However, the two hydrophobic subunits in PGY1 are only 23% homologous (19), which parallels the sequence diversity between PSF1 and PSF2 in these regions. When compared to other transporters, these share significant homology only with the N- and C-terminal hydrophobic domains of PGY1 (19–23%) (19) and STE6 (18–20%), the two last-named proteins being equally related (20–22%) (24).

Although the general topological organization of PSF1 and PSF2 is characteristic of all eukaryotic and prokaryotic transporters, the hydropathy profiles can be best compared to the PGY and bacterial HlyB proteins (Fig. 4). Each half of PGY contains six putative transmembrane segments (19, 20) that also may be present in PSF1 and PSF2, which relative to PGY show N-terminal hydrophobic extensions of about 105 and 60 amino acids, respectively (Fig. 4). Thus, PSF1 and PSF2 are more alike than HlyB, in which a total of eight transmembrane segments has been experimentally demonstrated, differing from an earlier calculated number of six (43). Based on the algorithm by Eisenberg (37), PSF2 contains six α-helical membrane-spanning regions at the amino acid positions 11–31, 58–78, 100–120, 150–170, 186–206, and 278–298 (Figs. 1 and 4). However, as the PGY transmembrane segments V and VI are not matched, this prediction may be inaccurate. A tentative number of eight transmembrane regions in both PSF1 and PSF2 is therefore proposed with regard to their close resemblance to other proteins in this family (Fig. 4). Functional analysis of chimeric *PGY* gene constructs supports an involvement of the transmembrane domains in drug efflux (45). Moreover, a single glycine-for-valine substitution at position 185 in human PGY1 confers preferential resistance to colchicine, and several allelic var-

```
PSF1  MASSRCPAPRGCRCLPGASLAWLGTVLLLLADWVLLRTALPRIFSLLVPTALPLLRVWAVGLSRWAVLWLGACGVLRATVGSKSENAGAQGWLAALKPLAAALGLALPGLALFRELISWG  120
mtp1  --AHAW-TA.....AL----V--L---PV--G-------E.V--------------I-G--VR---GV-A-ARG.....-----Q--V----------S--K-SA--  97
HAM1
PSF2  -RLPDLRPW.....S---VDAAL-W.....LLQ.G--GTLLPQ--PGLWLEGTLRL-G-WGLLKLRG.....L-GFVGT-LLP-C--T-LTVSL-A-VAGA  86

PSF1  APGSADSTRLLHWGSHPTAFVVSYAAALPAAALWHKLGSLWVPGGQGGSGNPVRRLLGCLGSETRRLSLFLVLVVLSSLGEMAIPFFTGRLTDWILQDGSADTFTRNLTLMSILTIASAV  240
mtp1  -LREG-NAG----N-RLD---L--V-----------GF-A-S-HK-A-DMLC-M--F-D-KKG--H-V---LI--C----------I-------KT-PS-A--MW--C------T-  217
HAM1                                      +++-+--+--P++---Y-+---++--+----------+-------++V++----I+---------+A  69
PSF2  SRAPPARVASAP-SWLLVGYGAAGLSWSLW-V-SPPGAQ.EKEQD-VNNKVLMW---KLSRPDLPL-VAAFFFL--AV---TL--HYS--VI-ILGG-FDPHA-ASAIFF-CLFSFG-SL  205

PSF1  LEFVGDGIYNNTMGHVHSHLQGEVFGAVLRQETEFFQQNQTGNIMSRVTEDTSTLSDSLSENLSLFLWYLVRGLCLLGIMLWGSVSLTMVTLITLPLLFLLPKKVGKWYQLLEVQVRESL  360
mtp1  ---A------I----M--RVH----R---H---G--LK-P--S-T---------NVCE-I-DK-N------G------AF-I---FY--V---LS--------RRL--V--S-A-K-Q---  337
HAM1  ---+S-----+----+-G+++R---+-------+--++-+A-+-+-------A++++-+-+T---L----+-A----V+-F---P+--L----N----------+--+H-+-+-+---  189
PSF2  SAGCRG-CFTY--SRINLRIREQL-SSL---DLG---ETK--ELN--LSS--TLM-NW-PL-ANVL-RS--KVVG-Y-F--SI-PR--LLS-LHM-FTIAAE-VYNTRH-EVLREIQDAV  325

PSF1  AKSSQVAIEALSAMPTVRSFANEEGEAQKFREKLQEIKTLNQKEAVAYAVNSWTTSISGMLLKVGILYIGGQLVTSGAVSSGNLVTFVLYQMQFTQAVEVLLSIYPRVQKAVGSSEKIFE  480
mtp1  ---T---L----------------------Q--E-M-P--K---L--VTEV--M-V----------L-----VR---------S-----L---R----------SM--S--A------  457
HAM1  ---+---+----------------------+--+-+----+---+--+A++----+----------+-----I+-T-------+-----+------Q----L--++----------  309
PSF2  -RAG--VR--VGGLQ-----GA--H-VCRYK-A-EQCRQ-YWRRDLER-LYLLIRRVLHLGVQMLM-SC-L-QMQD-ELTQ-S-LS-MI--ESVGSY-QT-VY--GDMLSN--AA--V-S  445
                                                                  — WA —
PSF1  YLDRTPRCPPSGLLTPLHLEGLVQFQDVSFAYPNRPDVLVLQGLTFTLRPGEVTALVGPNGSGKSTVAALLQNLYQPTGGQLLLDGKPLPQYEHRYLHRQVAAVGQEPQVFGRSLQENIA  600
mtp1  ------CS-L--S-A--NMK---K----------H-N-Q---------Y--K----------------------------KV----E--V--D-H---T----------LL----FR----  577
HAM1  ------++-+--+-+-+-S+++---E----------Q-K-+---------H--T----------------------------QR-+---+-+---+---------++----+----  429
PSF2  -M--Q-NL-SP-T-A-TT-Q-V-K-------------RP--K-------------------------------------------V---E--IS----C---S--VS-----VL-SG-VRN---  565
                                                                  —— WB ——
PSF1  YGLTQKPTMEEITAAAVKSGAHSFISGLPQGYDTEVDEAGSQLSGGQRQAVALARALIRKPCVLILDDATSALDANSQLQVEQLLYESPERYSRSVLLITQHLSLVEQADHILFLEGGAI  720
mtp1  ----RT--------V-ME----D----F--------G-T-N-------------------RL-----------GN--R-QR-------WA--T------Q---A-R-H-----KE-SV  697
HAM1  ---N++---------+--+----+----+---------+-+-+-------------------L+----------++--+-++------K-+--+------+---+---+-----R+-++  549
PSF2  ---QSCED.DKVM---QAAH-DD--QEMEH-IY-D-G-K----AA--K-RL-I----V-D-R-----E------....V-C--A-QDWNS-GD-T--V-AHR-QA-QR-HQ--V-QE-KL  680

PSF1  REGGTHQQLMEKKGCYWAMVQAPADAPE*  748
mtp1  C-Q---L----RG---RS--E-L-APSD*  725
HAM1  G-+---+---K++---+---+-+-++A+*  577
PSF2  QKLAQL-EGQDLYSRLVQQRLM.....D*  703
```

FIG. 3. Comparison of PSF2 to PSF1 and the corresponding rat MTP1 and mouse Ham-1 amino acid sequences. Sequences as available were aligned with PSF1 by using a program by Feng and Doolittle (33). Dashes and dots represent identical amino acids and gaps, respectively. Residues shared by MTP1 and Ham-1 are indicated by a plus. Numbers at the right refer to sequence coordinates. See the text for references and further explanation. WA and WB, Walker A and B motifs.

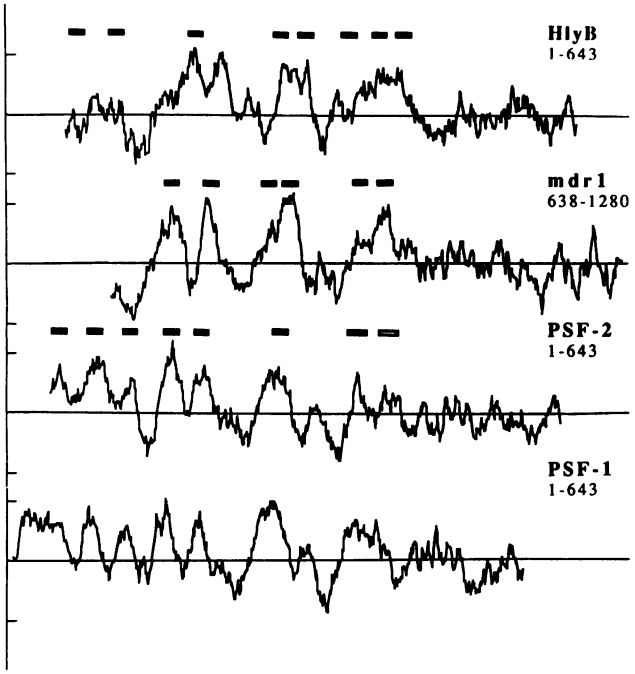

FIG. 4. Comparison of hydropathy profiles of PSF2 and PSF1 to HlyB and the C-terminal half of human MDR1. Plots were established by using a program by Kyte and Doolittle (34) with a window of 21 amino acids and were visually aligned. Of both PSF1 and PSF2, residues 1–643 are shown. Solid bars indicate transmembrane segments, which in HlyB have been experimentally demonstrated (43) and which in MDR1 (19) and PSF2 are predicted by the Eisenberg algorithm (37). Open bars above the PSF2 profile are proposed as additional membrane-spanning segments based on the structural similarities shared with HlyB and MDR1.

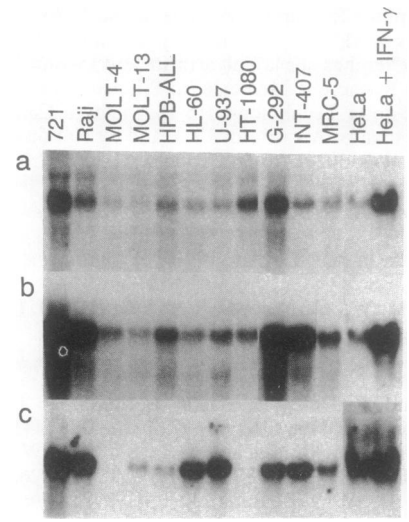

FIG. 5. Expression of PSF1, PSF2, and HLA-B mRNA in various cell lines and up-regulation by IFN-γ in HeLa cells. Blots of total RNA samples were hybridized with cDNA probes for PSF2 (*a*), PSF1 (*b*), and a locus-specific genomic probe for HLA-B (*c*). Filters were exposed for 4 days in *a* and *b* and for 18 hr in *c*. The human cell lines included B cells (721, Raji), T cells (MOLT-4, MOLT-13, HPB-ALL), promyelocytes (HL-60), monocytes (U-937), and epithelial or fibroblastic cells (HT-1080, G-292, INT-407, MCR-5, and HeLa). For further details, see text.

iants of *pfmdr1* in *Plasmodium falciparum* are specifically associated with chloroquine resistance (46, 47). A role of the putative transmembrane domains in PSF1 and PSF2 in substrate selection and transport seems certain, although as in other transporters, the underlying mechanisms are completely unknown.

Genetic evidence suggests that PSF1 and PSF2 may not function individually but instead by association in a heterodimer (16). This would be equivalent to the composite structure of PGY, pfmdr, CFTR, and STE6, in which two homologous subunits are integral. Both hydrophobic domains in PGY have been implicated in substrate binding and may effect drug efflux in a cooperative manner (48). The necessary functioning of both ATP-binding folds is likely to be concerted, perhaps by alternate hydrolysis of ATP coupled to sequential conformation changes (41). This organization may be universal in all transporter systems, except that the four domains reside on different numbers of protein subunits. For example, the mammalian peroximal membrane protein PMP70, the *Drosophila* white and brown gene products, and the bacterial HlyB, CyaB, and LktB proteins carry single transmembrane and ATP-binding domains, like PSF1 and PSF2, and may form homodimers; in the bacterial ribose permease, two ATP-binding folds are fused in RbsA and in the Opp oligopeptide permease, all four domains are on distinct polypeptides (23, 26, 28, 29). In the absence of further genetic and/or biochemical data, the possibility that PSF1 and PSF2 are complementary subunits is supported by their similar transcriptional regulation. In a panel of 12 cell lines representing several different lineages, the relative PSF1 and PSF2 steady-state mRNA levels corresponded to some degree to each other and in most samples also to those of HLA-B (Fig. 5). Moreover, both the PSF1 and PSF2 mRNA levels were strongly increased by IFN-γ (Fig. 5 *a* and *b*). However, at present it cannot be ruled out that PSF1 and PSF2 function independently, either in different cellular compartments or with distinct substrate specificities, or both.

The probable location at least of PSF1 is in the ER membrane, with the hydrophilic ATP-binding fold facing the cytosol. Potential N-glycosylation sites likely to be exposed to the ER lumen are present in PSF1 at the positions 250, 279, and 302 but are missing in PSF2. In some transmembrane proteins, such as the adenovirus E3/19-kDa protein and UDP-glucuronyl transferase, two consecutive lysine residues three to five positions from the C terminus confer retention in the ER (49). This motif is absent in both PSF1 and PSF2, in which the C-terminal 25 amino acids are entirely unrelated.

In summary, the available functional and genetic data and the close evolutionary relationship between PSF1 and PSF2 suggest that both of these transport peptides, together as heterodimers or independently as homodimers, or both. Formally, PSF2 could be related to a defect in several LCL mutants that are unable to process exogenous antigen and express conformationally labile class II heterodimers (50), but there is no current conceptual support for this role. It is plausible that PSF2 may function complementary to PSF1 in peptide-loading of class I molecules. Presumably, the substrate peptides are of at least nine amino acids to bind effectively to class I chains and might have a much greater upper size limit to avoid constraining class I allele-specific peptide selection (10). Peptide-loading could involve a direct contact of class I chains with PSF1 and/or PSF2, as free intracellular peptides appear to be extremely short-lived (ref. 10 and references therein). Finally, some variability may exist in peptide transport, as suggested by the characteristics of the as-yet-unidentified *CIM* locus in the rat MHC, which determines the alloantigenicity of the RT1.A^a class I molecule and maps to a class II region interval containing the *MTP1* and *MTP2* transporter genes (36, 51). Thus, some allelic variants of *PSF1* and/or *PSF2* could restrict antigen presentation and be related to several autoimmune diseases that are associated with genes in the MHC class II region (30).

We thank the National Institutes of Health-supported Molecular

Biology Computer Research Resource at the Dana–Farber Cancer Institute. This work was supported by fellowships from the Association de Recherches sur la Polyarthrite and Institut National de la Sante et de la Recherche Médicale (to S.B.) and the Ministere de la Recherche et de la Technologie (to D.A.), by a Cancer Research Institute Investigator Award (to T.S.), and by National Institutes of Health Grants DK-30241 (to J.L.S.) and AI-30581 (to T.S.).

1. Townsend, A. & Bodmer, H. (1989) *Annu. Rev. Immunol.* **7**, 601–624.
2. Bjorkman, P. J. & Parham, P. (1990) *Annu. Rev. Biochem.* **59**, 253–288.
3. Moore, M. W., Carbone, F. R. & Bevan, M. J. (1988) *Cell* **54**, 777–785.
4. Sweetser, M. T., Morrison, L. A., Braciale, V. L. & Braciale, T. J. (1989) *Nature (London)* **342**, 180–182.
5. Townsend, A., Ohlen, C., Bastin, J., Ljunggren, H.-G., Foster, L. & Karre, K. (1989) *Nature (London)* **340**, 443–448.
6. Nuchtern, J. G., Bonifacino, J. S., Biddison, W. E. & Klausner, R. D. (1989) *Nature (London)* **339**, 223–226.
7. Yewdell, J. W. & Bennink, J. R. (1989) *Science* **244**, 1072–1075.
8. Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L. & Wiley, D. C. (1987) *Nature (London)* **329**, 512–518.
9. Van Bleek, G. M. & Nathenson, S. G. (1990) *Nature (London)* **348**, 213–216.
10. Falk, K., Rotzschke, O., Stevanovic, T., Jung, G. & Rammensee, H.-G. (1991) *Nature (London)* **351**, 290–296.
11. DeMars, R., Rudersdorf, R., Chang, C., Petersen, J., Strandtmann, J., Korn, N., Sidwell, B. & Orr, H. T. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 8183–8187.
12. Salter, R. D. & Cresswell, P. (1986) *EMBO J.* **5**, 943–949.
13. Cerundolo, V., Alexander, J., Anderson, K., Lamb, C., Cresswell, P., McMichael, A., Gotch, F. & Townsend, A. (1990) *Nature (London)* **345**, 449–452.
14. Hosken, N. A. & Bevan, M. J. (1990) *Science* **248**, 367–370.
15. Spies, T., Bresnahan, M., Bahram, S., Arnold, D., Blanck, G., Mellins, E., Pious, D. & DeMars, R. (1990) *Nature (London)* **348**, 744–747.
16. Spies, T. & DeMars, R. (1991) *Nature (London)* **351**, 323–324.
17. Anderson, K., Cresswell, P., Gammon, M., Hermes, J., Williamson, A. & Zweerink, H. (1991) *J. Exp. Med.* **174**, 489–492.
18. Trowsdale, J., Hanson, I., Mockridge, I., Beck, S., Townsend, A. & Kelly, A. (1990) *Nature (London)* **348**, 741–744.
19. Chen, C.-j., Chin, J. E., Ueda, K., Clark, D. P., Pastan, I., Gottesman, M. M. & Roninson, I. B. (1986) *Cell* **47**, 381–389.
20. Gerlach, J. H., Endicott, J. A., Juranka, P. F., Henderson, G., Sarangi, F., Deuchars, K. L. & Ling, V. (1986) *Nature (London)* **324**, 485–489.
21. Ueda, K., Cardarelli, C., Gottesman, M. M. & Pastan, I. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 3004–3008.
22. Riordan, J. R., Rommens, J. M., Kerem, B.-S., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J.-L., Drumm, M. L., Iannuzzi, M. C., Collins, F. S. & Tsui, L.-C. (1989) *Science* **245**, 1066–1073.
23. Kamijo, K., Taketani, S., Yokota, S., Osumi, T. & Hashimoto, T. (1990) *J. Biol. Chem.* **265**, 4534–4540.
24. Kuchler, K., Sterne, R. E. & Thorner, J. (1989) *EMBO J.* **8**, 3973–3984.
25. McGrath, J. P. & Varshavsky, A. (1989) *Nature (London)* **340**, 400–404.
26. Blight, M. A. & Holland, I. B. (1990) *Mol. Microbiol.* **4**, 873–880.
27. Higgins, C. F., Hiles, I. D., Salmond, G. P. C., Gill, D. R., Downie, J. A., Evans, I. J., Holland, I. B., Gray, L., Buckel, S. D., Bell, A. W. & Hermodson, M. A. (1986) *Nature (London)* **323**, 448–450.
28. Hiles, I. D., Gallagher, M. P., Jamieson, D. J. & Higgins, C. F. (1987) *J. Mol. Biol.* **195**, 125–142.
29. Hyde, S. C., Emsley, P., Hartshorn, M. J., Mimmack, M. M., Gileadi, U., Pearce, S. R., Gallagher, M. P., Gill, D. R., Hubbard, R. E. & Higgins, C. F. (1990) *Nature (London)* **346**, 362–365.
30. Nepom, G. T. & Erlich, H. (1991) *Annu. Rev. Immunol.* **9**, 493–525.
31. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
32. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410.
33. Feng, D. F. & Doolittle, R. F. (1987) *J. Mol. Evol.* **25**, 351–360.
34. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
35. Monaco, J. J., Cho, S. & Attaya, M. (1990) *Science* **250**, 1723–1726.
36. Deverson, E. V., Gow, I. R., Coadwell, W. J., Monaco, J. J., Butcher, G. W. & Howard, J. C. (1990) *Nature (London)* **348**, 738–741.
37. Eisenberg, D. (1984) *Annu. Rev. Biochem.* **53**, 595–623.
38. Kozak, M. (1987) *Nucleic Acids Res.* **15**, 8125–8132.
39. Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. (1982) *EMBO J.* **1**, 945–951.
40. Horio, M., Gottesman, M. M. & Pastan, I. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 3580–3584.
41. Azzaria, M., Schurr, E. & Gros, P. (1989) *Mol. Cell. Biol.* **9**, 5289–5297.
42. Cutting, G. R., Kasch, L. M., Rosenstein, B. J., Zielenski, J., Tsui, L.-C., Antonarakis, S. E. & Kazazian, H. H. (1990) *Nature (London)* **346**, 366–369.
43. Wang, R., Seror, S. J., Blight, M., Pratt, J. M., Broome-Smith, J. K. & Holland, I. B. (1991) *J. Mol. Biol.* **217**, 441–454.
44. van der Bliek, A. M., Kooiman, P. M., Schneider, C. & Borst, P. (1988) *Gene* **71**, 401–411.
45. Buschman, E. & Gros, P. (1991) *Mol. Cell. Biol.* **11**, 595–603.
46. Choi, K., Chen, C.-j., Kriegler, M. & Roninson, I. B. (1988) *Cell* **53**, 519–529.
47. Foote, S. J., Kyle, D. E., Martin, R. K., Oduola, A. M. J., Forsyth, K., Kemp, D. J. & Cowman, A. F. (1990) *Nature (London)* **345**, 255–258.
48. Bruggemann, E. P., Germann, U. A., Gottesman, M. M. & Pastan, I. (1989) *J. Biol. Chem.* **264**, 15483–15488.
49. Shin, J., Dunbrack, R. L., Lee, S. & Strominger, J. L. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1918–1922.
50. Mellins, E., Smith, L., Arp, B., Cotner, T., Celis, E. & Pious, D. (1990) *Nature (London)* **343**, 71–74.
51. Livingstone, A. M., Powis, S. J., Diamond, A. G., Butcher, G. W. & Howard, J. C. (1989) *J. Exp. Med.* **170**, 777–795.
52. Klinger, H. P. (1990) *Human Gene Mapping 10.5: Update to the Tenth International Workshop on Human Gene Mapping* (Karger, Basel).