Human Proton/Oligopeptide Transporter (POT) Genes: Identification of Putative Human Genes Using Bioinformatics

Submitted February 25, 2000; accepted May 11, 2000; published June 7, 2000

Christopher W. Botka, Thomas W. Wittig, Richard C. Graul, Carsten Uhd Nielsen, Wolfgang Sadée
Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, University of California San Francisco, San Francisco, California, USA

Kazutaka Higaki, Gordon L. Amidon
Department of Pharmaceutics, College of Pharmacy, The University of Michigan, Ann Arbor, Michigan, USA

ABSTRACT **Purpose**: The proton-dependent oligopeptide transporters (POT) gene family currently consists of ~70 cloned cDNAs derived from diverse organisms. In mammals, two genes encoding peptide transporters, *PepT1* and *PepT2* have been cloned in several species including humans, in addition to a rat histidine/peptide transporter (*rPHT1*). Because the *Candida elegans* genome contains five putative POT genes, we searched the available protein and nucleic acid databases for additional mammalian/human POT genes, using iterative BLAST runs and the human expressed sequence tags (EST) database. The apparent human orthologue of *rPHT1* (expression largely confined to rat brain and retina) was represented by numerous ESTs originating from many tissues. Assembly of these ESTs resulted in a contiguous sequence covering ~95% of the suspected coding region. The contig sequences and analyses revealed the presence of several possible splice variants of *hPHT1*. A second closely related human EST-contig displayed high identity to a recently cloned mouse cDNA encoding cyclic adenosine monophosphate (cAMP)-inducible 1 protein (gi:4580995). This contig served to identify a PAC clone containing deduced exons and introns of the likely human orthologue (termed *hPHT2*). Northern analyses with EST clones indicated that *hPHT1* is primarily expressed in skeletal muscle and spleen, whereas *hPHT2* is found in spleen, placenta, lung, leukocytes, and heart. These results suggest considerable complexity of the human POT gene family, with relevance to the absorption and distribution of cephalosporins and other peptoid drugs.

Corresponding author: Wolfgang Sadee, Department of Biopharmaceutical Sciences and Pharmaceutical Chemistry, University of California San Francisco, San Francisco, California, USA. 94143-0446. sadee@cgl.ucsf.edu

INTRODUCTION

Small peptides and peptidelike drugs are often too polar to cross lipid bilayers by simple diffusion. Therefore, translocation across membranes depends on transport by a suitable carrier, and orally administered peptoid drugs would be poorly absorbed unless transported (1). The main intestinal H^+/dipeptide transporter protein, PepT1, is thought to play a critical role in oral bioavailability of peptidelike drugs (2-7). *hPepT1* is a member of a well defined small gene family, the proton-dependent oligopeptide transporters (POT, also referred to as PTR), with ancestral roots that can be traced to bacterial, fungal, and plant peptide transporters (8-11). This class of secondary active transporters has broad selectivity for di- and tripeptides, whereas ability to transport longer peptides decreases drastically with increasing length. Most of the POT members share a common structural architecture with ~12 predicted transmembrane domains (TMDs), but among the dipeptide transporters in distant phyla, variations on this theme do occur. We have confirmed the transmembrane topology of at least a portion of the main intestinal transporter, hPepT1, using an epitope tagging approach (12).

Until recently, only two POT genes had been identified in mammalian species, *PepT1* and *PepT2*, the main renal peptide transporters (6,13). cDNA's of

the respective human orthologues have been cloned (orthologue refers to the same gene in different species) (13,14). These transporters display overlapping, broad substrate selectivity and interact with numerous drugs, some with chemical structures quite distinct from peptides. Substrates include important drug classes, such as β-lactam and cephalosporin antibiotics, renin inhibitors, ACE inhibitors, and 5'-nucleoside esters of amino acids, such as valcyclovir (15).

With a few exceptions, single amino acids are not substrates. In 1997, a third cDNA was cloned, encoding the rat peptide-histidine transporter, rPHT1 (16), with use of sequence alignments of POT members with the expressed sequence tags (EST) database. Currently cloned mammalian POT cDNAs and deduced proteins are summarized in Table 1.

Whereas rPHT1 is mainly expressed in rat brain and retina, the identity and tissue distribution of its human orthologue remain unknown. We have found numerous human ESTs that on the basis of high sequence identity among them appear to represent the human orthologue of the rat transporter gene *rPHT1*. Many of these ESTs were isolated from tissues other than brain, eg, human colon carcinoma, suggesting that the tissue distribution of this gene product in humans might differ from that in rat. This finding cautions against assuming that only one H^+/dipeptide transporter is expressed in a tissue of interest. Nakanishi et al (1997) have postulated the presence of a distinct peptide carrier in a human fibrosarcoma cell line (17). To understand peptide transport, all relevant genes need to be identified and characterized.

Table 1. Molecular characteristics of cloned POT cDNAs*

Transporter	**hPepT1**	**hPepT1-RF†**	**hPepT2**	**rPHT1**
Family	POT	POT	POT	POT
Accession No.	U13173	AB001328	PS01023 g2833279	02208839(AB000280)
Chromosome location	13q.34-q.35		3q13.3-q21	
Protein length	708 a.a.	208 a.a.	729 a.a.	
TMD	12		12	12
Substrate	oligopeptides peptoid drugs	-	oligopeptides peptoid drugs	oligopeptides histidine
Tissue distribution	intestines	intestines	kidney	brain (rat)
Function	H^+-dependent transport	pH-sensing regulatory factor	H^+-dependent transport	H^+-dependent transport

**POT indicates proton-dependent oligopeptide transporters; TMD indicates transmembrane domain.*

†Possible alternative splicing product of hPepT1

Comprehensive sequence analysis reveals the presence of at least five possible members of the POT family in *C elegans* (this report). Therefore, one would expect several POT genes to exist in the human genome, in addition to *hPepT1* and *hPepT2*. Because the POT family shares limited sequence similarity with other transporter families, any newly identified sequences with significant similarity to POT proteins are likely to be members of the POT family. Therefore, our overall objective was to identify all human homologues of the POT family and determine their tissue distribution. In this study, we show results obtained from a bioinformatics analysis of the available databases, supplemented by assays of gene expression in human tissues.

To find all members of a gene family, we have developed a Java-based program, which iteratively searches sequence databases for homologous genes using BLAST. Called INCA (iterative neighborhood cluster analysis; http://itsa.ucsf.edu/~gram/home/inca/), this program identifies a complete cluster of sequence neighbors in the database (18). By applying INCA to search the

nonredundant protein sequence database and subsequently the human EST database for novel dipeptide transporters, we have identified two new human genes as possible members of POT and several ESTs representing candidate genes of additional human peptide transporters of the POT family. A greater repertoire of the dipeptide transporter gene family in humans must be considered in the interpretation of pharmacological studies with peptoid drugs and could also serve for targeting drugs to specific tissues. Moreover, sequence variations in these transporters could account for interindividual genetic differences in the disposition of peptoid drugs.

METHODS

Iterative Neighborhood Cluster Analysis (INCA) using BLAST

Scanning available databases for all genes and proteins related to each other requires multiple BLAST runs (basic local alignment search tool; http://www.ncbi.nlm.nih.gov/BLAST/). Each individual run within an INCA analysis used gapped BLAST of the v2.0.9 family of programs (19). With gapped BLAST, we found local alignments of similar sequence fragments while allowing for gaps in the alignment. This approach enhances detection of sequence similarities and exceeds nongapped BLAST searches. We have developed an iterative BLAST search termed INCA (http://itsa.ucsf.edu/~gram/home/inca/) (18), which automatically performs BLAST on all sequences identified in the first run with a maximal Expect value (10^{-6} in the present study). In subsequent runs, INCA further tabulates any newly identified sequences scoring with $E < 10^{-6}$, until no further sequences are found. The Expect value estimates the probability that a given alignment score describing sequence similarity could have arisen by chance in the database searched. $E = 10^{-6}$ is often taken as a cutoff below which homology is considered possible while 10^{-9} indicates probable homology (cutoff E values may change as a function of the type of protein and the analysis used). The INCA results are tabulated so that sequences found in subsequent runs are listed with the sequence in the first run yielding the lowest E value (highest similarity). Thus, a first pass with multiple iterative BLASTs identifies all known protein sequences belonging to the POT family, regardless of the starter sequence. In a second pass, we used this entire neighborhood cluster, performing BLAST with each sequence, to scan the human EST database (BLAST, database: 'human ests'). EST clones represent individual mRNA species extracted from target tissues and converted to cDNAs that are then subcloned and partially sequenced. Upon completion of the second INCA pass, the program tabulates the results with each EST assigned to the protein sequence giving the lowest E value. Thereby, INCA provides a simple and exhaustive means of finding all relevant ESTs and assigning them to their most closely related protein sequence in the core cluster of POT sequences. This facilitates the search for new genes that might be represented by ESTs in the accessible databases.

Sequence Analysis

Sequence analysis programs used are available at http://www.sacs.ucsf.edu/. Database homology searches were carried out using the NCBI BLAST v2.0.9 family of programs (19). Contigs of EST sequences (the term contig as used here refers to a contiguous sequence assembled from several overlapping sequence fragments) were assembled using CAP3 (20) and viewed and edited using Sequencher3.0 (www.genecodes.com). The structure of the *hPHT2* gene was characterized with FGENESH (Genefinder: A.A. Salamov and V.V. Solovyev, unpublished data, 1999). Multiple alignments of protein sequences were produced with the Pileup GCGv10 (Wisconsin Package Version 10.0; Genetics Computer Group (GCG), Madison, WI), using oldpep matrix with default parameters. Hydropathy plots were generated using Kyte and Doolittle hydropathy measure (21) and a window size of 7 by Pepplot GCGv10 (Wisconsin Package Version 10.0; Genetics Computer Group (GCG)) with local modifications. The following transmembrane prediction tools were used to produce computer predicted TM topology:

TMHMM(http://www.cbs.dtu.dk/services/TMHMM-1.0/)
TMPRED(http://www.ch.embnet.org/software/TMPRED_

form.html)
HMMTOP(http://www.enzim.hu/hmmtop/), and
MEMSTAT(http://globin.bio.warwick.ac.uk/psipred/)

The transmembrane topology schematic was rendered using TOPO (S.J. Johns and R.C. Speth, Transmembrane protein display software, http://www.sacs.ucsf.edu/TOPO/topo.html, unpublished data). Sequence identities were calculated using the Smith Waterman algorithm (22) by the program ssearch3, a component of the FASTA programs (23).

Reverse Transcriptase-Polymerase Chain Reaction Analysis

Poly $(A)^+$ RNA samples extracted from human intestinal biopsy specimen were analyzed by reverse transcriptase-polymerase chain reaction (RT-PCR). RT-PCR was performed with the GeneAmp RNA PCR Kit Part No. N808-0017 from Perkin Elmer (Wellesley, MA) using 0.5ul AmpliTaq® DNA Polymerase. Alternatively, cDNA samples from skeletal muscle and pancreas (purchased from CLONTECH Laboratories, Palo Alto, CA) were analyzed without RT treatment. The thermocycle included heating at 95°C, annealing at 60°C, reaction temperature at 70°C, for 35 cycles. Primers 5'MHPHT1 (CGTTAGGTGGCATTGCCTAT, left primer at position 562) and 3'MHPHT1 (GAGGATGAGCACAGCATCAA, right primer, at position 1071) were designed with the Primer3 program from the Whitehead Institute for Biomedical Research, Massachusetts Institute of Technology, Cambridge, MA. The expected product size is 509 base pairs. The amplified product was electrophoresed on a 1% Agarose gel, extracted, and sequenced by the University of California San Francisco Human Genetics Sequencing Service, San Francisco, CA.

Northern Blot Analysis of hPHT1 and hPHT2 Expression in Human Tissues

Membrane blots containing size-fractionated poly(A)+ mRNA from 12 tissues of human origin were purchased from CLONTECH Laboratories. The accession codes of the ESTs used as probes are W53019 and AA242853 for *hPHT1* and *hPHT2*, respectively. These cDNA's were labeled with α-^{32}P-dATP (3000 Ci/mmol; Amersham, Piscataway, NJ) according to the random priming method, using a kStrip-EZ DNA kit (Ambion, Austin, TX). Hybridization was performed at 42°C overnight after purifying ^{32}-P-cDNA on a Sephadex G-50 spin column (mini Quick Spin Columns; Boehringer Mannheim, Indianapolis, IN). The blots were washed twice at 42°C for 5 min with low stringency solution (Ambion) and for 15 min with high stringency solution (Ambion). The membranes were exposed against x-ray film at -80°C for 3 to 7 days with intensifying screens. The hybridized probe was removed from the membranes by using a Strip-EZ DNA kit (Ambion) before rehybridization.

RESULTS

Iterative Neighborhood Cluster Analysis (INCA) of the POT Family, First Pass: Nonredundant Protein Databases

Regular BLAST analysis of the protein databases confirmed the presence of three mammalian members reported to belong to the POT family: PepT1, PepT2, and rPHT1. To compile the entire POT family, we then ran INCA with hPepT1 as starter sequence (any sequence is suitable as INCA yields the same sequence cluster regardless of the starter sequence). As shown in Table 2, there are 68 members of the core cluster of POT sequences (as of November 99), each connected to at least one other core sequence with $E = 10^{-6}$ or lower. In the first iteration, 63 sequences were identified by BLAST as possible/probable homologues of hPepT1, while in the second iteration 5 additional sequences appeared in the core cluster of POT.

Table 2. Iterative BLAST analysis using INCA (18). The protein sequence of hPepT1 served as the starter sequence for pass 1 searching the nr protein databases, and two iterations were done. The core cluster contains all sequences scoring with Expect values E < 10-6. These are numbered 1-63, or with a second number to indicate the ranking order of a newly added sequence in iteration 2 (eg, 42.41). Underlining indicates human core sequences; bold-face indicates multi-drug efflux transporter in POT core cluster; bold-face with italics indicates putative C elegans transporters. In pass 2, each of the 68 members of the core cluster was run against the human EST database, and the results tabulated such that the identified ESTs were listed with the core sequence providing the highest score. Note the inclusion of a bacterial drug-resistance transporter in the protein core cluster (bold-face); this sequence would have identified more members of the core cluster in a third iteration (not done here). These include bacterial drug resistance transporters which are currently listed outside the core cluster. Putative POT members from C elegans are shown in bold-face with italics. Further, note that the human ESTs mainly cluster with sequences 11 (hPepT2), 50 (rPHT), and 51 (mouse cAMP-inducible 1 protein).

INCA parameters: pass = 1, iterations 2, similar = 0.0, significant = 1.0E-6, minimum = 0.0, maximum = 10.0. PROGRAM = blastp, DATALIB = nr; ran BLAST 64 times, found 596 total neighbors, 68 inside cluster. Scores include the No. bits and the E value (see http://www.ncbi.nlm.nih.gov/BLAST/)

Core Cluster (E < 10^{-6})

1 04827008 solute carrier family 15 (oligopeptide transp.) [Homo sap] 1430 0.0
2 02832268 (AF043233) Caco-2 oligopeptide transporter [Homo sap] 1430 0.0
3 01136776 (D50306) proton-coupled dipeptide cotransporter [Homo sap] 1213 0.0
4 02143888 oligopeptide transport protein PepT1 [Rat norv] 1210 0.0
5 01730492 OLIGOPEPT. TRANSP., SMALL INTEST. [Rat norv] 1206 0.0
6 00548474 OLIGOPEPT. TRANSP., SMALL INTEST. [Oryct Cunic] 1167 0.0
7 00535426 (U13707) oligopeptide transporter [Oryctolagus cu] 1162 0.0
8 01082041 (L46873) proton-dependent peptide transporter [Rat n]. 792 0.0
9 01585806 peptide transporter [Rattus norvegicus] 760 0.0
10 01172436 OLIGOPEPT. TRANSP., KIDNEY ISOf. [Oryct cunic] 651 0.0
11 02833272 OLIGOPEPT. TRANSP., KIDNEY ISOF. [Homo sap] 636 0.0
12 02499990 OLIGOPEPT. TRANSP., KIDNEY ISOF. [Rat norv] 622 1e-177
13 05901826 (AF181635) BcDNA.GH06717 [Drosoph mel] 495 1e-139
14 03449109 (AL031130) EG:EG0002.1 [Drosophila melanogaster] 435 1e-121
15 03449108 (AL031130) alternatively spliced form [Drosoph mel] 435 1e-121
16 04115344 (U97114) opt1 short [Drosophila melanogaster] 432 1e-120
17 04115343 (U97114) opt1 long [Drosophila melanogaster] 432 1e-120
18 02829749 OLIGOPEPt. TRANSP. 1 (YIN PROTEI.] [Dros mel] 424 1e-117
19 02811011 HYPOTHET. OLIGOPEPT. TRANSP .[C elegans] 397 1e-109
20 02506043 (AB001328) pH-sensing regulat. factor of pept.tr. [Homo sap] 362 4e-99
21 03241977 (AF000417) high-affinity peptide transp. [C elegans]357 1e-97
22 02833290 HYPOTHET. OLIGOPEPT. TRANSP. [C elegans] 357 1e-97
23 03241979 (AF000418) low-affinity peptide transporter [C elegans] 351 1e-95
24 02833308 HYPOTHET. OLIGOPEPT. TRANSP.[C elegans] 319 3e-86
25 01172704 PEPTIDE TRANSPORTER PTR2-B [A thaliana] 177 3e-43
26 01076331 histidine transport protein – [Arabidopsis thal] 176 4e-43
27 04406784 (AC006532) putative oligopept. transport prot.[A thaliana] 175 1e-42
28 04102839 (AF016713) LeOPT1 [Lycopersicon esculentum] 175 1e-42
28.26 03335358 (AC003028) putative peptide transporter [Arabid thal thal] 185 8e-46
29 02655098 (AF023472) peptide transporter [Hordeum vulgare] 168 1e-40
30 02160144 (AC000375) Strong similarity to Arabidopsis olig] 168 1e-40
31 04490321 (AJ011604) nitrate transporter [Arabidopsis] 136 6e-31

32 02760834 (AC003105) putative nitrate transporter [Arab] 136 6e-31
33 01504097 (Y07561) proton-dependent peptide transporter... 135 1e-30
34 01172703 PEPTIDE TRANSPORTER PTR2-A > gi|575427 132 1e-29
35 04895194 (AC007661) putative peptide trans... 130 3e-29
36 02213590 (AC000348) T7N9.10 [Arabidopsis thaliana] 123 6e-27
37 00602292 (U17987) RCH2 protein [Brassica napus] 123 6e-27
38 02651310 (AC002336) putative PTR2-B pept. Transp. [A thal] 122 7e-27
39 00544018 NITRATE/CHLORATE TRANSPORTER [A thaliana] 121 2e-26
40 05734721 (AC008075) Similar to gb|AF023472... 118 1e-25
41 04490323 (AJ131464) nitrate transporter [A thaliana] 118 2e-25
42 02829802 HYPOTHET. 53.3 KD PROTEIN IN SFP-GE[B subtilis] 116 7e-25
42.41 02828140 (AF040248) erythroid different.-related factor [Homo sap] 63 4e-09
43 04455276 (AL035527) peptide transporter-like prot.[A thaliana] 114 3e-24
44 02213586 (AC000348) T7N9.6 [Arabidopsis thaliana] 111 1e-23
45 03377517 (AF073361) nitrate transporter NTL1 [Arabid thal] 109 9e-23
46 00548630 PEPTIDE TRANSP. PTR2 [S cerevis]. 108 1e-22 53/61:54/62
47 00453646 (L11994) permease [Saccharomyces cerevisiae] 1081e-22
48 02829624 HYPOTHETICAL 54.2 KD PROTEIN IN PHRB-N... 107 4e-22
48.2 04062301 (D90709) Hypothetical protein f485 [E coli] 631 1e-180
49 00731995 HYPOTHET. 53.1 KD PROT. IN LYSU-C [E coli] 102 1e-20
50 02208839 (AB000280) peptide/histidine transporter [Rat n]. 100 5e-20
51 04580995 (AF121080) cAMP inducible 1 protein [Mus m] 100 6e-20
52 02367418 (AF000392) peptide transporter [Lotus japonicus] 97 4e-19
53 04678332 (AL049658) putative peptide transporter [Ar thal] 96 7e-19
54 03859684 (AL033503) peptide transport protein [Candida] 88 2e-16
55 02829654 HYPOTHET. 54.0 KD PROT. IN NTH-GS [E coli] 88 2e-16
55.45 04193955 (AF113952) multidrug-efflux transp. [Camp jejuni] 57 3e-07
56 01172741 PEPTIDE TRANSPORTER PTR2 [Candida alb] 86 1e-15
57 01495366 (Z69370) nitrite transporter [Cucumis sativus] 83 9e-15
57.27 05360083 (AF154930) transporter-like protein 88 2e-16
58 02507268 HYPOTHET. 53.7 KD PROT. IN USPA-PRLC [E coli] 82 1e-14
59 01073520 hypothetical protein o489 - E coli 82 1e-14
60 05080808 (AC007258) Similar to nitrate tr. [A thaliana] 79 1e-13
61 04322327 (AF080545) peptide transporter [Nepenthes alata] 76 1e-12
62 02811053 DI-/TRIPEPTIDE TRANSPORTER [Lactobac helv] 64 3e-09
63 00544192 DI-/TRIPEPTIDE TRANSP. [Lact lactis] 62 1e-08
Displaying neighbors outside cluster (10^{-6} > E < 0.01)
13.65 00071400 dermal gland protein APEG precursor - African c... 47 6e-04
13.66 00731172 SKIN SECRETORY PROTEIN XP2 PRECURSOR (A... 47 6e-04
13.67 05802676 (AF177977) serum opacity factor pr... 46 0.001
13.68 05701582 (AF026205) No definition line foun... 45 0.002
13.69 03877270 (Z77662) predicted using Genefinder [Caenor... 45 0.002
13.70 02435546 (AF026205) No definition line found [Caenorh... 45 0.002
13.71 05748800 (AF141140) serum opacity factor pr... 45 0.003
13.72 05002375 (AF153315) serum opacity factor pr... 44 0.004
13.73 02435547 (AF026205) No definition line found [Caenorh... 44 0.004
13.74 00477578 sialidase - Actinomyces viscosus > gi|141852 (L0... 44 0.004
13.75 02275336 (AF001978) differentially expressed in relation ... 44 0.005

13.76 02507049 HYPHAL WALL PROTEIN 1 (CELL ELONGATION... 44 0.005
13.77 05139301 (AF157555) serum opacity factor pr... 43 0.007
13.78 01781122 (Z83864) hypothetical protein Rv3835 [Mycobac... 43 0.007
13.79 01480457 (U42640) latex allergen [Hevea brasiliensis] > gi... 43 0.007
13.80 01420865 (X80397) orf1 [Streptococcus pyogenes] 43 0.009
22.64 02621135 (AE000800) multidrug transporter homolog [Methan... 52 1e-05
42.55 03820455 (AJ007367) multi-drug resistance efflux pump ... 51 2e-05
42.59 02695718 (AJ001694) putative membrane protein [Thermot... 48 2e-04
42.61 04467970 (X76640) hypothetical protein [Myxococcus xan... 46 6e-04
42.62 02618837 (AF017113) YvkA [Bacillus subtilis] > gi|2636047|... 450.001
42.64 05457646 (AJ248283) MULTIDRUG RESISTANCE PROTEIN [Py... 44 0.003
45.59 02827716 (AL021684) predicted protein [Arabidopsis tha... 43 0.005
48.52 02828202 BILE ACID TRANSPORTER >gi|1381569 (U57... 54 3e-06
48.55 01842056 (U87258) cis,cis-muconate transport protein MucK... 49 1e-04
48.58 04885441 Na/PO4 cotransporter >gi|4587207... 44 0.003
48.59 02225983 (Z97193) hypothetical protein Rv1877 [Mycobac... 43 0.005
48.61 00586828 HYPOTHETICAL 44.2 KD PROTEIN IN COTF-T... 43 0.005
49.57 00401611 D-GALACTONATE TRANSPORTER >gi|290540 (... 45 0.001
49.58 00586812 HYPOTHETICAL 43.2 KD PROTEIN IN DNAC-R... 43 0.007
53.56 02500934 SHIKIMATE TRANSPORTER >gi|1736645|dbj|... 45 0.001
55.53 04753872 (AL049754) putative transmembrane efflux pr... 49 7e-05
55.61 02078013 (Z95207) efpA [Mycobacterium tuberculosis] 46 8e-04
55.62 01161051 (L39922) efflux protein [Mycobacterium tuberculo... 46 8e-04
55.64 04455672 (AL035472) putative transmembrane efflux pr... 44 0.002
55.65 00586001 METHYL VIOLOGEN RESISTANCE PROTEIN SMV... 44 0.003
55.66 02808775 (AL021411) integral membrane protein [Strepto... 43 0.004
55.67 02695836 (AL021006) hypothetical protein Rv1250 [Mycob... 43 0.004
55.68 02808785 (AL021411) putative transport protein [Strept... 43 0.005
58.43 02650264 (AE001079) oxalate/formate antiporter (oxlT-2) [... 53 5e-06
59.45 02127150 nitrate transporter - Bacillus subtilis (fragment) 506e-05
59.46 01171658 NITRATE TRANSPORTER >gi|1437473|dbj|BA... 50 6e-05
59.53 02222715 (Z97179) hypothetical protein MLCL383.37 [Myc... 46 5e-04
59.56 03257679 (AP000005) 372aa long hypothetical protein [P... 45 0.001
59.58 05777416 (AJ249180) sugar efflux transpoter [Erwinia... 43 0.004
59.59 01787304 (AE000207) orf, hypothetical protein [Escherichi... 43 0.004
59.60 02501163 HYPOTHETICAL 44.4 KD PROTEIN IN GRXB-R... 43 0.004
59.61 02982930 (AE000678) nitrate transporter [Aquifex aeolicus] 43 0.005
63.49 02612908 (AF015825) hexuronate transporter-like protein [... 51 2e-05
63.52 05457697 (AJ248283) TRANSPORTER [Pyrococcus abyssi] 48 1e-04
63.53 03116222 (AB007122) transporter [Arthrobacter sp.] 47 3e-04
63.57 01346939 ANTISEPTIC RESISTANCE PROTEIN > gi|7733... 45 0.001 0/0:0/3
63.58 00097843 probable transport protein qacA - Staphylococcu... 44 0.003
63.59 03327943 (AF053771) multidrug efflux protein QacB [Staphy... 43 0.004
63.60 03097809 (L49465) hypothetical metabolite transport prote... 43 0.004
63.62 05650769 (AF089813) nitrate transporter [Sy... 43 0.007
63.63 03327948 (AF053772) multidrug efflux protein QacB [Staphy... 43 0.007
63.64 01706916 FOSMIDOMYCIN RESISTANCE PROTEIN >gi|212... 43 0.007

Pass = 2, iteration = 1, similar = 0.0, significant = 1.0E-6, minimum = 0.0, maximum = 10.0. PROGRAM = tblastn, IPROGRAM = tblastx, DATALIB = est_human. Ran BLAST 68 times, found 125 total neighbors, 73 inside cluster, 52 outside cluster.

Core human EST cluster, 73 neighbors ($E < 10^{-6}$)

11.1 03163329 am93c10.s1 Stratagene schizo brain S11... 277 1e-75
11.2 01544978 zf48e05.r1 Soares retina N2b4HR Homo s... 213 2e-62
11.3 02575676 ns35d11.s1 NCI_CGAP_GCB1 Homo sapiens ... 118 5e-32
11.4 00575005 H. sapiens partial cDNA sequence; clon... 118 1e-25
11.5 04736822 wb22d10.x1 NCI_CGAP_GC6 Homo sapiens... 106 1e-21
11.6 00574634 H. sapiens partial cDNA sequence; clon... 65 7e-17
11.8 01886510 zs10d12.s1 NCI_CGAP_GCB1 Homo sapiens ... 63 8e-09
42.1 03894461 ap21e10.x1 Schiller oligodendroglioma ... 65 2e-09
48.1 01747695 zp18b11.r1 Stratagene fetal retina 937... 98 3e-19
50.1 01349645 zc48d11.r1 Soares senescent fibroblasts Nb... 344 1e-93
50.2 05396524 wf64b06.x1 Soares_NFL_T_GBC_S1 Homo ... 303 3e-81
50.3 05663458 wo90e12.x1 NCI_CGAP_Kid11 Homo sapie... 301 1e-80
50.4 05675802 wp71b07.x1 NCI_CGAP_Brn25 Homo sapie... 298 1e-79
50.5 05109278 wg25b07.x1 Soares_NSF_F8_9W_OT_PA_P_... 296 4e-79
50.6 06037457 xb65b04.x1 Soares_NFL_T_GBC_S1 Homo ... 285 9e-76
50.7 04896654 wa76f08.x1 Soares_NFL_T_GBC_S1 Homo ... 265 7e-70
50.8 05395461 wf66c05.x1 Soares_NFL_T_GBC_S1 Homo ... 263 4e-69
50.9 05396659 wf65d06.x1 Soares_NFL_T_GBC_S1 Homo ... 250 2e-65
50.10 05450343 wf09b03.x1 Soares_NFL_T_GBC_S1 Homo ... 247 3e-64
50.11 05391793 td11g01.x1 NCI_CGAP_CLL1 Homo sapien... 243 3e-63
50.12 00816183 yg78d02.r1 Homo sapiens cDNA clone 39680 5... 227 5e-62
50.13 03899028 ql64f04.x1 Soares_NhHMPu_S1 Homo sapie... 238 1e-61
50.14 03842747 qh34g06.x1 Soares_NFL_T_GBC_S1 Homo ... 230 3e-59
50.15 01046571 ys10b07.r1 Homo sapiens cDNA clone 214357 5'. 227 3e-58
50.16 01024648 yr69b03.r1 Homo sapiens cDNA clone 210509 ... 218 7e-56
50.17 03895536 ql54e10.x1 Soares_NhHMPu_S1 Homo sapie... 211 1e-53
50.18 06075659 xd72f06.x1 Soares_NFL_T_GBC_S1 Homo ... 207 2e-52
50.19 03933090 qm02b04.x1 Soares_NhHMPu_S1 Homo sapie... 205 8e-52
50.20 05542610 tc51g08.x1 Soares_NhHMPu_S1 Homo sap... 205 1e-51
50.21 05448546 wf04d03.x1 Soares_NFL_T_GBC_S1 Homo ... 199 4e-50
50.22 03804209 qg95d09.x1 Soares_NFL_T_GBC_S1 Homo sa... 199 6e-50
50.23 05437154 wk14c11.x1 NCI_CGAP_Lym12 Homo sapie... 197 2e-49
50.24 04739863 tt41a11.x1 NCI_CGAP_GC6 Homo sapiens... 166 3e-48
50.25 01976192 EST26707 Cerebellum II Homo sapiens cD... 185 7e-46
50.28 01616355 zm91d04.r1 Stratagene ovarian cancer (... 104 9e-32
50.30 02779368 nz03d12.s1 NCI_CGAP_GCB1 Homo sapiens ... 132 6e-30
50.31 01933517 zs50g05.s1 NCI_CGAP_GCB1 Homo sapiens ... 99 7e-29
50.33 01470970 zi07e01.r1 Soares fetal liver spleen 1... 124 2e-27
50.34 02849230 aa66d09.s1 NCI_CGAP_GCB1 Homo sapiens ... 122 7e-27
50.35 03253724 ow73h02.s1 Soares_fetal_liver_spleen_1... 122 1e-26
50.36 00750175 ye72b12.r1 Homo sapiens cDNA clone 123263 5'. 112 1e-23
50.37 02783232 ny25b10.s1 NCI_CGAP_GCB1 Homo sapiens ... 112 1e-23
50.38 00900981 yp44h09.r1 Homo sapiens cDNA clone 190337 5'. 109 5e-23
50.39 00900971 yp44f09.r1 Homo sapiens cDNA clone 190313 5'. 109 5e-23

50.42 02018481 EST77078 Pancreas tumor III Homo sapie... 95 2e-18
50.44 01989422 EST41878 Endometrial tumor Homo sapien... 92 1e-17
50.45 01024562 yr69b03.s1 Homo sapiens cDNA clone 210509 3'. 90 4e-17
50.46 01324288 zc82h02.r1 Pancreatic Islet Homo sapiens c... 90 5e-17
50.47 01616244 zm91d04.s1 Stratagene ovarian cancer (... 90 7e-17 0/0:2/3
50.49 03076109 om08f11.s1 Soares_NFL_T_GBC_S1 Homo sa... 81 3e-14
50.50 01858092 zr49e07.r1 Soares NhHMPu S1 Homo sapie... 51 6e-14
50.51 04688089 ts89f09.x1 NCI_CGAP_GC6 Homo sapiens... 80 8e-14
50.53 01959622 EST178192 Colon carcinoma (HCC) cell l... 75 1e-12
50.55 03838659 qh35f01.x1 Soares_NFL_T_GBC_S1 Homo ... 69 1e-10
50.57 03049244 om32h08.s1 Soares_NFL_T_GBC_S1 Homo sa... 65 2e-09
50.58 01056010 ys10b07.s1 Homo sapiens cDNA clone 214357 3'. 57 2e-09
51.1 01873671 zr64d03.r1 Soares NhHMPu S1 Homo sapie... 238 1e-67
51.2 03649141 ot08d05.x1 NCI_CGAP_GC3 Homo sapiens c... 234 2e-60
51.3 03920067 qt87b05.x1 NCI_CGAP_Co14 Homo sapiens ... 210 4e-53
51.5 05863189 UI-H-BI0-aac-g-02-0-UI.s1 NCI_CGAP_S... 169 8e-41
51.6 00668137 yc09b07.r1 Homo sapiens cDNA clone 80149 5'. 104 4e-39
51.7 03162566 oq02f07.s1 NCI_CGAP_Lu5 Homo sapiens c... 160 4e-38
51.12 01192150 yy94d05.s1 Homo sapiens cDNA clone 281193 3'. 140 4e-32
51.14 02899359 od60b01.s1 NCI_CGAP_GCB1 Homo sapiens ... 122 3e-29
51.24 03155247 oq67h11.s1 NCI_CGAP_Kid6 Homo sapiens ... 112 1e-23
51.32 04112719 qy04e10.x1 NCI_CGAP_Brn23 Homo sapiens... 104 2e-21
51.37 02163000 zx07d11.r1 Soares total fetus Nb2HF8 9... 97 5e-19
51.39 03675523 oz69f11.x1 Soares_senescent_fibroblast... 95 2e-18
51.40 00761243 yf26a08.r1 Homo sapiens cDNA clone 127958 ... 94 3e-18
51.41 03180750 ou49a11.s1 NCI_CGAP_Br2 Homo sapiens c... 94 5e-18
51.42 01887135 zr64d03.s1 Soares NhHMPu S1 Homo sapie... 89 2e-16
51.49 00668009 yc09b07.s1 Homo sapiens cDNA clone 80149 3'. 70 5e-11
51.52 03918968 qw07b08.x1 NCI_CGAP_Ut3 Homo sapiens c... 62 2e-08
Displaying human EST neighbors outside cluster (10^{-6} < E < 0.01)
50.61 02788570 ny06b02.s1 NCI_CGAP_GCB1 Homo sapiens ... 55 2e-06
50.65 02820068 nz64g10.s1 NCI_CGAP_GCB1 Homo sapiens ... 50 8e-05
50.66 02752650 nw56h09.s1 NCI_CGAP_GCB1 Homo sapiens ... 50 8e-05
50.67 02167640 zx45c08.r1 Soares testis NHT Homo sapi... 45 0.003
55.45.1 05130954 cn18b12.x1 Normal Human Trabecular B... 55 2e-06

Relationship of POT Family to Other Transporters

Earlier INCA runs had yielded similar results but with fewer sequences in the core cluster. The same analysis performed a year earlier revealed only 46 core sequences and converged after two iterations; that is, further iterations did not reveal any new sequences scoring with $E < 10^{-6}$. This had suggested that the POT family shows rather unique sequence characteristics, with no sequence from other transporter families scoring with $E = 10^{-6}$ or lower. However, continued deposition of new sequences has enlarged the POT family and general databases considerably. This could result in the discovery of links to other transporter families possibly related to POT in evolution. Indeed, the recent INCA run shown in Table 2 identified a distinct sequence with $E < 10^{-6}$ belonging to the family of bacterial drug resistance transporters, a multidrug-efflux transporter of *Campylobacter* (Table 2, core cluster; bold-face type). To avoid including numerous multidrug-efflux transporters with the POT core family in a third

iteration of BLAST, Table 2 contains only two iterations. A number of these drug-resistance transporters of the major facilitator type transporter family appear in the list of neighbor sequences outside the core cluster (Table 2). Several distinct types of transporters reach E values of ~10^{-5} (eg, the bile acid transporter gi/1381596 [sequence 48.52, Table 2], and the hexuronate transporter Af015825 [sequence 63.49] reported earlier (24)). These results support the supposition that the peptide transporter family POT is indeed related in evolution to other transporter families. We will pursue these relationships to better understand the structure and evolution of the peptide transporter family.

Search for New POT Members

The core cluster contains 5 putative POT genes from the completely sequenced genome of *C elegans* (Table 2, bold-face and italics). Each of these deduced proteins has high similarity to hPepT1. This finding suggests that the human genome may also contain more POT members than are currently cloned.

Table 2 includes a number of deposited sequences encoding the main intestinal and renal transporters, hPepT1 (sequences 1-3) and hPepT2 (sequence 12), and their orthologues in other mammalian species. The pH sensing regulatory factor of peptide transporter (sequence 20) (25) appears to represent a possible truncated splice isoform of hPepT1. Also shown in the core cluster of Table 2, the rat peptide/histidine transporter (rPHT1) (sequence 50; E = 5×10^{-50} against hPepT1) is adjacent to a new sequence with high similarity to it, namely the mouse cAMP-inducible protein 1 (sequence 51). The latter was recently cloned from a lymphoid cell line and had not been suspected to belong to the POT family (26). We will demonstrate below that both rPHT1 and mouse cAMP-inducible 1 protein have apparent human orthologues, which we will term hPHT1 and hPHT2, respectively. This suggests that rPHT1 and mouse cAMP-inducible 1 protein represent two distinct but closely related genes belonging to the POT family.

The core cluster of POT sequences contains one additional human sequence, namely, erythroid differentiation-related factor 2 (sequence 42.41, Table 2). The E value (4×10^{-9}) suggests probable homology to a putative POT transporter of *Bacillus subtilis* (sequence 42; 13 predicted TMDs). However, this sequence is rather short (107 residues), showing good sequence similarity with TMD11 and adjacent loop of the *B subtilis* transporter. More studies are needed to discover the biological relevance of this finding.

In summary, the INCA analysis of the protein sequence databases revealed the presence of 4 mammalian sequences as probable members of POT: PepT1, PepT2, PHT1, and cAMP-inducible 1 protein (the latter two termed PHT1 and PHT2 in our nomenclature).

Second Pass INCA: Scanning the Human EST Database

All 68 sequences in the core cluster (Table 2) served in a second INCA pass to search the human EST database. The resultant list of core ESTs ($E < 10^{-6}$), sorted by highest similarity to one of the 68 core protein sequences, is also included with Table 2. Curiously, no EST scored best with PepT1, presumed to be the main intestinal transporter in rodents and, by inference, in humans (intestinal ESTs may be underrepresented in the available EST databases). Further, seven ESTs assorted with the cloned human hPepT2; however, it remains to be seen whether all seven are indeed representatives of hPepT2. With E scores exceeding 10^{-10}, it is possible that these ESTs with moderate similarity could represent distinct genes or splice variants (not analyzed further). Numerous human ESTs scored best with the rat peptide histidine transporter, rPHT (sequence 50), and with mouse cAMP-inducible 1 protein (sequence 51). We therefore assembled these ESTs into contiguous sequences to identify the respective human orthologous gene products, termed hPHT1 and hPHT2.

In the human EST core cluster (Table 2), two single ESTs scored best with putative peptide transporters from bacteria (sequences 42 and 48 of the core

proteins of POT). Additional analysis will be required to ascertain whether the respective human genes suggested by these two ESTs could represent yet additional members of the human POT gene family.

hPHT1 and hPHT2 Sequences Assembled from Human ESTs

Scanning the human EST database, we have identified numerous ESTs that appear to represent the human orthologues of rPHT and cAMP-inducible 1 protein. Table 3 lists the ESTs firmly assigned to the presumed human orthologue of rPHT1 (hPHT1) and of cAMP-inducible 1 protein (termed here hPHT2) (cutoff E value ~10^{-20}). For comparison, Table 3 also displays the ESTs assigned to hPepT2, whereas no ESTs appeared to represent hPepT1. The tissue source of the numerous ESTs representing *hPHT1* is of considerable interest because in the rat, rPHT expression is largely confined to the brain and retina. Yet, the *hPHT1*-related ESTs derive from many body tissues, including colon carcinoma. This raises the possibility that hPHT1 is broadly expressed in many human tissues.

Table 3. ESTs from BLAST (blastn) assigned to hPepT2, hPHT1, and hPHT2. Criteria of inclusion for alignments are >90% identity in sequences longer than 60 bps.

hPepT2 vs human EST:

gi|3163329|gb|AA984804 am93c10.s1 Stratagene schizo br... 277 1e-75
gi|1544978|gb|AA054054 zf48e05.r1 Soares retina N2b4HR... 213 2e-62
gi|2575676|gb|AA649247 ns35d11.s1 NCI_CGAP_GCB1 Homo s... 118 5e-32
gi|575005 |emb|Z45771|HSCZTG061H. sap. partial cDNA sequence 118 1e-25

hPHT1 vs human EST

(ESTs from single tissues are bold-face)

gi|1349645|gb|W53019 zc48d11.r1 Soares *senescent fibroblasts* 1056 0.0
gi|5396524|gb|AI809958.1 wf64b06.x1 Soares_NFL_T_GBC_S... 955 0.0
gi|5663458|gb|AI927494.1 wo90e12.x1 NCI_CGAP_*Kidney* Homo sap 949 0.0
gi|5109278|gb|AI740990.1 wg25b07.x1 Soares_NSF_F8_9W_O... 947 0.0
gi|5675802|gb|AI936932.1 wp71b07.x1 NCI_CGAP_*Brain* Homo sap 945 0.0
gi|5450343|gb|AI829672.1 wf09b03.x1 Soares_NFL_T_GBC_S... 859 0.0
gi|5396659|gb|AI810093.1 wf65d06.x1 Soares_NFL_T_GBC_S... 859 0.0
gi|4896654|gb|AI685360.1 wa76f08.x1 Soares_NFL_T_GBC_S... 857 0.0
gi|5395461|gb|AI808895.1 wf66c05.x1 Soares_NFL_T_GBC_S... 850 0.0
gi|3899028|gb|AI276754 ql64f04.x1 Soares_NhHMPu_S1 Hom... 800 0.0
gi|1470970|gb|AA009923 zi07e01.r1 Soares *fetal liver spleen* 792 0.0
gi|5391793|gb|AI805139.1 td11g01.x1 NCI_CGAP_CLL1 Homo... 790 0.0
gi|3933090|gb|AI290316 qm02b04.x1 Soares_NhHMPu_S1 Hom... 753 0.0
gi|3842747|gb|AI247350.1 qh34g06.x1 Soares_NFL_T_GBC_S... 747 0.0
gi|1024648|gb|H65908 yr69b03.r1 Homo sapiens cDNA clone ... 745 0.0
gi|1046571|gb|H73031 ys10b07.r1 Homo sapiens cDNA clone ... 743 0.0
gi|4739863|gb|AI655884.1 tt41a11.x1 NCI_CGAP_GC6 Homo ... 699 0.0
gi|1616355|gb|AA076486 zm91d04.r1 *Stratagene ovarian cancer* 681 0.0
gi|3895536|gb|AI273268 ql54e10.x1 Soares_NhHMPu_S1 Hom... 665 0.0
gi|816183 |gb|R54281 yg78d02.r1 Homo sapiens cDNA clone 3... 624 e-177
gi|5437154|gb|AI818075.1|wk14c11.x1 NCI_CGAP *lymph node* H sap 590e-166
gi|3804209|gb|AI222006 qg95d09.x1 Soares_NFL_T_GBC_S1 ... 564 e-158
gi|5448546|gb|AI827875.1 wf04d03.x1 Soares_NFL_T_GBC_S... 547 e-153
gi|1324288|gb|W40166 zc82h02.r1 *Pancreatic Islet Homo sap.* 543 e-152

gi|1976192|gb|AA323865 EST26707 *Cerebellum II Homo sap.* 531 e-149
gi|2779368|gb|AA740776 nz03d12.s1 NCI_CGAP_GCB1 Homo s… 525 e-147
gi|1933517|gb|AA287835 zs50g05.s1 NCI_CGAP_GCB1 Homo s… 432 e-119
gi|750175 |gb|R00439 ye72b12.r1 Homo sapiens cDNA clone 1... 420 e-115
gi|3049244|gb|AA909954 om32h08.s1 Soares_NFL_T_GBC_S1 ... 412 e-113
gi|3253724|gb|AI032598 ow73h02.s1 *Soares_fetal_liver_spleen* 368 1e-99
gi|900971 |gb|H30061 yp44f09.r1 Homo sapiens cDNA clone 1... 329 5e-88
gi|900981 |gb|H30071 yp44h09.r1 Homo sapiens cDNA clone 1... 329 5e-88
gi|2783232|gb|AA743881 ny25b10.s1 NCI_CGAP_GCB1 Homo s... 325 8e-87
gi|1858092|gb|AA234173 zr49e07.r1 Soares NhHMPu S1 Hom... 323 3e-86
gi|2849230|gb|AA789110 aa66d09.s1 NCI_CGAP_GCB1 Homo sap 293 4e-77
gi|1989422|gb|AA337185 EST41878 *Endometrial tumor Homo sap* 287 3e-75
gi|1959622|gb|AA307073 EST178192 *Colon carcinoma* (HCC)... 265 1e-68
gi|1616244|gb|AA076314 zm91d04.s1 *Stratagene ovarian cancer* 261 2e-67
gi|1024562|gb|H65822 yr69b03.s1 Homo sapiens cDNA clone ... 230 3e-58
gi|3076109|gb|AA927212 om08f11.s1 Soares_NFL_T_GBC_S1 ... 228 1e-57
gi|4688089|gb|AI636759.1 ts89f09.x1 NCI_CGAP_GC6 Homo ... 198 2e-48
gi|1056010|gb|H77921 ys10b07.s1 Homo sapiens cDNA clone ... 196 6e-48
gi|2752650|gb|AA731761 nw56h09.s1 NCI_CGAP_GCB1 Homo s... 154 4e-35
gi|2788570|gb|AA748612 ny06b02.s1 NCI_CGAP_GCB1 Homo s... 151 2e-34
gi|2820068|gb|AA768830 nz64g10.s1 NCI_CGAP_GCB1 Homo s... 147 3e-33
gi|3075545|gb|AA926648 om28b11.s1 Soares_NFL_T_GBC_S1 ... 119 8e-25
gi|3094496|gb|AA936578 on78f12.s1 Soares_NFL_T_GBC_S1 ... 111 2e-22
gi|3052867|gb|AA913475 ol30h05.s1 Soares_NFL_T_GBC_S1 ... 109 9e-22

hPHT2 vs human EST :

gi|1873671|gb|AA242853 zr64d03.r1 Soares NhHMPu S1 Hom... 795 0.0
gi|668137 |gb|T64272 yc09b07.r1 Homo sapiens cDNA clone 8... 541 e-152
gi|2163000|gb|AA448980 zx07d11.r1 Soares total fetus N... 379 e-103
gi|3162566|gb|AA984041 oq02f07.s1 NCI_CGAP_Lu5 Homo sa... 311 1e-82
gi|761243 |gb|R09320 yf26a08.r1 Homo sapiens cDNA clone 1... 301 1e-79
gi|3649141|gb|AI141684 ot08d05.x1 NCI_CGAP_GC3 Homo sa... 155 1e-35

Approximately 50 ESTs served to assemble a contig DNA sequence of the presumed *hPHT1* nucleotide sequence Figure 1A and Figure 1B; Table 4). A schematics of the *hPHT1* contig assembly Figure 1A contains the minimum number of EST's spanning the length of the deduced *hPHT1* sequence. To view the EST coverage of each segment, click on the EST/region of interest. This will reveal segments with numerous overlapping ESTs. Because multiple ESTs cover the same regions of *hPHT1*, one can deduce possible sequence variations in the human population where EST sequences are not identical. Clearly, many of these variations may be due to sequencing errors, but in a few cases a variation occurs more than once at the same location. These variations increase the probability that a genetic variant might be involved. The contig sequence is closely related to that of *rPHT1*; however, the first ~50 5'-terminal base pairs are missing. (It appears that there is a rare *Not*I site at the 5'-end, which could have caused truncation during preparation for EST sequence analysis.) Thus, the EST contig is likely to represent >95% of the coding region of *hPHT1*. The deduced hPHT1 amino acid sequence is shown in Table 4.

Table 4. Deduced amino acid and nucleotide sequences of hPHT1 and hPHT2

hPHT1 peptide sequence

hpht1-contig from 1 to 1671, generated symbols 1 to 557

```
   1  AAAAAAGAFA GRRAACGAVL LTELLERAAF YGITSNLVLF LNGAPFCWEG
  51  AQASEALLLF MGLTYLGSPF GGWLADARLG RARAILLSLA LYLLGMLAFP
 101  LLAAPATRAA LCGSARLLNC TAPGPDAAAR CCSPATFAGL VLVGLGVATV
 151  KANITPFGAD QVKDRGPEAT RRFFNWFYWS INLGAILSLG GIAYIQQNVS
 201  FVTGYAIPTV CVGLAFVRFL CGQSVFITKP PDGSAFTDMF KILTYSCCSQ
 251  KRSGERQSNG EGIGVFQQSS KQSLFDSCKM SHGGPFTEEK VEDVKALVKI
 301  VPVFLALIPY WTVYFQMQTT YVLQSLHLRI PEISNITTTP HTLPAAWLTM
 351  FDAVLILLLI PLKDKLVDPI LRRHGLLPSS LKRIAVGMFF VMCSAFAAGI
 401  LESKRLNLVK EKTINQTIGN VVYHAADLSL WWQVPQYLLI GISEIFASIA
 451  GLEFAYSAAP KSMQSAIMGL FFFFSGVGSF VGSGLLALVS IKAIGWMSSH
 550  TDFGNINGCY LNYYFFLLAA IQGATLLLFL IISVKYDHHR DHQRSRANGV
 551  PTSRRA*
```

hPHT1 nucleotide sequence

Length: 1779

```
   1  GCGGCCGCGG CGGCGGCTGG GGCGTTCGCG GGCCGGCGCG CGGCGTGCGG
  51  GGCCGTGCTG CTGACGGAGC TGCTGGAGCG CGCCGCTTTC TACGGCATCA
 101  CGTCCAACCT GGTGCTATTC CTGAACGGGG CGCCGTTCTG CTGGGAGGGC
 151  GCGCAGGCCA GCGAGGCGCT GCTGCTCTTC ATGGGCCTCA CCTACCTGGG
 201  CTCGCCGTTC GGAGGCTGGC TGGCCGACGC GCGGCTGGGC CGGGCGCGCG
 251  CCATCCTGCT GAGCCTGGCG CTCTACCTGC TGGGCATGCT GGCCTTCCCG
 301  CTGCTGGCCG CGCCCGCCAC GCGAGCCGCG CTCTGCGGTT CCGCGCGCCT
 351  GCTCAACTGC ACGGCGCCTG GTCCCGACGC CGCCGCCCGC TGCTGCTCAC
 401  CGGCCACCTT CGCGGGGCTG GTGCTGGTGG GCCTGGGCGT GGCCACCGTC
 451  AAGGCCAACA TCACGCCCTT CGGCGCCGAC CAGGTTAAAG ATCGAGGTCC
 501  GGAAGCCACT AGGAGATTTT TTAATTGGTT TTATTGGAGC ATTAACCTGG
 551  GAGCGATCCT GTCGTTAGGT GGCATTGCYT ATATTCAGCA GAACGTYAGC
 601  TTTGTCACTG GTTATGCGAT CCCCACTGTC TGCGTCGGCC TTGCTTTTGT
 651  GGYCTTCCTC TGTGGCCAGA GCGTTTTCAT CACCAAGCCT CCTGATGGCA
 701  GTGCCTTCAC CGAYATGTTC AAGATACTGA CGTATTCCTG CTGTTCCCAG
 751  AAGCGAAGTG GAGAGCGCCA GAGTAATGGT GAAGGCATTG GAGTCTTTCA
 801  GCAATCTTCT AAACAAAGTC TGTTTGATTC ATGTAAGATG TCTCATGGTG
 851  GGCCATTTAC AGAAGAGAAA GTGGAAGATG TGAAAGCTCT GGTCAAGATT
 901  GTCCCTGTTT TCTTGGCTTT GATACCTTAC TGGACAGTGT ATTTCCAAAT
 951  GCAGACAACA TATGTTTTAC AGAGTCTTCA TTTGAGGATT CCAGAAATTT
1001  CAAATATTAC AACCACTCCT CACACGCTCC CTGCAGCCTG GCTGACCATG
1051  TTTGATGCTG TGCTCATCCT CCTGCTCATC CCTCTGAAGG ACAAACTGGT
1101  CGATCCCATT TTGAGAAGAC ATGGCCTGCT CCCATCCTCC CTGAAGAGGA
1151  TCGCCGTGGG CATGTTCTTT GTCATGTGCT CGGCCTTTGC TGCAGGAATT
1201  TTGGAGAGTA AAAGGCTGAA CCTTGTTAAA GAGAAAACCA TTAATCAGAC
1251  CATCGGCAAC GTCGTCTACC ATGCTGCCGA TCTGTCGCTG TGGTGGCAGG
1301  TGCCGCAGTA CTTGCTGATT GGGATCAGCG AGATCTTTGC AAGTATCGCA
1351  GGCCTGGAAT TTGCATACTC AGCTGCCCCC AAGTCCATGC AGAGTGCCAT
```

1401 AATGGGCTTG TTCTTTTTCT TCTCTGGCGT CGGGTCGTTC GTGGGTTCTG
1451 GACTGCTGGC ACTGGTGTCT ATCAAAGCCA TCGGATGGAT GAGCAGTCAC
1501 ACAGACTTTG GTAATATTAA CGGCTGCTAT TTGAACTATT ACTTTTTCCT
1551 TCTGGCTGCT ATTCAAGGAG CTACCCTCCT GCTTTTCCTC ATTATTTCTG
1601 TGAAATATGA CCATCATCGA GACCATCAGC GATCAAGAGC CAATGGCGTG
1651 CCCACCAGCA GGAGGGCCTG ACCTTCCTGA GGCCAYGTGC GGTTTCTGAR
1701 GCTGACATGT CAGTAACTGA CTGGGGTGCA CTGAGAACAG GCAAGACTTT
AAATTCCCAT AAAATGTCTG ACTTCACTG

hPHT2 peptide sequence

8 exon (s) 22937 - 36932 chain +
hPHT2.pep, length: 580

1 MPAPRAREQP RVPGERQPLL PRGARGPRRW RRAAGAAVLL VEMLERAAFF
51 GVTANLVLYL NSTNFNWTGE QATRAALVFL GASYLLAPVG GWLADVYLGR
101 YRAVALSLLL YLAASGLLPA TAFPDGRSSF CGEMPASPLG PACPSAGCPR
151 SSPSPYCAPV LYAGLLLLGL AASSVRSNLT SFGADQVMDL GRDATRRFFN
201 WFYWSINLGA VLSLLVVAFI QQNISFLLGY SIPVGCVGLA FFIFLFATPV
251 FITKPPMGSQ VSSMLKLALQ NCCPQLWQRH SARDRQCARV LADERSPQPG
301 ASPQEDIANF QVLVKILPVM VTLVPYWMVY FQMQSTYVLQ GLHLHIPNIF
351 PANPANISVA LRAQGSSYTI PEAWLLLANV VVVLILVPLK DRLIDPLLLR
401 CKLLPSALQK MALGMFFGFT SVIVAVLEME RLHYIHHNET VSQQIGEVLY
451 NAAPLSIWWQ IPQYLLIGIS EIFASIPGLE FAYSEAPRSM QGAIMGIFFC
501 LSGVGSLLGS SLVALLSLPG GWLHCPKDFG NINNCRMDLY FFLLAGIQAV
551 TALLFVWIAG RYERASQGPA SHSRFSRDRG

hPHT2 nucleotide sequence, deduced coding region, length: 1740

>AC004126 ASSEMBLE August 2, 1999 12:06
ATGCCCGCGCCGCGCGCCCGGGAGCAGCCCCGCGTGCCCGGGGAGCGCCAGCCGCTGCTG
CCTCGCGGTGCGCGGGGCCCTCGACGGTGGCGGCGGGCGGCGGGCGCGGCCGTGCTGCTG
GTGGAGATGCTGGAGCGCGCCGCCTTCTTCGGCGTCACCGCCAACCTCGTGCTGTACCTC
AACAGCACCAACTTCAACTGGACCGGCGAGCAGGCGACGCGCGCCGCGCTGGTATTCCTG
GGCGCCTCCTACCTGCTGGCGCCCGTGGGCGGCTGGCTGGCCGACGTGTACCTGGGCCGC
TACCGCGCGGTCGCGCTCAGCCTGCTGCTCTACCTGGCCGCCTCGGGCCTGCTGCCCGCC
ACCGCCTTCCCCGACGGCCGCAGCTCCTTCTGCGGAGAGATGCCCGCGTCGCCGCTGGGA
CCTGCCTGCCCCTCGGCCGGCTGCCCGCGCTCCTCGCCCAGCCCCTACTGCGCGCCCGTC
CTCTACGCGGGCCTGCTGCTACTCGGCCTGGCCGCCAGCTCCGTCCGGAGCAACCTCACC
TCCTTCGGTGCCGACCAGGTGATGGATCTCGGCCGCGACGCCACCCGCCGCTTCTTCAAC
TGGTTTTACTGGAGCATCAACCTGGGTGCTGTGCTGTCGCTGCTGGTGGTGGCGTTTATT
CAGCAGAACATCAGCTTCCTGCTGGGCTACAGCATCCCTGTGGGCTGTGTGGGCCTGGCA
TTTTTCATCTTCCTCTTTGCCACCCCCGTCTTCATCACCAAGCCCCCGATGGGCAGCCAA
GTGTCCTCTATGCTTAAGCTCGCTCTCCAAAACTGCTGCCCCCAGCTGTGGCAACGACAC
TCGGCCGACCGTCAATGTGCCCGCGTGCTGGCCGACGAGAGGTCTCCCCAGCCAGGGGCT
TCCCCGCAAGAGGACATCGCCAACTTCCAGGTGCTGGTGAAGATCTTGCCCGTCATGGTG
ACCCTGGTGCCCTACTGGATGGTCTACTTCCAGATGCAGTCCACCTATGTCCTGCAGGGT
CTTCACCTCCACATCCCAAACATTTTCCCAGCCAACCCGGCCAACATCTCTGTGGCCCTG
AGAGCCCAGGGCAGCAGCTACACGATCCCGGAAGCCTGGCTCCTCCTGGCCAATGTTGTG
GTGGTGCTGATTCTGGTCCCTCTGAAGGACCGCTTGATCGACCCTTTACTGCTGCGGTGC
AAGCTGCTTCCCTCTGCTCTGCAGAAGATGGCGCTGGGGATGTTCTTTGGTTTTACCTCC

GTCATTGTGGCAGTCCTGGAGATGGAGCGCTTACACTACATCCACCACAACGAGACCGTG
TCCCAGCAGATTGGGGAGGTCCTGTACAACGCGGCACCACTGTCCATCTGGTGGCAGATC
CCTCAGTACCTGCTCATTGGGATCAGTGAGATCTTTGCCAGCATCCCACTGGAGTTTGCC
TACTCAGAGGCCCCGCGCTCCATGCAGGGCGCCATCATGGGCATCTTCTTCTGCCTGTCG
GGGGTGGGCTCACTGTTGGGCTCCAGCCTAGTGGCACTGCTGTCCTTGCCCGGGGGCTGG
CTGCACTGCCCCAAGGACTTTGGGAACATCAACAATTGCCGGATGGACCTCTACTTCTTC
CTGCTGGCTGGCATTCAGGCCGTCACGGCTCTCCTATTTGTCTGGATCGCTGGACGCTAT
GAGAGGGCGTCCCAGGGCCCAGCCTCCCACAGCCGTTTCAGCAGGGACAGGGGC

Figure 1: Deduced cDNA sequence of *hPHT1*. (A) Schematics of EST assembly. To view all ESTs aligned with the cDNA contig, click on the representative EST in the schematics. Possible sequence variations in the alignments are indicated by an asterix. + indicates the location of an uncertain nucleotide. (B) Schematics of *hPHT1* splice variants deduced from EST alignments and identified by RT-PCR (see second band in Figure 5A with lower molecular weight). Click on the indicated variants to view alignments with the contig sequence shown in 1A. Links to an HTML format of deduced cDNA sequences are also provided for each putative variant.

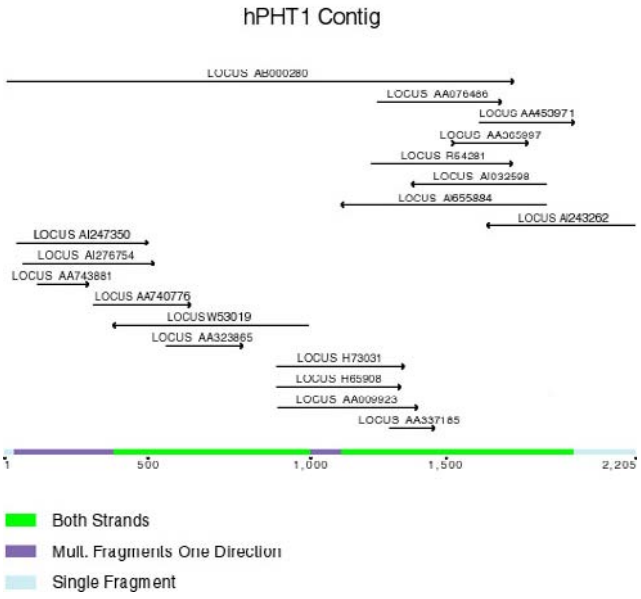


Figure 1a. Deduced cDNA sequence of hPHT1. (A) Schematics of EST assembly.

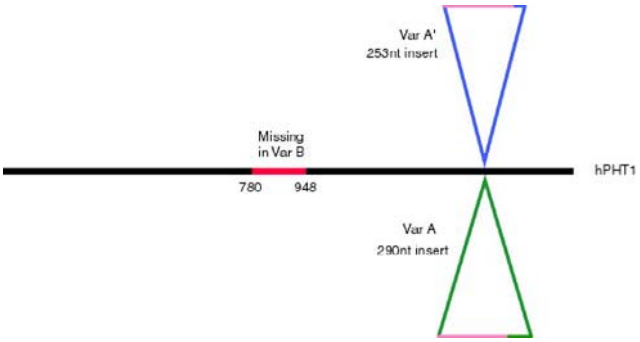


Figure 1b. Deduced cDNA sequence of hPHT1. (A) Schematics of EST assembly. To view all ESTs aligned with the cDNA contig, click on the representative EST in the schematics. Possible sequence variations in the alignments are indicated by an asterix. + indicates the location of an uncertain nucleotide. (B) Schematics of hPHT1 splice variants deduced from EST alignments and identified by RT-PCR (see second band in Figure 5A with lower molecular weight).

Putative Splice Variants of hPHT1

Several ESTs appear to span a sequence insert, suggesting the presence of possible splice variants shown in Figure 1B. Several variant ESTs assemble into a contig sequence containing an additional coding region of ~100 amino acid residues in the predicted extracellular loop between TMDs 11 and 12 (Variant A). It will be interesting to see the functional consequences of this insertion. Yet 3 additional ESTs -; also representing *hPHT1* -; assembled into a contig sequence suggestive of a variant form with a somewhat smaller insertion in the same location Figure 1B. Lastly, RT-PCR analysis of several tissues resulted in two bands, both found to be related to *hPHT1* upon sequence analysis. The faster migrating band contained a gap of 169 bps in

the middle of *hPHT1* Figure 1B, Variant B; for RT-PCR results see Figure 5A). Some of the putative *hPHT1* splice variants may introduce a frame shift and would not be expected to result in a functionally active transporter. These variants need to be cloned individually and tested experimentally. The information presented here is therefore important for guiding cloning efforts to produce functional hPHT1 protein.

Genomic hPHT2 Sequence.

The human EST contig sequence corresponding to rat cAMP-inducible 1 protein served to scan the human nr nucleotide databases. This revealed a PAC clone (P1-derived artificial chromosome) containing human genomic sequences closely related to the mouse cDNA-encoding cAMP-inducible 1 protein. Using the cDNA sequence of cAMP-inducible 1 protein, we were able to identify the likely introns and exons representing the presumed *hPHT2* gene Figure 2; click on the introns and exons to display DNA sequence). Each of the intron-exon boundaries are flanked by GT. . . AG in the intron sequence. Thus, the intron structure follows the GT-AG rule, where GT is called the splice donor and AT is called the splice acceptor.

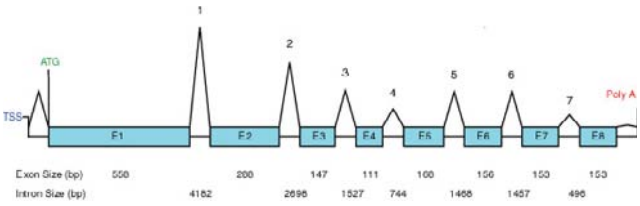


Figure 2. Predicted hPHT2 gene structure (A). Click on the introns and exons to view genomic sequence

The deduced cDNA coding sequence and protein sequence are shown in Table 4. Table 5 lists the identities and similarities among the protein sequences of the main mammalian members of the POT family. While hPepT1 and hPEPT2 represent one branch of this family, hPHT1, hPHT2, rPHT, and mouse cAMP-inducible 1 protein are closely related and form a second branch. hPHT1 has 89% identity to rPHT1, while hPHT2 is 81% identical to mouse cAMP-inducible 1 protein. Multiple sequence alignments are provided in Figure 3, including either the PHT branch only (3A), or both branches (3B).

Table 5. Identities and similarities among six POT protein sequences*

	hPepT1	**hPepT2**	**hPHT1**	**hPHT2**	**rPHT1**	**cAMP-inducible 1 protein**
hPepT1		51.34	20.39	25.96	22.84	24.05
hPepT2	51.34		22.54	22.88	24.17	24.52
hPHT1	20.39	22.54		53.90	89.27	51.96
hPHT2	25.96	23.41	53.90		54.00	80.94
rPHT1	22.84	24.17	89.27	54.00		51.81
cAMP-inducible 1 protein	24.05	24.52	51.96	80.94	51.81	

**POT indicates proton-dependent oligopeptide transporters.*

Hydropathy analysis corroborates the close similarity within the PHT and PepT1 branches of the POT family Figure 4A. These hydropathy profiles predict the presence of 11-12 transmembrane domains, as reported earlier. Topological predictions are shown in Figure 4B for hPHT1 and hPHT2.

After completion of this work, a cDNA sequence nearly identical to *hPHT2* was deposited by K. Ishiabshi and M. Imai: AB020598, Homo sapiens mRNA for peptide transporter 3, complete cds, 2113 bps in length. This defines the 3' and 5' ends of *hPHT1* as having a coding sequence of 1740 bps. The

deduced putative cDNA coding sequence is the identical length of the coding sequence suggested by our genomic *hPHT2* sequence, and it is identical to *hPHT2* over a large portion of the presumed coding region. However, there are also several remarkable differences. First, a fragment of 50 bps in the *hPHT2* coding region from position 61-110 is replaced in the cloned cDNA by a 50-bp fragment of low complexity (cg-rich), which is excluded from BLAST analysis by a low complexity filter. This 50-bp cDNA fragment did not recognize any sequence in the PAC clone containing the *hPHT2* genomic sequence, but it did recognize fragments in a number of unrelated genes, therefore, possibly representing a low complexity repeat fragment. The remainder of the cDNA sequence was identical to that of *hPHT2*, except for an insertion of three nucleotides each in three different locations of *hPHT2* (at positions 837, 1271, and 1428 of *hPHT2*). These sequence variations would indicate the presence of three additional amino acids at these respective positions, without disturbing the overall reading frame. It remains to be seen how these changes from our deduced coding sequence came about and whether they are of functional significance. In any case, comparing the cDNA and genomic sequences reveals many details of the possible protein structure not available otherwise.

After completion of this work, a cDNA sequence nearly identical to *hPHT2* was deposited by K. Ishiabshi and M. Imai: AB020598, Homo sapiens mRNA for peptide transporter 3, complete cds, 2113 bps in length. This defines the 3' and 5' ends of *hPHT1* as having a coding sequence of 1740 bps. The deduced putative cDNA coding sequence is the identical length of the coding sequence suggested by our genomic *hPHT2* sequence, and it is identical to *hPHT2* over a large portion of the presumed coding region. However, there are also several remarkable differences. First, a fragment of 50 bps in the *hPHT2* coding region from position 61-110 is replaced in the cloned cDNA by a 50-bp fragment of low complexity (cg-rich), which is excluded from BLAST analysis by a low complexity filter. This 50-bp cDNA fragment did not recognize any sequence in the PAC clone containing the *hPHT2* genomic sequence, but it did recognize fragments in a number of unrelated genes, therefore, possibly representing a low complexity repeat fragment. The remainder of the cDNA sequence was identical to that of *hPHT2*, except for an insertion of three nucleotides each in three different locations of *hPHT2* (at positions 837, 1271, and 1428 of *hPHT2*). These sequence variations would indicate the presence of three additional amino acids at these respective positions, without disturbing the overall reading frame. It remains to be seen how these changes from our deduced coding sequence came about and whether they are of functional significance. In any case, comparing the cDNA and genomic sequences reveals many details of the possible protein structure not available otherwise.

RT-PCR and Northern Blot Analysis of hPHT1 and hPHT2 mRNA From Human Tissues

The presence of numerous ESTs from many tissues representing the presumed *hPHT1* in the databases suggested that *hPHT1* might be expressed in many human tissues. RT-PCR analysis had revealed detectable expression of *hPHT1* in each of three tissues tested: intestines, skeletal muscle, and pancreas (for intestines, see Figure 5A). In each of these tissues, the presence of a shorter band suggested a possible splice isoform. The experimentally determined sequence Figure 1B shows a deletion of 169 bases.

For Northern analysis we used two ESTs representing *hPHT1* and *2*, which were labeled to detect the presence of mRNA in 12 human tissues Figure 5B. *hPHT1* was mainly expressed in skeletal muscle, followed by kidney, heart, and liver, with relatively little expression in colon and brain. mRNA bands were detected at apparent molecular weight 2.8 kb and 5.1 kb, indicating the presence of possible mRNA variants.

The mRNA tissue distribution of *hPHT2* differed significantly from that of *hPHT1* Figure 5B. A single major band appeared at 2.4 kb, with highest expression in spleen, placenta, lung, and leukocytes, followed by heart, kidney, and liver.

Figure 3. Multiple protein sequence alignments. PHT branch only (A), plus PepT branch (B). Blue: 50% or higher identity; red, green: conservative and similar substitutions.

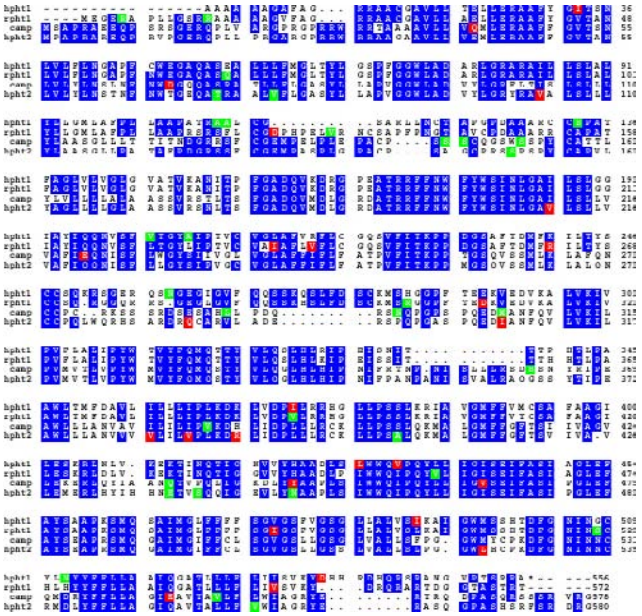


Figure 3a. PHT branch.

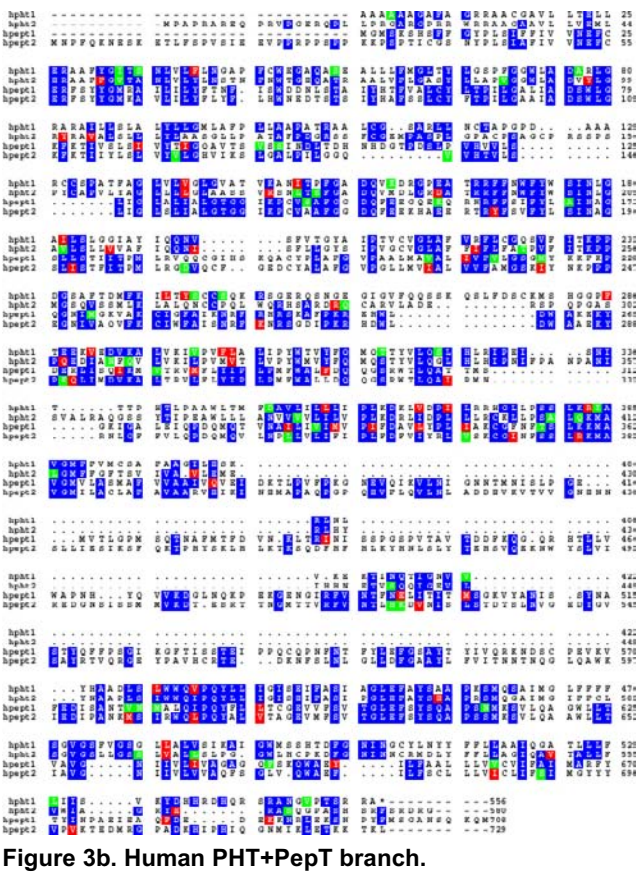


Figure 3b. Human PHT+PepT branch.

Figure 4: Hydropathy plots for hPepT1, hPepT2, rPHT, cAMP-inducible 1 protein, hPHT1, and hPHT2 (A), and predicted topological representation (B) of hPHT1 and hPHT2

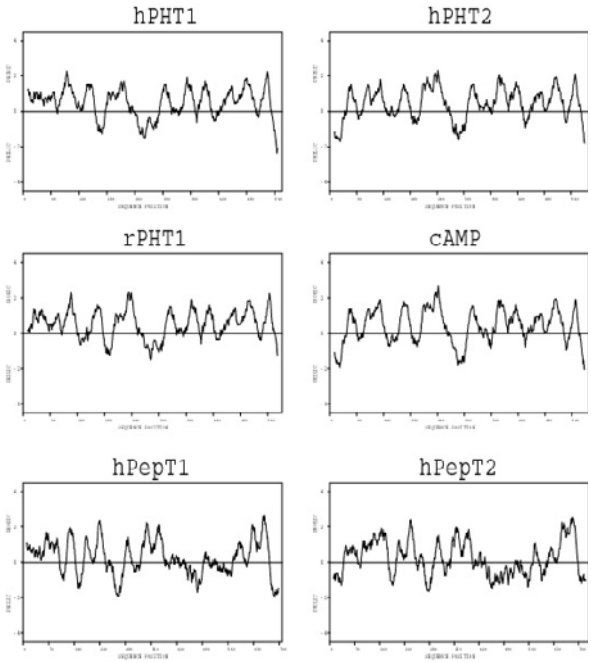


Figure 4a. Kyte Doolittle method, window size 7.

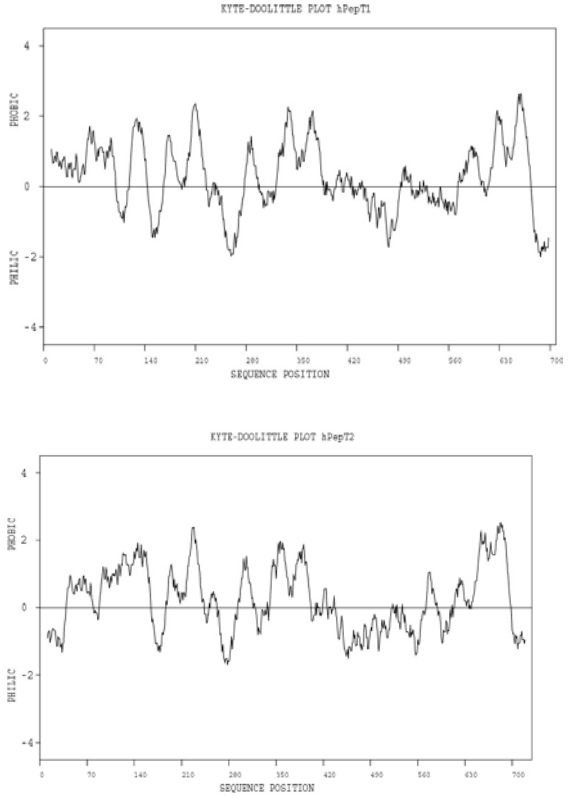


Figure 4b. Topo representations

Figure 5: (A) RT-PCR analysis of *hPHT1* extracted from human small intestinal tissue specimen. The control lane contained a control mRNA with unrelated primers. The main band in the mRNA lane is 509 bps long (PCR-1), and its determined sequence corresponds to *hPHT1*, whereas the faster migrating band is ~340 bp long, representing a possible splice variant of *hPHT1* (see Figure 1B for experimentally determined sequence). This latter band is also seen in skeletal muscle and panceas. (B) Northern blot analysis of *hPHT1* and *hPHT2* expression, using respective EST clones as the probe. An actin probe served as the control. Human mRNA was extracted from the indicated tissues

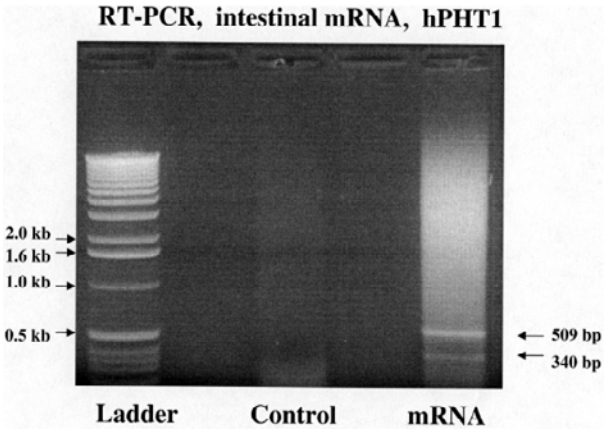
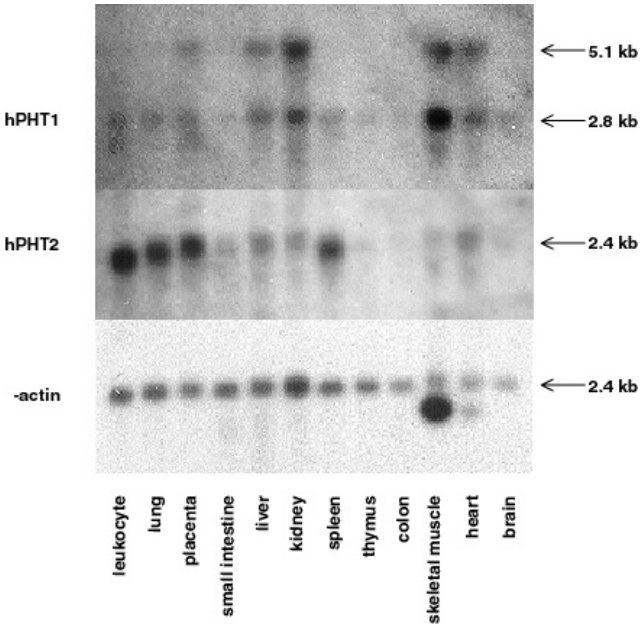


Figure 5a. RT-PCR analysis of hPHT1 extracted from human small intestine tissue sample.



5b. Northern blot analysis of hPHT1, hPHT2, and actin expression in human tissues.

DISCUSSION

This study identifies several new putative members of the POT gene family, using a bioinformatics analysis of available sequence databases, including the human EST database. Because the POT family seemingly stands separate from other transporter gene families, searching for POT-related ESTs is facilitated. With an increasing number of sequences present in the publicly accessible databases, however, we now have begun to find transporters outside the POT gene family with alignment scores that support a finding of possible homology. The INCA search, or any other exhaustive search tool, should be performed periodically to finds the missing links that can document POT's probable homology to other gene families. This will greatly facilitate the study of structure and function of these transporters because one would expect molecular architecture and functional domains to recur in homologous proteins. In this study, we have used gapped BLAST analyses (19), which permits inclusion of gaps in local alignments. This carries the possible disadvantage that the permutation matrix used to calculate similarity scores may be inappropriate, in particular for the TMD sections. The use of psi BLAST (19) could overcome this problem by generating a position-specific matrix that would account for mutational drift in TMDs and loops separately. However, our iterative INCA approach could result in overmodeling of the matrix, and thus, possible inclusion of unrelated sequences. Moreover, the primary goal here was to identify new human genes closely related to the known POT family members, rather than probing the most distant relationships.

By far, the largest number of human ESTs aligned best with rPHT1, a peptide-histidine transporter mainly expressed in rat brain. Assembly of numerous ESTs into a contig sequence termed *hPHT1* permits a number of observations on the putative coding sequence. More than 95% of the coding sequence expected from comparison to rPHT1 can be derived from the assembled ESTs. In several regions of the deduced *hPHT1* sequence, multiple ESTs overlap, thereby providing a first glimpse of possible sequence variations in the human population. While EST sequencing is not rigorously quality controlled

and single nucleotide variants occur only sporadically, multiple overlapping ESTs could nevertheless assist in finding single nucleotide polymorphisms (SNPs). There are several such candidate SNPs in the EST alignments: future work will determine whether these do, indeed, represent human sequence variations. However, the presence of possible splice variants of *hPHT1* was clearly indicated by an insert or gap in two regions of the *hPHT1* coding region. These were detected either by EST alignemnts or experimentally by RT-PCR Figure 1B and Figure 5A. It will be of interest to determine the tissue expression and function of the splice variants. The insert into the loop between TMDs 11 and 12 introduces an additional 100 residue-sequence fragment into hPHT1. This region contains no homology to any other known protein.

A possible splice variant has previously been detected for hPepT1. Inue and colleagues (25) have cloned a cDNA from human duodenum with 1704 bp, encoding a predicted protein of only 208 residues (hPepT1-RF). Of these, residues 18-195 are identical to an equivalent region of hPepT1. This truncated hPepT1 protein appears to lack ability to transport peptides; however, cotransfection of *hPepT1-RF* with *hPepT1* affected the pH sensitivity of peptide transport by hPepT1, suggesting a regulatory function for hPepT1-RF of yet unknown mechanism. The functions of splice variants of *hPHT1* remain to be determined.

Assembly of ESTs into a second contig sequence revealed high identity to the mouse cAMP-inducible 1 protein, which was isolated from lymphocytes as one of the upregulated mRNAs after stimulating with cAMP (26). Close similarity to PHT1 suggests that this gene also encodes a peptide transporter. Using this human contig sequence, we have identified a full-length genomic sequence in a PAC clone from which introns and exons can be predicted for the presumed *hPHT2* gene. High identity to the entire mouse cAMP-inducible 1 protein suggests that hPHT2 is the human orthologue, closely related to the PHT-like branch of the POT family. This will facilitate the cloning and testing of this putative new member of the human POT family.

Our comprehensive INCA search uncovered several additional ESTs with sequences similar to members of the POT family. Using BLAST analyses with these ESTs as the query suggested the possibility of additional human POT genes. Further work is needed to complete the human POT family. Also, we cannot preclude that the proposed *hPHT 1* and *2* genes, although highly similar to genes encoding rPHT1 and cyclic AMP-inducible 1 protein, are not the immediate orthologues, but that there are as yet additional closely related human genes not represented in the EST database.

The strong representation of *hPHT1* in the EST database suggests that the presumed hPHT1 is widely expressed in human tissues, largely in the CNS, in contrast to its restricted expression in rats. One of these ESTs stems from a human colon carcinoma, an indication that hPHT1 may also be expressed in human intestines. We have corroborated this supposition with the use of RT-PCR; however, Northern blot analysis has revealed that *hPHT1* and *hPHT2* are not highly expressed in intestines relative to other tissues. Protein expression and functional studies are required to determine whether these transporters, in addition to hPepT1, could play a role in intestinal peptoid drug absorption. This could be of considerable pharmacological interest, as previous studies have suggested that PepT1 was the sole peptide transporter in rodent intestines, whereas no definitive experimental evidence has been reported on the responsible transporter subtype in humans. Bioavailability studies demonstrate that peptoid drugs, such as certain antibiotics, largely depend on intestinal absorption by peptide transporters to enter the systemic circulation. For example, coadministration of a dipeptide reduced the oral bioavailability of amoxicillin by 80% in human subjects providing strong evidence that a saturable dipeptide transport process is involved (27). Whether hPHT1 or 2 could play a role in oral antibiotic bioavailability remains to be seen.

Our Northern blot analysis revealed strong expression for *hPHT1* in skeletal muscle and kidney while *hPHT2* was highly expressed in leukocytes, lung placenta, and spleen. Detectable expression of both genes in organs, such as the heart, may be of

interest in understanding the efficacy of antibiotic treatment of localized infections -; particularly if the infectious agent resides intracellularly. The tissue distribution of gene expression differs from that of *hPepT1* (mainly intestinal) and *hPepT2* (mainly renal) which underscores the relevance of our findings to targeting therapy to specific organs. It is curious to note that neither hPHT1 nor hPHT2 are strongly expressed in the brain, while rPHT1 is highly expressed in the CNS. It will be of interest to establish the role of any peptide transporters expressed in human brain.

Overall, our results indicate that the human POT family contains at least 4 genes encoding possible peptide transporters. Each displays a distinct pattern of tissue expression, providing a possible avenue for drug targeting to select tissues. Tissue distribution of POT gene expression is of particular interest for achieving oral bioavailability or targeting drugs to tumor tissues. For example, Tsuji et al have identified a fibrosarcoma cell line expressing an unusual H^+/dipeptide transporter activity (17) . We have determined that hPepT1 is highly expressed in pancreatic (28) and colon adenocarcinomas, including liver metastases (M.Y. Covitz and W.S., unpublished data), considerably above the level seen in surrounding normal tissues. Moreover, each member of the peptide transporter family is likely to exhibit distinct selectivities for peptides and peptoid drugs. Thus, amino acid esters of 5'nucleosides (eg, valcyclovir) (15) are recognized by hPepT1, even though the structure of these prodrugs is quite distinct from dipeptides. However, affinity of these nucleoside prodrugs for other peptide transporters remains to be determined. One could envisage a large variety of amino acid-nucleoside prodrugs of antivirals or anticancer agents to enhance oral bioavailability or target the drug to tumor tissues, as a function of which peptide transporter is expressed in the target tissue. Phenotypic characterization of tumor tissues as to which transporters are expressed and to what extent will become an important question that extends beyond the peptide transporter family, and indeed should be applied to all known drug transporters.

ACKNOWLEDGMENT

C.U.N. was supported by the Center for Drug Design and Transport, grant from Danish Medical Research Council.

REFERENCES

1. Yang CY, Dantzig AH, Pidgeon C. Intestinal peptide transport systems and oral drug availability. Pharm Res. 1999;16:1331-1343.
2. Dantzig AH, Bergin L. Uptake of the cephalosporin, cephalexin, by a dipeptide transport carrier in the human intestinal cell line, Caco-2. Biochim Biophys Acta 1990;1027:211-217.
3. Matsumoto S, Saito H, Inui K. Transcellular transport of oral cephalosporins in human intestinal epithelial cells, Caco-2: interaction with dipeptide transport systems in apical and basolateral membranes. J Pharmacol Exp Ther. 1994;270:498-504.
4. Wenzel U, Gebert I, Weintraut H, Weber WM, Clauss W, Daniel H. Transport characteristics of differently charged cephalosporin antibiotics in oocytes expressing the cloned intestinal peptide transporter PepT1 and in human intestinal Caco-2 cells. J Pharmacol Exp Ther. 1996;277:831-839.
5. Kramer W, Girbig F, Gutjahr U, Kleemann HW, Leipe I, Urbach H, Wagner A. Interaction of renin inhibitors with the intestinal uptake system for oligopeptides and beta-lactam antibiotics. Biochim Biophys Acta 1990;1027:25-30.
6. Fei YJ, Kanai Y, Nussberger S, Ganapathy V, Leibach FH, Romero MF, Singh SK, Boron WF, Hediger MA. Expression cloning of a mammalian proton-coupled oligopeptide transporter. Nature 1994;368:563-566.
7. Thwaites DT, Brown CD, Hirst BH, Simmons NL. Transepithelial glycylsarcosine transport in intestinal Caco-2 cells mediated by expression of H^+-coupled carriers at both apical and basal membranes. J Biol Chem. 1993;268:7640-7642.
8. Graul RC, Sadée W. Sequence Alignments of the H^+-Dependent Oligopeptide Transporter Family. Inferences on Structure and Function of the Intestinal PET1 Transporter. Pharm Res. 1997;14:388-400.
9. Fei YJ, Ganapathy V, Leibach FH. Molecular and structural features of the proton-coupled oligopeptide transporter superfamily. Prog Nucleic Acid Res Mol Biol. 1998;58:239-261.
10. Paulsen IT, Skurray RA. The POT family of transport proteins. Trends Biochem Sci. 1994;19:404.
11. Steiner HY, Naider F, Becker JM. The PTR family: A new group of peptide transporters. Mol Microbiol. 1995;16:825-834.
12. Covitz KM, Amidon GL, Sadée W. Membrane topology of the human dipeptide transporter, hPEPT1, determined by epitope insertions. Biochemistry 1998;37:15214-15221.
13. Liu W, Liang R, Ramamoorthy S, Fei YJ, Ganapathy ME, Hediger MA, Ganapathy V, Leibach FH. Molecular cloning of PEPT 2, a new member of the H^+/peptide cotransporter

family, from human kidney. Biochim Biophys Acta. 1995;1235:461-466.

14. Liang R, Fei YJ, Prasad PD, Ramamoorthy S, Han H, Yang-Feng TL, Hediger MA, Ganapathy V, Leibach FH. Human intestinal H^+/peptide cotransporter: Cloning, functional expression, and chromosomal localization. J Biol Chem. 1995;270:6456-6463.
15. Han H, de Frueh RLA, Rie JJ, Covitz KM, Smith PL, Lee C-P, Sadée W, Amidon GL. 5'-Amino acid esters of antiviral nucleosides, acyclovir and AZT, are absorbed by the intestinal hPEPT1 peptide transporter. Pharm Res. 1998;15:1154-1159.
16. Yamashita T, Shimada S, Guo W, Sato K, Kohmura E, Hayakawa T, Takagi T, Tohyama M. Cloning and functional expression of a brain peptide/histidine transporter. J Biol Chem. 1997;272:10205-10211.
17. Nakanishi T, Tamai I, Sai Y, Sasaki T, Tsuji A. Carrier-mediated transport of oligopeptides in the human fibrosarcoma cell line HT1080. Cancer Res. 1997;57:4118-4122.
18. Graul RC, Sadée W. Evolutionary relationships among proteins probed by an iterative neighborhood cluster analysis (INCA): Alignment of bacteriorhodopsins with the yeast sequence YRO2. Pharm Res. 1997;14:1533-1541.
19. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389-3402.
20. Huang X, Madan A. CAP3: A DNA sequence assembly program. Genome Res. 1999;9:868-877.
21. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of proteins. J Mol Biol. 1982;157:105-132.
22. Smith TF, Waterman MS. Identification of common molecular subsequences. J Mol Biol. 1981;147:195-197.
23. Pearson WR. Searching protein sequence libraries: Comparison of the sensitivity and selectivity of the Smith-Waterman algorithms. Genomics 1991;11:635-650.
24. Sadée W, Graul RC, Lee AY. Classification of membrane transporters. In: Amidon GL and Sadée W, eds. *Membrane Transporters as Drug Targets*. New York: Plenum Press, 1999. pp 29-58.
25. Saito H, Motohashi H, Mukai M, Inui K. Cloning and characterization of a pH-sensing regulatory factor that modulates transport activity of the human H^+/peptide cotransporter, PEPT1. Biochem Biophys Res Commun. 1997;237:577-582.
26. Miayata M, Takahashi Y, Zheng P, Smith HD. Identification of three genes markedly indiced by cAMP treatment of RAW264 cells. *Biochem Genet Metab*. submitted.
27. Chulavatnatol S, Charles B. Antagonism of amoxycillin absorption in man following oral administration with glycyl peptides. Europ J Pharmacol. 1994;40:374-378.
28. Gonzales DE, Covitz KM, Sadée W, Mrsny RJ. An oligopeptide transporter is expressed at high levels in the pancreatic carcinoma cell lines AsPc-1 and Capan-2. Cancer Res. 1998;58:519-525.