

Human Membrane Transporter Database: A Web-Accessible Relational Database for Drug Transport Studies and Pharmacogenomics

Submitted June 21, 2000; accepted June 27, 2000; published July 14, 2000

Qing Yan, Wolfgang Sadée

Program in Medical Information Science, University of California, San Francisco, CA 94143-0446

ABSTRACT The human genome contains numerous genes that encode membrane transporters and related proteins. For drug discovery, development, and targeting, one needs to know which transporters play a role in drug disposition and effects. Moreover, genetic polymorphisms in human membrane transporters may contribute to interindividual differences in the response to drugs. Pharmacogenetics, and, on a genome-wide basis, pharmacogenomics, address the effect of genetic variants on an individual's response to drugs and xenobiotics. However, our knowledge of the relevant transporters is limited at present. To facilitate the study of drug transporters on a broad scale, including the use of microarray technology, we have constructed a human membrane transporter database (HMTD). Even though it is still largely incomplete, the database contains information on more than 250 human membrane transporters, such as sequence, gene family, structure, function, substrate, tissue distribution, and genetic disorders associated with transporter polymorphisms. Readers are invited to submit additional data. Implemented as a relational database, HMTD supports complex biological queries. Accessible through a Web browser user interface via Common Gateway Interface (CGI) and Java Database Connection (JDBC) (<http://128.218.208.23/transporter/trans.html>), HMTD also provides useful links and references, allowing interactive searching and downloading of data. Taking advantage of the features of an electronic journal, this paper serves as an interactive tutorial for using the database, which we expect to develop into a research tool.

Corresponding author: Qing Yan, 513 Parnassus Ave., S-926, Box 0446, University of California, San Francisco, CA 94143-0446; qyan@itsa.ucsf.edu

INTRODUCTION

Membrane transporters play direct roles in the absorption, distribution, and elimination of nutrients, ions, and many drugs (1). The term "transporter" as used here includes a variety of membrane proteins with diverse functions and structures, such as ion exchangers and certain non-transporter ATP-binding cassette membrane proteins. The importance of human membrane transporters is shown in genetic disorders caused by transporter malfunction, such as glucose malabsorption (2,3) and insulin-resistant glucose transport (4). Multiple drug resistance (MDR) genes encoding transporter molecules are implicated in native and acquired resistance to antineoplastic agents (5-7).

While the role of membrane transporters in drug effects has generated much recent interest, the relevant transporters remain unknown for most drugs. Even if a transporter is known to interact with a drug, it is uncertain as to whether as yet unknown transporters also recognize the same drug. Extrapolating from the fully sequenced genomes of an increasing number of organisms, one would expect the number of human transporters and related proteins to count into the thousands. A primary goal of current research in drug discovery and development is therefore to understand which transporters recognize a drug candidate and which transporters can be exploited to target the drug to the intended tissue in the body.

Pharmacogenetics and pharmacogenomics address the role of genetic variants in 1 or more genes in individual response to drugs and xenobiotics (8). Interindividual differences in therapeutic efficacy or adverse/toxic drug reactions (9) may be attributable to hereditary variation in proteins that affect the absorption, distribution, metabolism, and excretion of

compounds, their interaction with the receptor/protein target, and finally downstream effects. Because drug effects depend on multiple proteins in the body, a search for the responsible genes must be genome-wide, going beyond the candidate-gene approach of earlier pharmacogenetic studies. Thus, pharmacogenomics is an emerging discipline that may open new avenues for improving drug therapy. Specifically, polymorphisms in transporters may affect drug absorption, distribution, elimination, and tissue targeting. Polymorphisms are allelic variants in genes that exist stably in the population, conventionally with an allele frequency above 1%. This definition of polymorphisms is somewhat flawed because allelic frequencies differ among distinct ethnic groups. Therefore, we prefer the term "sequence variant." Such variants may alter the activity of the encoded protein relative to the wild-type sequence. Moreover, mutations can result in severe genetic disorders, many of which are included with the database presented here. Most of these mutations are rare (considerably below 1%), but nevertheless of considerable medical importance.

At this point, very little information is available as to whether and in which cases transporter polymorphisms are a main cause of inter-individual differences in drug response. P-glycoprotein (MDR) is considered to act as an active efflux pump relevant to drug absorption and elimination, access to the central nervous system, and resistance of tumor cells to chemotherapeutic agents (5-7,10). However, this activity is largely associated with the level of transporter protein expression in tissues rather than any documented sequence variants. Protein expression depends on numerous possible factors and events, including variants in the promoter and enhancer regions, sequence variants that interfere with protein stability and cellular trafficking, somatic changes (eg, in tumors by gene amplification), and induction of transcription. This type of information needs to be organized such that it is readily accessible for the user. In the year 2000, we expect completion of sequencing the entire human genome, with full annotation following shortly thereafter. Once this is achieved, we can catalogue all transporter genes and complete the human transporter database, including tissue distribution and functions of each gene, and

their sequence variants. This large endeavor requires new technologies to implement it fully.

The systematic application of microarray gene technology to the analysis of membrane transporters may help determine tissue expression and allelic distributions on an individual basis (11). Such technology may also facilitate the identification of putative defective alleles and the exclusion of affected patients from therapy with a given drug. However, it is difficult to determine which transporters are relevant, not just because of the multitude of genes, but also because of their tremendous structural and functional heterogeneity. With the amount of available data rapidly increasing and the pressing demand for microarray gene technology, the need for a database becomes increasingly urgent. The construction of a human membrane transporter database (HMTD) is an effort to meet such needs.

FEATURES OF HMTD

The HMTD currently contains data on more than 250 human membrane transporters and related proteins, their structure, function, sequence variants, and substrates, especially drugs. Structural information provided in the database includes gene and protein size, sequence, and exon number. General characteristics of human membrane transporters are included, ie, chromosome location, membrane topology such as transmembrane domain, gene family, and tissue distribution. Because our database is intended to support pharmacogenomic studies, it provides information on sequence variants, altered functions caused by polymorphisms/mutations, and the (patho)physiological role and associated disease. HMTD also supplies relevant links to external databases and references. The database collects and integrates different data types such as text (eg, gene family), number (eg, protein size), and hyperlink (eg, reference).

HMTD is implemented as a relational database, which facilitates dynamic queries and extraction of information. Several Web sites contain information on transporters, such as the page maintained by Michael Müller in the Netherlands (12). However in

nonrelational systems, it is difficult to find data beyond the logical structure that was originally defined. For example, to find a piece of information that is derived from raw data, the user has to go over all the forms or articles. This makes it difficult to obtain direct answers to questions such as "Which transporters are expressed in the kidney?" It is also time consuming to update such systems. One of the distinct features of HMTD is that it is maintained as a relational database that overcomes these obstacles. Such a database provides a powerful way to organize and maintain information. In our relational database, data are structuralized and standardized to enable further application of decision support such as ad hoc queries and data mining. Adjustments in the relational database can be easily made, and all the data elements can be queried flexibly (the details of querying the database will be described below). In addition, views of the data do not have to be predefined, and relationships can be built on a moment's notice.

Another important feature of HMTD is that the database is published on the Web and is accessible through the Internet. Such an application allows remote public access, interactive searching, and easy downloading of data. The Web offers standardization of the user interface and makes the interface user-friendly. Users can make queries without knowing any computer languages or details about the database. The Web also provides high transaction volume, the ability to work across platforms and applications, and a common way to display multimedia data. Users can search for information in the database whether they are using PC or MAC format, and more than 1 person can make queries at the same time.

Terminology regarding genes often causes confusion. If you want to look for information on neutral amino acid transporters, should you type "SATT" or "ASCT1"? If you search these 2 gene symbols from the nomenclature database (13), they are also termed "SLC1A4." In our database, we use the "Nomenclature of Mammalian Transporter Genes" as our reference for gene names (14). To facilitate use,

our database can be searched by its symbol(s) as transporter names. For example, to search for information about neutral amino acid transporters, you can type in "SATT," "ASCT1," or "SLC1A4" as the key word.

CONTENTS AND STRUCTURE OF HMTD

One of the construction goals of HMTD is to provide an overview of human membrane transporters. Therefore the database tries to contain transporter genes as exhaustively as possible and will be updated regularly. A list of transporters in the current database can also be found at <http://128.218.208.23/transporter/translist.html>.

Because of the fragmented current knowledge, one needs to understand that the current database is incomplete in many of its possible entries, and it soon will have to greatly expand with the sequencing of the human genome. By posting the database on relevant Web pages, we invite contributions from readers with relevant information to update the database on a continuous basis. Currently our database is updated monthly.

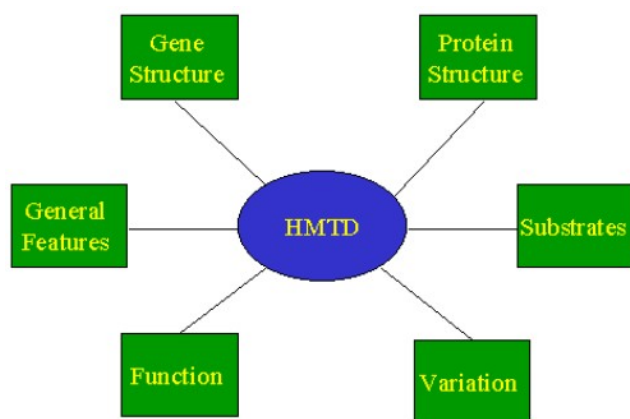
To ensure the accuracy of information in our database, our data sources include peer-reviewed papers and published sequences (such as from GenBank). HMTD collects relevant and useful links to other databases and Web sites. Some of the links provided in the database are listed in Table 1. Among these links, dbSNP is a database about single nucleotide polymorphisms (SNPs) recently established by National Center for Biotechnology Institute (NCBI, currently containing ~25,000 SNPs). SNPs are the most common variations that occur about once every 100 to 300 bases, which are expected to facilitate large-scale association genetics studies. Updating with useful links is another goal of our database development.

The information in HMTD is organized in multiple tables and implemented in a relational database management system. Such a system facilitates categorization of data and supports complex queries.

Table 1. Sample Database Hyperlinks in HMTD

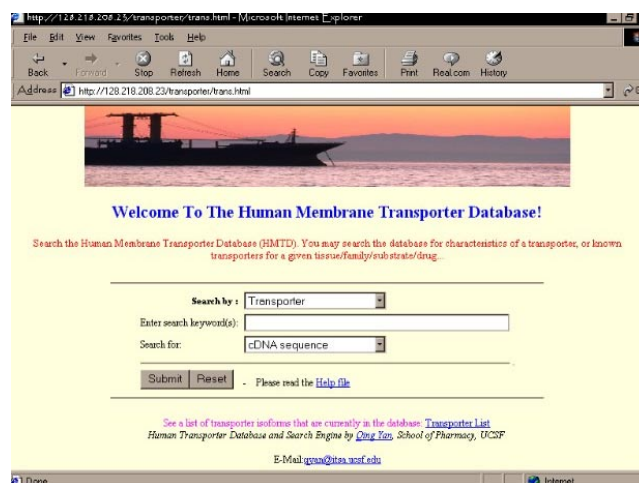
LINK	URL
Entrez	http://www.ncbi.nlm.nih.gov/Entrez/
OMIM (Online Medialian Inheritance in Man)	http://www.ncbi.nlm.nih.gov/Omim/
HGMD (Human Gene Mutation Database)	http://www.uwcm.ac.uk/uwcm/mg/hgmd0.html
GDB (Genome Database)	http://www.gdb.org/
PubMed	http://www.ncbi.nlm.nih.gov/PubMed/
Genes and Disease Map	http://www.ncbi.nlm.nih.gov/disease/Transporters.html
dbSNP (Database of Single Nucleotide Polymorphisms)	http://www.ncbi.nlm.nih.gov/SNP/index.html

The body of data is divided into 6 tables that provide information about gene and protein structure, general characteristics of the transporter, substrates of the transporter, function, and variation (Figure 1). For example, in the entity General Features, the attributes include family, topology, and tissue distribution. Homology of the transporter, pharmacokinetic information, and transport mechanism are also contained in the database (if incomplete at this stage). To standardize our data, we use Peng and Hediger's "Nomenclature of Mammalian Transporter Genes" as our reference for classification of gene family and transporter type (14).

**Figure 1. The structure of HMTD.**

PROGRAM ARCHITECTURE, USER INTERFACE, AND APPLICATION

HMTD is accessible through the Web. Its user interface is written in standard HTML that can be rendered in any Web browser, such as Netscape Navigator and Microsoft Internet Explorer (IE) (Figure 2). The Web database application is based on a 3-tiered client-server architecture. As shown in Figure 3, the client tier is a thin client HTML page on a Web browser (such as Netscape or IE). The server tier is a relational database (HMTD) management system. The middle tier consists of a Web server and a Common Gateway Interface (CGI) program written in Java. The connection from the CGI program to the database is through the JDBC (Java Database Connection) driver.

**Figure 2. Database Front End: the HTML page.**

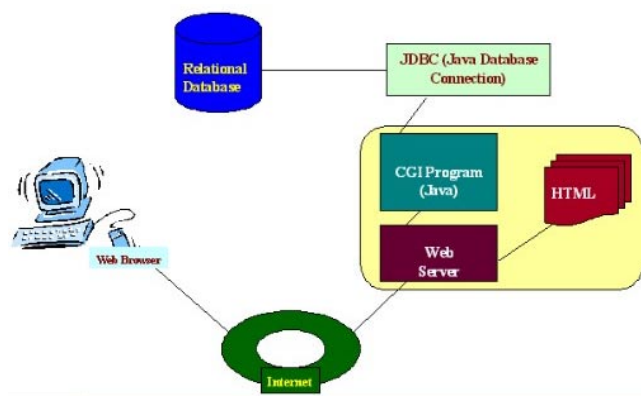


Figure 3. The architecture of HMTD: the three-tiered Web database application.

A database is the storage of a large amount of data to be retrieved as useful information. Such retrieval of information is accomplished through queries. Queries can be performed through filling out the form at the Web page (Figure 2, also <http://128.218.208.23/transporter/trans.html>).

Detailed instructions on using this form can be found at <http://128.218.208.23/transporter/helpfile.html>.

Some sample queries for retrieving information from HMTD are shown in Table 2.

Table 2. Sample Queries for HMTD

-
- » For a given transporter
 - What are the substrates?
 - What polymorphisms are known?
 - What are the gene and protein sequences?
 - What is the chromosomal location?
 - What are the associated diseases and disorders?
 - In which tissues can it be found?
 - » For a substrate/drug
 - Which transporters are known to interact with it?
 - » For a given tissue
 - Which transporters are expressed in it?
-

For example, a useful query might be “Which tissues express transporter ‘OATP’?” (see the answer in Figure 4). A reverse query is also commonly asked, such as “Which transporters are known to be expressed in a given tissue?” The answer to this question may help identify the possible main drug transporter candidates in a target tissue. Table 3 shows some sample results of the query from HMTD, given intestine, kidney, and brain as sample tissues. Gaining information from such queries allows us to make judgments on the basis of which *known* transporters are prevalent in available tissues.

Table 3. Some Sample Results of the Queries: Which transporters are expressed in the intestine (or kidney, or brain) tissue?

	INTESTINE	KIDNEY	BRAIN		
GLUT2	MDR1	GLUT2	KCC3	ENT2	SVMT
GLUT5	NCCT	GLUT5	CNT1	GLUT3	NCX1
SGLT1	MCT7	SGLT1	GLVR1	SGLT1	OATP1
SGLT2	FOLT	SGLT2	EAAT3	OATP	KCC3
NTCP2	NRAMP2	OATP	OCTN1	MCT2	ZNT-1
PEPT1	rBAT	NTCP2	OAT1	GAT-1	ZNT-3
SVCT2	NHE2	TAUT	OAT3	GAT-3	GLYT-2
CNT1	NHE3	LAT-3	MDR1	LAT-1	CNT1
CNT2	SDCT1	NCX1	NTCP1	HTT	CNT2
EAAT3	OCT1	PEPT1	AE1	TAUT	CAT-4
PGT	ATBo	SVCT2	AE2	PROT	MCT6

Such information may have relevance in pharmacogenomics studies. For example, the resultant information can be used in the design of new drugs that need to be targeted to certain tissues. To determine which drug transporters are candidates for genotyping, this information may also be helpful. In addition, such information may assist in guiding drug therapy in individual patients on the basis of the drug transporter genotype and phenotype.

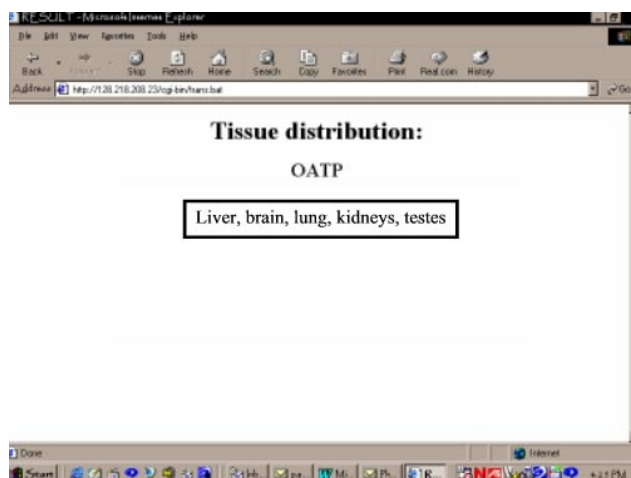


Figure 4. The sample result of the query “What tissue is OATP expressed in?”

To query the database, users fill out the form on the Web page by selecting the type of data they would like to query (such as “polymorphism” of a “transporter”); we can update the user interface if needed to expand the types of data to be searched), typing in the keyword (such as “TAP1”), and clicking the button “submit” (try it at <http://128.218.208.23/transporter/trans.html>). The Web browser sends the Web page requests (data requests) to the Web server. The Web server processes the page requests and passes the data requests to the CGI program. The CGI program analyzes the input data and constructs an SQL (structured query language) query. The query is then sent to the HMTD database through a JDBC interface. After the query is executed in HMTD, a result set is returned to the CGI program. Finally, the CGI program dynamically constructs the HTML page and sends it back to the Web server, which in turn passes it to the Web browser. The users then see the query results (Figure 5).

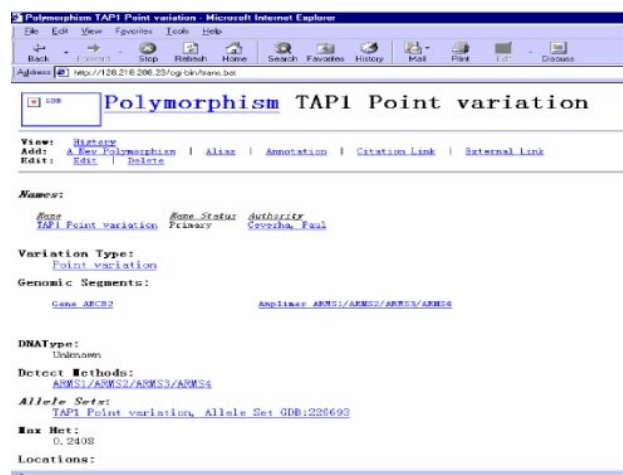


Figure 5. The sample result of the query “What’s known about polymorphisms of TAP1?”

SUMMARY AND FUTURE DEVELOPMENT

HMTD is a human membrane transporter database created to facilitate studies on human membrane transporters for drug discovery, development, and therapy, and for pharmacogenetics and pharmacogenomics. As a Web-accessible relational database, HMTD supports dynamic searches for structures, expression, functions, substrates, variants, disorders, and references of human membrane transporters. HMTD is an effort to organize the multitude of transporter genes, their tremendous structural heterogeneity, and various functions. The database can also be used as a prototype for pharmacogenomics research in other areas.

With the advancement of research about transporters and the increase of our knowledge, more data will be added to the database. More complicated queries will also be possible with the inclusion of chemical similarity algorithms, eg, “Which substrates of known transporters are most similar in chemical structures?” A query that is highly pertinent to drug discovery and development could be presented, eg, “Which transporters are most likely to interact with this new drug?”.

The progression of research may also enable us to include the secondary and tertiary structures (as more membrane transport proteins are crystallized) of membrane transporter genes in our database.

Structural data are needed to study the associations between structure and function. To facilitate researchers' submission of data for querying, and for viewing of the results in the context of published data, security access could be implemented in the database (currently researchers are invited to submit their data through e-mail to qyan@itsa.ucsf.edu).

Future enhancement of HMTD may also include the development of data-mining tools to make use of the database effectively. Such tools may help answer meaningful pharmacogenomics questions such as: "Given the numerous known disease mutations in transporters, can we establish a pattern as to which ones affect the response to what drugs? Are there any common motifs that are disrupted in these transporters with altered functions, to establish structure-function relationships?" In this fashion, databases will eventually transform into research tools.

REFERENCES

1. Sadée W, Druebbisch V, Amidon GL. Biology of membrane transporter proteins. *Pharm Res.* 1995;12:1823-1837.
2. Levin RJ. Digestion and absorption of carbohydrates -- from molecules and membranes to humans. *Am J Clin Nutr.* 1994;59(3 Suppl):690S-698S.
3. Martin MG, Lostao MP, Turk E, Lam J, Kreman M, Wright EM. Compound missense mutations in the sodium/D-glucose cotransporter result in trafficking defects. *Gastroenterology.* 1997;112:1206-1212.
4. Heart E, Choi WS, Sung CK. Glucosamine-induced insulin resistance in 3T3-L1 adipocytes. *Am J Physiol Endocrinol Metab.* 2000;278:E103-E112.
5. Rund D, Azar I, Shperling O. A mutation in the promoter of the multidrug resistance gene (MDR1) in human hematological malignancies may contribute to the pathogenesis of resistant disease. *Adv Exp Med Biol.* 1999;457:71-75.
6. Gerlach JH, Kartner N, Bell DR, Ling V. Multidrug resistance. *Cancer Surv.* 1986;5:25-46.
7. Dalton WS, Miller TP. Multidrug resistance. In: DeVita VT Jr, Hellman S, Rosenbery SA, eds. *Cancer: Principles and Practice of Oncology*. PPO Updates. Philadelphia: J.B. Lippincott; 1991:1-13.
8. Persidis A. The business of pharmacogenomics. *Nat Biotechnol.* 1998;16:209-210.
9. Regalado14. Peng J-B, Hediger MA. Nomenclature of mammalian transporter genes. 1999. Private communication.
10. Inventing the pharmacogenomics business. *Am J Health-Syst Pharm.* 1999;56:40-50.
11. Evans WE, Relling MV. Pharmacogenomics: translating functional genomics into relational therapeutics. *Science.* 1999;286:487-491.
12. Sadée W, Graul RC, Lee A. Classification of membrane transporters. In: Amidon GL, Sadée W, eds. *Membrane Transporters as Drug Target (Pharmaceutical Biotechnology, Vol. 12)*. Plenum; 2000:29-58.
13. <http://www.med.rug.nl/mdl/tab3.htm>
14. <http://www.gene.ucl.ac.uk/cgi-bin/nomenclature/searchgenes.pl>