Preface

Studies of membrane transporters have had great impact on our understanding human diseases and the design of effective drugs. About 30% of current clinically marketed drugs are targeting membrane transporters or channels. *Membrane Transporters: Methods and Protocols* provides various practical methodologies for the ongoing research on membrane transporters. To provide readers the most up-to-date information, several emerging fields and methodologies are embraced in this book, including pharmacogenomics, bioin-formatics, and microarray technology. Pharmacogenomics studies of membrane transporters are useful in drug discovery and in predicting drug responses in the clinic. In this volume, the current status of pharmacogenomics studies of transporters is reviewed and research methodologies in this field are described.

Transporter classification is important in studying the structure and function of membrane transporters and has thus triggered intensive interest in recent years. *Membrane Transporters: Methods and Protocols* provides a systematic classification of all transmembrane transport proteins found in living organisms on Earth. This classification system will be helpful for further studies on various aspects of membrane transporters, especially for such large-scale gene expression studies as those employing microarray technologies.

Bioinformatics is frequently used in transporter studies and has become indispensable for all kinds of research methods. Commonly used bioinformatics methods, such as databases and tools for sequence analysis and motif studies, are explained in order to facilitate membrane transporters research. Because of heterogeneous sources and tremendous amounts of data, data integration has become one of the most important issues in transporter studies. The brief introduction to data integration methodology offered here can help researchers manage their data to facilitate further knowledge discovery.

In *Membrane Transporters: Methods and Protocols*, the authors not only provide methods and protocols, but also share their valuable hands-on experience with readers. Microarray technology has just begun to bloom in recent years. At present, shared experience in using this relatively new technology is especially helpful. Our book provides guidelines as well as the authors' experience in both microarray experiments and data analysis. We believe our readers will find them useful for understanding, designing, and carrying out their own microarray tests in membrane transporters. Laser capture microdissection

is also described as another recently developed technology useful for the study of transporter gene expression.

Because structural and functional studies have been the main issues in transporter studies and are also essential in pharmacogenomics, methodologies and protocols from various points of view are provided to tackle structure–function problems. For example, for studying the structural biology of membrane transporters, a series of techniques and methodologies are discussed, including small-angle X-ray scattering (SAXS), nuclear magnetic resonance (NMR), and molecular modeling. Some methods for studying the structure–function correlation are described, such as site-directed mutagenesis, immunocytochemistry, confocal microscopy, and kinetics studies, including equilibrium binding.

Some methods go beyond the structure–function study and may have poten-tial implications for development of novel therapeutics, as well as for studying gene–drug interactions and improving drug efficacy. For example, fluorescence techniques can be applied to study interactions between drugs and P-glycoprotein multidrug transporter (Pgp), a transporter protein that plays an important role in drug resistance in many cancer therapies. The adenovirus-mediated gene transfer method in electrophysiological studies of ion channels in mammalian myocardium may help develop therapeutics against heart diseases.

Readers are encouraged to explore integrated views and comprehensive methodologies from different chapters of this book; the methods are not presented as isolated techniques but are complementary. One method often also includes descriptions of several related techniques. For example, studying the expression system of *Xenopus* oocytes uses immunocyto-chemical, electrophysiological, and kinetic methods. Reconstitution allows detailed characterization of membrane transporters in further depth and allows the use of other techniques such as fluorescence spectroscopy. To measure intracellular pH, which is important for understanding the role of membrane transporters in cellular processes, NMR and fluorescence techniques may also be used.

Membrane Transporters: Methods and Protocols strives to deliver to readers not only a collection of practical protocols that can be used immediately in the lab but also critical surveys of key topics by leading researchers in this field. Readers can develop their own workable schemes for their personal studies based on the application of these powerful methodologies. Biomedical researchers in various fields who are interested in membrane transporters, including biochemists, molecular biologists, geneticists, physiologists,

microbiologists, immunologists, bioinformatics researchers, pharmaceutical scientists, and clinical researchers, will find the book useful.

I would like to thank all of the authors for sharing their valuable experience and insights with the research community at large. I would also like to thank series editor John Walker for his help in reviewing the manuscripts.

***Qing Yan,** MD, PhD*

2

The IUBMB-Endorsed Transporter Classification System

Wolfgang Busch and Milton H. Saier, Jr.

1. Introduction

Transport systems are essential to every living cell. They (1) allow the entry of all essential nutrients into the cell and its compartments, (2) regulate the cytoplasmic concentrations of metabolites by excretion mechanisms, (3) provide physiological cellular concentrations of ions that can differ by several orders of magnitude from those in the external medium, (4) export macromolecules such as complex carbohydrates, proteins, lipids, and DNA, (5) catalyze export and uptake of signaling molecules that mediate intercellular communications, (6) prevent toxic effects of drugs and toxins by mediating active efflux, and (7) participate in biological warfare by exporting biological active agents that insert into or permeate the membranes of target cells. Transport is an essential aspect of all life-endowing processes: metabolism, communication, biosynthesis, reproduction, and both cooperative and antagonistic interorganismal behaviors.

This chapter provides a summary of the recently developed transporter classification (TC) system ***(1–3)*** formally adopted by the International Union of Biochemistry and Molecular Biology (IUBMB) in June 2001. The development of a classification system for transport proteins has allowed us to comprehensively view transport systems from structural, functional, and evolutionary standpoints ***(2,4,5)***. This development has been strongly influenced by recent progress in computational biology and genome sequencing. Since our last description of the TC system ***(3)***, we have expanded the transporter classification system by (1) introducing new families and classes of transporters, (2) expanding the memberships of pre-existing families, (3) pro-

From: *Methods in Molecular Biology, vol. 227: Membrane Transporters: Methods and Protocols*
Edited by: Q. Yan © Humana Press Inc., Totowa, NJ

Table 1
TC System Overview

1. Channels/Pores
- A. α-Helical protein channels
- B. β-Barrel protein porins
- C. Toxin channels
- D. Nonribosomally synthesized channels
- E. Holins

2. Electrochemical Potential-Driven Transporters
- A. Porters
- B. Nonribosomally synthesized porters
- C. Ion-gradient-driven energizers

3. Primary Active Transporters
- A. P–P-bond hydrolysis-driven systems
- B. Decarboxylation-driven systems
- C. Methyltransfer-driven systems
- D. Oxidoreduction-driven systems
- E. Light-absorption-driven systems

4. Group Translocators
- A. Phosphotransfer-driven systems

5. Transmembrane Electron Carriers
- A. Two-electron transfer carriers
- B. One-electron transfer carriers

8. Accessory Factors Involved in Transport
- A. Auxiliary transport proteins

9. Incompletely Characterized Transport Systems
- A. Recognized transporters of unknown biochemical mechanism
- B. Putative uncharacterized transport proteins
- C. Functionally characterized transporters lacking identified sequences

Note: The hierarchical system for classifying families of transporters in the TC system is presented. Classes are presented in bold and subclasses are presented below the class designations.

viding more detailed annotation of these families and proteins, (4) updating reference citations relevant to proteins described in the TC system, and (5) creating an interactive database, which we have named TCDB. The results of our analyses, made possible by these updates, are summarized here. For a more detailed account of these studies, *see* **ref. *6***.

2. Classes of Transporters

The properties of the different hierarchical units that comprise the TC system are described briefly in this section and illustrated in **Table 1**. For more extensive descriptions, visit the websites and database (http://tcdb.ucsd.edu).

2.1. Class 1. Channels/Pores

Class 1 consists of channel-type facilitators. Transmembrane channel proteins span the lipid bilayer as either α-helices or β-strands. The transport mode of these systems usually involves the unencumbered passage of molecules across membranes in a process related to passive diffusion. Thus, channel-mediated transport occurs by facilitated diffusion, an energy-independent process in which the substrate passes through the transmembrane aqueous pore or channel without the coupling of the translocation process to another chemical or vectorial process.

2.1.1. 1.A. α-Type Channels

Channel proteins of subclass 1.A usually consist of bundles of transmembrane α-helices that form α-helical aqueous pores or channels. Rarely, β-strands contribute to the channel. These channels are found ubiquitously in the membranes of all types of organism.

2.1.2. 1.B. β-Barrel Porins

The transmembrane pores of subclass 1.C proteins consist exclusively of β-strands that form β-barrels. These channels are found in the outer membranes of bacteria, mitochondria, and plastids.

2.1.3. 1.C. Pore-Forming Toxins

Polypeptides of subclass 1.C attack target cells other than the producer cell by inserting into the target cell membrane, usually forming oligomeric transmembrane pores. The toxic effects are caused by allowing the free flow of electrolytes and other small molecules across the membrane. Polypeptides of this subclass are probably synthesized universally by all types of living cell.

2.1.4. 1.D. Nonribosomally Synthesized Channels

Subclass 1.D systems usually consist of small molecular building blocks such as L- and D-amino acids and hydroxy acids. Assembly of the molecular building blocks allows construction of oligomeric transmembrane ion channels. “Depsipeptides” and amino acid-free substances of this class usually provide a function related to biological warfare. Most of these substances are synthesized by bacteria and fungi.

2.1.5. 1.E. Holins

The many families of channel-forming holins comprise subclass 1.E. They do not exhibit significant sequence similarity between families, but all holins display common structural and functional characteristics. The primary func-

tion of holins is to export murein hydrolases across the cytoplasmic membranes of bacteria where they hydrolyze the cell wall polymer as a prelude to cell lysis. Holins may also facilitate leakage of electrolytes and nutrients from the cell cytoplasm, thereby promoting cell death. They are encoded within the genomes of Gram-positive and Gram-negative bacteria as well as the bacteriophage of these organisms.

2.2. Class 2. Electrochemical Potential-Driven Transporters

Class 2 transport systems, also called secondary carrier-type facilitators, usually exhibit strict stereospecifity and are energy coupled to the proton motive force (pmf) or the sodium motive force (smf).

2.2.1. 2.A. Porters

Subclass 2.A consists of transport systems that utilize carrier-mediated processes to catalyze uniport (a single species is transported either by facilitated diffusion or in a membrane potential-dependent process if the solute is charged), antiport (two or more species are transported in opposite directions in a tightly coupled process, not coupled to a direct form of energy other than chemiosmotic energy), and/or symport (two or more species are transported together in the same direction in a tightly coupled process, not coupled to a direct form of energy other than chemiosmotic energy). These systems are ubiquitous, being found in all living organisms.

2.2.2. 2.B. Nonribosomally Synthesized Porters

Like class 1.D, nonribosomally synthesized channels, molecules of subclass 2.B may be depsipeptides or non-peptide-like substances. They usually facilitate translocation by complexing an ion in their hydrophilic interior, exposing their hydrophobic exterior and moving from one side of the bilayer to the other. Transport can be electrophoretic if the free porter can cross the membrane in the uncomplexed form, or it can be electroneutral if only the complex can cross the membrane. Most of these molecules are products of bacteria and fungi.

2.2.3. 2.C. Ion-Gradient-Driven Energizers

Class 2.C energizers use the proton or sodium motive force across the cytoplasmic membrane. The mechanism is poorly understood, but they undoubtedly couple proton (H^+) or sodium (Na^+) fluxes to the energized process. Currently recognized energizers can drive bacterial flagellar rotation or active transport across the outer membranes of Gram-negative bacteria. They belong to a single family.

2.3. Class 3. Primary Active Transporters

Class 3 transporters use a primary source of energy as compared with a secondary (chemiosmotic) source of energy to drive active transport of solutes against concentration gradients.

2.3.1. 3.A. P–P-bond Hydrolysis-Driven Transporters

Transport systems of subclass 3.A hydrolyze the diphosphate bond of inorganic pyrophosphate or a nucleoside triphosphate to drive the active uptake and/or extrusion of a solute or solutes. The transport protein may or may not be transiently phosphorylated, but the substrate is not chemically modified. Members of this subclass are found in all domains of the living world.

2.3.2. 3.B. Decarboxylation-Driven Transporters

Transport systems that drive solute uptake or extrusion by decarboxylation of a cytoplasmic substrate comprise subclass 3.B. These multisubunit transporters are currently thought to be restricted to prokaryotes and belong to a single family.

2.3.3. 3.C. Methyltransfer-Driven Transporters

A single characterized multisubunit protein family currently falls into subclass 3.C, the Na^+-transporting methyltetrahydromethanopterin: coenzyme M methyltransferases. These transporter complexes have been found only in archaea.

2.3.4. 3.D. Oxidoreduction-Driven Transporters

Subclass 3.D is comprised of transport systems that drive transport of a solute (H^+ or Na^+) energized by the exothermic flow of electrons from a reduced substrate to an oxidized substrate. These multisubunit systems are distributed in all domains of the living world.

2.3.5. 3.E. Light-Absorption-Driven Transporters

Transport systems that utilize light energy to drive transport of an ion are included in subclass 3.E. These systems and their homologs are distributed in all three domains of life.

2.4. Class 4. Group Translocators

Class 4 systems include transporters that chemically alter the substrate during transport across a membrane so that the species released into the cytoplasm differs from the one that was taken up.

2.4.1. 4.A. Phosphotransfer-Driven Group Translocators

Transport systems of the bacterial phosphoenolpyruvate:sugar phosphotransferase system are the only recognized group translocators of subclass 4.A. The product of the transport reaction, derived from extracellular sugar, is a cytoplasmic sugar-phosphate. The enzymatic constituents, which catalyze sugar phosphorylation, are superimposed on transport in a tightly coupled process.

2.5. Class 5. Transmembrane Electron Carriers

Class 5 proteins include systems that catalyze electron flow from one side of a biological membrane to the other. Thus, the electrons are transferred from donors localized to one side of the membrane to acceptors found on the other side. These systems contribute to or subtract from the membrane potential, depending on the direction of electron flow.

2.5.1. 5.A. Transmembrane Two-Electron Transfer Carriers

Subclass 5.A is restricted to systems that catalyze transfer of a pair of electrons across the membrane in one or more discrete steps without splitting the paired electrons.

2.5.2. 5.B. Transmembrane One-Electron Transfer Carriers

Subclass 5.B includes systems that catalyze the sequential transfer of single electrons across the membrane in a free-radical-type process.

2.6. Class 8. Accessory Factors Involved in Transport

Proteins that function with or are complexed to known transport proteins are included in category 8. In some cases, auxiliary proteins are considered to be an integral part of the transport system, and in such cases, the proteins are classified with the transporter. In this case, no distinct entry in category 8 is provided.

2.6.1. 8.A. Auxiliary Transport Proteins

Subclass 8.A consists of proteins that facilitate transport across one or more biological membranes but do not themselves participate directly in transport. These proteins always function in conjunction with one or more established transport system(s). They may provide a function connected with energy coupling to transport, play a structural role in complex formation, serve a biogenic or stability function, or play a regulatory role.

2.7. Class 9. Incompletely Characterized Transport Systems

Transport protein families for which insufficient information is available to allow classification in a defined class (e.g., TC classes 1–5) belong to category 9.

2.7.1. 9.A. Recognized Transporters of Unknown Biochemical Mechanism

Recognized families of transport proteins of unknown classification are grouped in subclass 9.A. These families include at least one member for which a transport function has been established, but either the mode of transport or the energy coupling mechanism is not known. They will be classified elsewhere when the transport mode and/or energy coupling mechanisms are characterized.

2.7.2. 9.B. Putative Uncharacterized Transport Proteins

Putative transport protein families are grouped into subclass 9.B if a transport function has been suggested for one or more members of the family, but evidence for such a function is not yet compelling. They will either be classified elsewhere when the transport function of a member becomes established or will be eliminated from the TC system if the proposed transport function is disproven.

2.7.3. 9.C. Functionally Characterized Transporters Lacking Identified Sequences

Transporters of particular physiological significance are included in category 9.C even though a family assignment cannot be made. When their sequences are identified, they will be assigned to an established family. This is the only TC subclass that includes individual proteins rather than protein families.

3. Practical Applications of Transporter Family Association

About 350 protein families were included in the TC system as of February 2002 (*see* TCDB). Affiliation with a family requires satisfying rigorous statistical criteria of homology ***(7)***. Briefly, a protein must exhibit a region of at least 60 residues in comparable portions of the protein that exhibit a comparison score in excess of 9 standard deviations (SD) with at least one established member of that family using any one of a number of programs such as GAP, RDF2, Los Alamos, or ALIGN (*see* **ref. *7***). Whereas the classes and subclasses distinguish functionally distinct types of transporters, the families and subfamilies provide a phylogenetic basis for classification. The TC system is thus a functional/phylogenetic system of classification. Families very rarely cross class or subclass lines.

Recognition of a phylogenetic relationship based on sequence similarity allows certain conclusions to be drawn regarding three-dimensional structural features. Any two proteins that can be shown to be homologous (i.e., that

exhibit sufficient primary and/or secondary structural similarity to establish that they arose from a common evolutionary ancestor) can be expected to exhibit strikingly similar three-dimensional structures, although a few exception have been noted ***(8)***. Therefore, extrapolation from one member of a family of known structure to all other members becomes justifiable. Extrapolation of structural data to other proteins should never be made if homology has not been established. Similar arguments apply to mechanistic considerations. Thus, the mechanism of solute transport is likely to be similar for all members of a permease family, and variations on a specific mechanistic theme will be greatest when the sequence divergence is greatest. By contrast, for members of any two independently evolving permease families, the transport mechanism may be strikingly different. Extensive experimental work has established that phylogenetic data can also be used to predict substrate specificity, polarity of transport, and even intracellular localization depending on the family and the degree of sequence divergence observed within that family ***(3,9)***.

Transport system families included in the current TC system are described in database format on the World Wide Web (http://tcdb.ucsd.edu). TCDB provides detailed descriptions of and reference citations for (1) TC classes, (2) subclasses, (3) families (4) subfamilies and, (5) individual proteins. Additionally, relevant research tools can be found on our website, facilitating examination of the world of transport proteins. TCDB is equipped with a search tool that allows the user to search by key word, gene name, family, or protein sequence. Any protein demonstrably homologous to a TC family member can be identified using TC-BLAST. TCDB is interconnected with other useful databases and websites.

4. Family Characteristics

Key features of the transporter families currently recognized in TCDB are summarized in Table 2 of **ref.** ***(6)***. The TC number of the family, the substrates transported, and the size ranges of the individual protein members within each family are presented. Additionally, the probable numbers of transmembrane segments in the integral membrane constituents of the family are predicted and the organismal groups in which members of each family have been identified are presented. The major conclusions summarized in this chapter are based on the data presented in **ref.** ***(6)***.

Several TC family entries are actually superfamilies. In such cases, TCDB indicates the numbers of subfamilies currently recognized within that superfamily. The VIC (1.A.1), MF (2.A.1.), and ABC (3.A.1) superfamilies are the largest and most diverse transporter superfamilies currently recognized, but several other TC families have achieved superfamily status. The interested

reader is referred to TCDB for further explanation. Some of these superfamilies are discussed in the following sections.

5. Transport Protein Topological Types

The topologies of proteins within the different families of the TC system have been predicted using topological prediction programs such as WHAT ***(10)*** and TOPPRED ***(11)***. In relatively few instances have protein topologies been experimentally established.

We previously proposed that channels and carriers are fundamentally different at both structural and functional levels, but that the former may have been the evolutionary precursors of the latter ***(4)***. In **Fig. 1**, the predicted topologies of α-type channels (subclasses 1.A+1.C+1.E) and carrier proteins (subclass 2.A) are shown. The numbers of families represented are plotted versus the numbers of transmembrane segments (TMSs) found in the protein constituent types.

As seen in **Fig. 1**, the topological types that comprise α-type channels differ fundamentally from secondary carriers. Most families of α-helical channels include proteins with one to six TMSs, whereas the vast majority of carrier-type families display 10–14 TMSs. Almost all proteins in subclasses 1.C and 1.E display just one or two TMSs (data not shown), but channel proteins of subclass 1.A can exhibit up to 24 TMSs per polypeptide.

The small numbers of TMSs in most channel-forming proteins reflect their oligomeric structures, whereas the larger numbers of TMSs in the carriers reflect their basically monomeric constructions. The average numbers of TMSs for subclasses 9.A and 9.B are more representative of channellike proteins than carriers, suggesting that the majority of the proteins in these categories that are transporters will prove to be channels. However, some will undoubtedly prove to be carriers, and a few may be found to function by novel mechanisms. The data presented in **Fig. 1** allow one to predict which class 9.A families are likely to be members of TC class 1 or 2.

It has been proposed that large complex transport systems arose progressively from smaller simpler ones ***(4)***. Subclass 1.A channels could have developed from toxinlike peptide channels of subclass 1.C or holinlike channels of subclass 1.E, and subclass 1.A channels may have been the evolutionary precursors of porters. The latter proteins have more TMSs, and by virtue of their increased structural and functional complexity, it is reasonable to propose that carriers arose from channels in processes that involved internal gene duplication events ***(7)***. In fact, sequence analyses have revealed the presence of internal repeat sequences in many of the proteins that comprise families of secondary and primary active transporters as well as some channel proteins

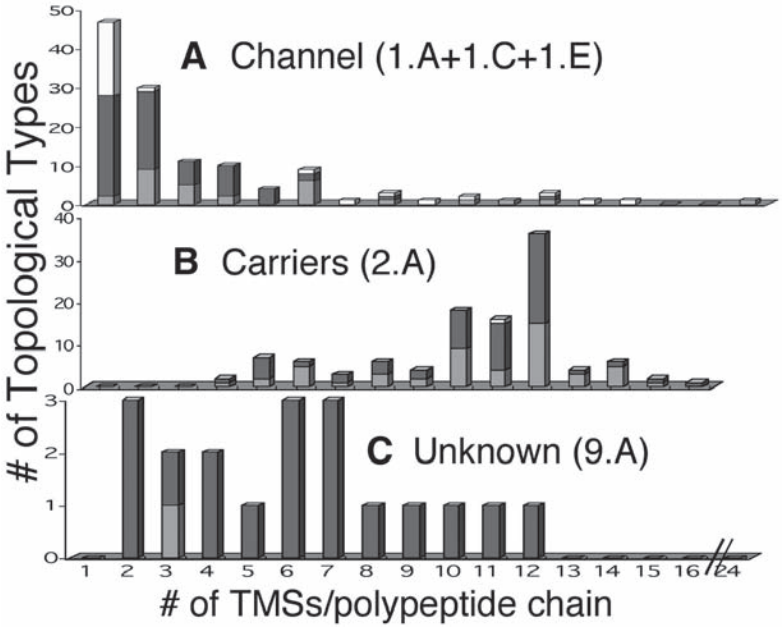

Fig. 1. Distribution of the various topological types of transporters in three subgroups of the TC system: **(A)** channels (TC classes 1.A plus 1.C plus 1.E); **(B)** porters (2.A); **(C)** transporter types of unknown mechanism (9.A). Grey, established; black, putative; white, uncertain.

(1,5,7). The repeat units of these complex transporters resemble the full-length sequences of the simpler channels ***(9)***. It is important to note that very few families of transporters include homologs that function in a capacity other than transport ***(5)***. Arguments that primary active transporters and group translocators resulted from superimposing catalytic proteins such as enzymes onto channels and carriers have been presented ***(4)***.

6. Size Variations in Transporters

We have evaluated the size ranges observed for the families that comprise the different subclasses of the TC system ***(6)***. Most families in TC subclass 1.A are intermediate in size (100–1000 residues per polypeptide chain), but a few are less than 100 or more than 1000 residues. By contrast, most transporter types in TC subclasses 1.C and 1.E are much smaller, and all of subclass 1.E proteins are small. Proteins of subclass 2.A are never smaller than 100 residues in length and most exceed 400 residues. Those of subclass 9.A more closely resemble the channels of subclass 1.A. These size differences reflect the topo-

logical differences noted earlier. Size variability is minimal for channel-forming subclasses 1.C and 1.E but substantially greater for subclass 1.A. The variance for proteins of subclass 2.A is comparable to that of the channels of subclass 1.A. Interestingly, that in subclass 9.A is minimal.

7. Transporter Family Sizes

Recognized transporter families differ over three orders of magnitude with respect to the numbers of currently sequenced proteins which comprise them. The vast majority of the TC families are of intermediate size, having between 6 and 500 currently sequenced members. As of February 2002, only 66 families have 5 members or less, and only 15 families include more than 500 sequenced protein members. Most of these large families are ubiquitous, having membership from all major domains of living organisms. They are probably ancient families dating back before the time that archaea and eukaryotes split apart from bacteria. Of the channels, the five largest families are the ubiquitous VIC, MIP, and HSP70 families (TC 1.A.1, 1.A.8, and 1.A.33, respectively) as well as the eukaryotic-specific TRP-CC and LIC families (TC 1.A.4 and 1.A.9, respectively). Of the secondary active carriers, the MF (2.A.1), RND (2.A.8), DMT (2.A.7), NSS (2.A.22), MC (2.A.29), and MATE (2.A.68) superfamilies have the largest membership. Among the primary active carriers, the ABC (3.A.1), P-type ATPase (3.A.3), and COX (3.D.4) families have the greater family membership with decreasing numbers of members in this order. One putative transporter family (FAT [9.B.17]), the acyl-CoA synthase family, includes thousands of sequenced proteins, but a role of these proteins in transport is not well established ***(12)***. If these enzymes couple fatty acid uptake to coenzyme A thio-esterification, such a process would provide an example of group translocation in which the substrate is modified during transport ***(13)***. These systems may, therefore, be group translocators, which, in addition to the families of the bacterial phosphotransferase system, would fall into TC class 4.

8. Family Representation in Bacteria, Archaea, and Eukaryotes

The occurrence of transporter types in the three domains of life was evaluated by creating a plot that shows the representation of the members of a family in the three domains of living organisms ***(6)***. Most ubiquitous families are found within subclass 2.A (*see* **Subheading 2.2.1.**).

We suggest that this fact in part reflects the larger polypeptide sizes of these usually monomeric proteins. The smaller oligomeric channel-forming proteins may have undergone more extensive sequence divergence, leading to the appearance of multiple families exhibiting insufficient degrees of sequence similarity to allow the establishment of homology. The large protein size

facilitates distant phylogenetic relationship detection, and requirements for retention of specific functional properties restrict the natural process of sequence divergence that occurs over evolutionary time.

Most of the channel families in class 1 are restricted to a single organismal type. This fact may in part reflect the ease with which simple channellike functions can be generated *de novo* from small peptides, but it may also reflect the absence of strict constraints preventing sequence divergence. It is also possible that most of these families arose late, after the three domains separated. If so, lateral transfer of genes encoding these channel proteins occurred very seldom.

Bacterial transport protein families are more prevalent than are those found only in eukaryotes (47% vs 26%). This could be a reflection of evolutionary pressure forcing bacteria to maintain diversity in order to remain adaptive in response to a wide range of environmental stress conditions. Multicellular eukaryotes generally create internal homeostatic environments that obviate the need for extensive cellular stress-response mechanisms. The greater diversity of prokaryotic transporters may also reflect the greater period of evolutionary time that these organisms have been on Earth. Eukaryotes may have evolved from a limited subgroup of primordial bacteria, and these bacteria may not have possessed the full complement of prokaryotic transporter families. Alternatively, eukaryotes may have lost families that were present in the eukaryotic progenitor. Although eukaryotes exhibit fewer families than prokaryotes, they have proliferated tremendous numbers of paralogs within certain families ***(6)***, probably for very specific purposes involving tissue-specific and organelle-specific functions. Some animals, for example, encode within their genomes hundreds of paralogs of a single family ***(14)***.

There are very few archaeal-specific transporter families. This observation may in part reflect the fact that functional data are sparse for archaeal proteins. Molecular biological research over the past six decades has focused almost exclusively on bacterial and eukaryotic systems. However, if archaea, like eukaryotes, arose from primordial bacteria, they may have acquired a restricted subset of proteins from the ancestral bacterium and they then would have had less time to diversify. This interesting postulate should be subject to empirical research. It is noteworthy that the recognition of a transporter family is facilitated by the availability of genomic sequence data only if functional data are available.

Ubiquitous families represent only 14% of the total found in the TC system, and those represented in two of the three domains of life are still less numerous. Those shared by bacteria and eukaryotes (8%) exceed those shared by archaea and bacteria (4%) by about twofold. Only one family is found in the archaeal and eukaryotic domains but not in the bacterial domain. Some of the families found in two but not three domains will undoubtedly prove to

be ubiquitous when more sequence data and more sensitive search tools become available.

9. Three-Dimensional Structural Data for Transporters

Detailed structural data for transport proteins will be necessary in order to gain an ultimate understanding of transport processes. Unfortunately, very few membrane proteins have yielded to the techniques of the X-ray crystallographer. Despite the fact that integral membrane proteins comprise about one-third of all proteins, less than 2% of the available three-dimensional structures reported in the PDB database are for such proteins. The 40 transporters for which high-resolution structural data are available are tabulated in **ref. *(6)***.

10. Conclusions and Perspectives

The TC system displayed in TCDB allows any researcher to easily gain access to the extensive body of knowledge available for transport systems. With the tools we provide on our websites, one can convincingly view the relationships between the established transporters in the TC system and novel proteins that have recently been or will soon be sequenced. A valuable tool for this purpose is TC-BLAST which performs a BLAST search against TCDB, revealing the nearest homologs and the families in the TC system to which the query sequences belong. There are also a number of other programs available on our website that help to bring to light the features of newly discovered transporters (e.g., WHAT, AveHAS, BBF, TV, etc.). The interested reader is invited to view our website for in-depth exposure to these tools.

Because of the nature of transporters as integral membrane proteins, we believe that computational approaches will prove particularly useful for their structural elucidation. Phylogenetic analyses should reveal structure–function relationships that greatly facilitate empirical research. With the availability of better tools, it will be easier to track phylogenetic relationships and to uncover the pathways by which transporters have evolved. Tracking the evolutionary pathways taken for the appearance of topologically dissimilar proteins within a family, and for families of transporters exhibiting dissimilar mechanisms of action, will prove to be a daunting but highly worthwhile endeavor. For this purpose, it will be important to create new and more reliable topological prediction programs as well as programs that allow detection of very distant phylogenetic relationships ***(15)***.

Another crucial aspect of analyzing relationships between families and classes of families will involve designing a dataset of proteins within each family in an accurate but automated way. We are currently designing such

software for the TC system. If the datasets for the different families are sufficiently extensive and reliable, we will be able to derive accurate sequence motifs and patterns that characterize the families and have structural–functional significance. New approaches for characterizing families will undoubtedly come to light.

Research into the molecular basis of transport processes will be greatly facilitated by the use of *in silico* approaches. Such approaches are likely to reveal characteristics dictated by primary protein sequences that are currently masked because of, first, our limited understanding of membrane proteins and, second, the inadequacies of currently available computational technologies. Bioinformatic advances should facilitate, for example, the development of transport protein-specific drugs that may comprise new classes of antibiotic, antiprotozoan, and antifungal substances. Many other unforeseen advances can be anticipated.

11. Summary

Over the past several years, we have designed and expanded the TC system. This system is a functional/phylogenetic system designed for the classification of all transmembrane transport proteins found in living organisms on Earth. It parallels but differs from the strictly functional EC system, developed decades ago by the Enzyme Commission of the IUBMB for the classification of enzymes.

The TC system has recently been adopted by the IUBMB as the internationally acclaimed system for the classification of transporters. Here, we described the classes and subclasses of the TC system. Based on the characteristics of the nearly 400 families of transport systems included in the TC system, we summarized statistical analyses of these families and their constituent proteins. Specifically, we reported analyses of various transporter types for size and topological differences and analyzed the families for their protein memberships as well as the numbers and organismal sources of their constituent proteins. We showed that channels and carriers exhibit distinctive structural and topological features. Bacterial-specific families outnumber eukaryotic-specific families about two to one, whereas ubiquitous families, found in all three domains of life, are about half as numerous as eukaryotic-specific families. We propose that the ubiquitous families are the ancient families that arose before archaea and eukaryotes segregated from bacteria. The results argue against appreciable horizontal transfer of genes encoding transporters between the three domains of life over the past two billion years.

Acknowledgments

We wish to thank Arnost Kotyk, Ian Paulsen, Can Tran, Nelson Yang, and YuFeng Zhai for useful discussions. We are particularly grateful to YuFeng Zhai for designing software that facilitated the analyses summarized in this chapter. Work in our laboratory is supported by grants from the National Institutes of Health.

References

1. Saier, M. H., Jr. (1998) Molecular phylogeny as a basis for the classification of transport proteins from bacteria, archaea and eukarya, in *Advances in Microbial Physiology* (Poole, R. K., ed.), Academic, San Diego, CA, pp. 81–136.
2. Saier, M. H., Jr. (1999) Genome archeology leading to the characterization and classification of transport proteins. *Curr. Opin. Microbiol.* **2,** 555–561.
3. Saier, M. H., Jr. (2000) A functional–phylogenetic classification system for transmembrane solute transporters. *Microbiol. Mol. Biol. Rev.* **64,** 354–411.
4. Saier, M. H., Jr. (2000) Vectorial metabolism and the evolution of transport systems. *J. Bacteriol.* **182,** 5029–5035.
5. Saier, M. H., Jr. (2001) Evolution of transport proteins, in *Genetic Engineering. Principles and Methods, Vol. 23* (J. K. Setlow, ed.), Kluwer Academic/Plenum, New York, pp. 1–10.
6. Busch, W. and Saier, M. H., Jr. (2002) The Transporter Classification (TC) System, **2002,** *Crit. Rev. Biochem. Mol. Biol.* **37,** 287–337.
7. Saier, M. H., Jr. (1994) Computer-aided analyses of transport protein sequences: gleaning evidence concerning function, structure, biogenesis, and evolution. *Microbiol. Rev.* **58,** 71–93.
8. Saier, M. H., Jr. and T.-T. Tseng (1999) Evolutionary origins of transmembrane transport systems, in *Transport of Molecules Across Microbial Membranes, Symposium 58, Society for General Microbiology* (Broome-Smith, J. K., Baumberg, S., Stirling, C. J., et al., eds.), Cambridge University Press, Cambridge, UK, pp. 252–274.
9. Saier, M. H., Jr. (2000) Families of proteins forming transmembrane channels. *J. Membr. Biol.* **175,** 165–180.
10. Zhai, Y. and Saier, M. H., Jr. (2001) A web-based program (WHAT) for the simultaneous prediction of hydropathy, amphipathicity, secondary structure and transmembrane topology for a single protein sequence. *J. Mol. Microbiol. Biotechnol.* **4,** 501–502.
11. Claros, M. G. and von Heijne, G. (1994) TopPred II: An improved software for membrane protein structure predictions. *CABIOS* **10,** 685–686.
12. Saier, M. H., Jr. and Kollman J. (1999) Is FatP a long chain fatty acid transporter? *Mol. Microbiol.* **33,** 670–672.

13. Faergeman, N. J., Black, P. N., Zhao, X. D., et al. (2001) The acyl-CoA synthetases encoded within FAA1 and FAA4 in Saccharomyces cerevisiae function as components of the fatty acid transport system linking import, activation, and intracellular utilization. *J. Biol. Chem.* **276,** 37,051–37,059.
14. *C. elegans* Sequencing Consortium. (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282,** 2012–2017.
15. Pei, J. and Grishin, N. V. (2001) AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics* **17,** 700–712.