Trends in Pharmacological Sciences

CellPress

Review

Classification Systems of Secondary Active Transporters

Emelie Perland[1] and Robert Fredriksson[1,*]

Membrane-bound solute carrier (SLC) transporter proteins are vital to the human body, as they sustain homeostasis by moving soluble molecule as nutrients, drugs, and waste across lipid membranes. Of the 430 identified secondary active transporters in humans, 30% are still orphans, and systematic research has been requested to elaborate on their possible involvement in diseases and their potential as drug targets. To enable this, the various classification systems in use must be understood and used correctly. In this review, we describe how various classification systems for human SLCs are constructed, and how they overlap and differ. To facilitate communication between researchers and to avoid ambiguities, everyone must clearly state which classification system they are referring to when writing scientific articles.

Trends

SLCs are crucial for maintaining homeostasis within the body as they control molecular trafficking across cellular lipid membranes. SLCs are also implicated in several diseases and they constitute possible drug targets.

There are fewer publications on SLCs than on other protein families, showing that the SLC family is understudied. Systematic research has been requested within the field.

There are several SLC classification systems in use, to facilitate efficient communication between researchers.

However, the many classification systems result in ambiguities and misunderstandings.

Understanding the different classification systems in use today aids a systematic approach on SLC research and should be considered when studying disease processes or novel potential drug targets.

A Unified View of SLC Classification Systems

The time is right for a coordinated research effort on SLC structures, specificities and functions [1], as research has shown they play crucial roles in health and disease. SLCs are secondary active or facilitated transporters that translocate soluble molecules across cellular membranes [2]. They can use ion gradients to drive uphill transports, work as exchangers, or facilitate passive diffusion of specific molecules [2]. SLCs are vital for maintaining homeostasis in the body and in individual cells, and genetic polymorphisms in SLCs are associated with several diseases, such as amyotrophic lateral sclerosis (ALS), Alzheimer's disease, and schizophrenia [3,4]. Furthermore, SLCs can function as drug targets [3], as well as constitute paths for drug absorption into specific organs [5].

The number of SLCs in humans is constantly rising. In 2004, there were 298 known SLCs [2] and 9 years later, this number had increased to 395 [4]. The most recent estimate is that there are 456 SLCs [1], although the authors of the referred article did not present the source for this number. In this review, we present a list of 430 unique secondary active transporters found in the human genome, of which, 400 so far have been classified as SLCs. We estimate that the expression profile and/or function remain unknown for 30% of the 430 unique sequences. In addition, there are 26 known pseudogenes encoding SLCs.

New proteins are constantly being discovered, therefore, it is important to have a standardized method for naming and classifying them. However, there are currently several systems for classifying SLCs, resulting in inconsistency. Our goal is therefore to clarify the various SLC classification systems containing human proteins (systems with human proteins only are summarized in Box 1, while multiorganism systems are in Box 2), and show how they are all inter-related. The purpose is to facilitate efficient communication between researchers, which will enable systematic and efficient research. Finally, with the use of the classification systems, we also identify 30 new proteins that are novel putative SLCs.

[1]Department of Pharmaceutical Bioscience, Molecular Neuropharmacology, Uppsala University, Uppsala SE 7512, Sweden

*Correspondence: Robert.Fredriksson@farmbio.uu.se (R. Fredriksson).

Box 1. Human Protein Classification Systems Including Secondary Active Transporters

The HGNC uses various characteristics, homology, functions, and structures to arrange human genes into families [6]. At present, there are 430 known SLCs, divided into 52 families, where most have been classified by HGNC. Here, SLCs with >20% protein sequence identity are considered members of the same family [2]. The SLC tables database lists all present 52 families (http://slc.bioparadigms.org/, 19 July 2016) and their members.

Box 2. Multiorganism Classification Systems for Transporter Proteins

There are two major multiorganism classification systems that include SLC proteins: the TCDB and the Protein Family systems. The TCDB system focuses only on transport proteins and incorporates both functional and phylogenetic information when grouping transporters [20], where characterized secondary active transporters belong to Group 2.A. TCDB aims to provide all transporters with distinct identity numbers. The Pfam system includes all types of proteins, including SLCs. Pfam is based on protein sequence alignments in which functional domains are identified and used to group proteins into families and larger clans [23]. SLC families are located within various Pfam clans, where the MFS, APC CPA/AT, and DMT clans contain more than one annotated SLC family.

Classification of Human Secondary Active Transporters

The HUGO Nomenclature Committee Gives Human SLCs Names Using a Family Root System

The HUGO Gene Nomenclature Committee (HGNC) provides human genes with unique symbols and manually curates the genes into larger families based on function, homology, or phenotype [6]. The database can be found at http://www.genenames.org/ and a search for solute carrier returns 391 SLCs and four pseudogenes (19 July 2016). In the HGNC database, genes are usually grouped under a common root symbol to indicate a family; solute carriers have been given the root SLC. HGNC ensures that a certain gene only will be given one approved symbol, so proteins will contain their original names; for example, major facilitator superfamily domain containing (MFSD)7 is clustered into SLC49 [7], and is therefore SLC49A3, but the protein is still named MFSD7.

The SLC gene nomenclature system was originally developed as a collaboration between the first chair of HGNC and the SLC tables (http://www.bioparadigms.org/slc/intro.htm), resulting in significant overlap between entries in these two databases. According to the SLC tables database, there are currently 397 SLCs and 22 pseudogenes, for a total of 417 SLCs, which are divided into 52 families. Some of the families are further subdivided by phylogenetic analyses into four clusters designated α, β, δ, and γ [8,9]. Attempts have also been made to classify transporters based on function; where proteins with similar substrates, structures, and mechanisms usually cluster together [10].

Table 1 summarizes the number of members per family, whereas the Online Supplemental Information Table S1 is an extended list including all proteins, systems, and symbols discussed in the review. Proteins in an SLC family have at least 20–25% amino acid sequence identity with at least one other family member [2].

There are some differences between the entries in the HGNC and SLC tables (see Online Supplemental Information Table S1 for details); for example, SLCO1B7 and SLC35E2B are listed in HGNC, but not in the SLC tables, and SLC6A10 is found in the SLC tables but not in HGNC. SLC6A10 is found as both a gene and a pseudogene in the SLC tables, but only the pseudogene is listed in HGNC. All this highlights discrepancies between these systems.

The human SLC naming system is also used for genes from other species, giving orthologous proteins the same name to facilitate scientific communication. This is because many gene annotation pipelines, such as the ENSEMBL pipeline [11], use the bases of human gene nomenclature systems when providing names for other species. However, it is important to