Database for renal collecting duct regulatory and transporter proteins

John Legato, Mark A. Knepper, Robert A. Star, and Raymond Mejia

Laboratory of Kidney and Electrolyte Metabolism, National Heart, Lung, and Blood Institute, Bethesda 20892-2690; Renal Diagnostics and Therapeutics, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda 20892-1268; and Mathematical Research Branch, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, Maryland 20892-5621

Submitted 14 February 2003; accepted in final form 6 March 2003

The mammalian kidney collecting duct plays an important role in the fine regulation of Na, K, water, and acid-base balance. Functional genomic and proteomic studies of the kidney offer new opportunities in the understanding of renal physiology and pathophysiology, and the collecting duct is an appropriate target tissue because of the relative simplicity of its cells and the ease of isolating or culturing large numbers of collecting duct cells. Study of the collecting duct includes assessment of gene expression and protein regulation and abundance. For example, DNA and protein microarrays can be used to quantify gene expression and protein regulation and abundance under varying physiological conditions. An Internet-accessible database has been devised for major collecting duct proteins involved in transport and regulation of cellular processes. The individual proteins included in this database are those culled from literature searches and from previously published studies involving cDNA arrays and serial analysis of gene expression (SAGE). Design of microarray targets for the study of kidney collecting duct tissues is facilitated by the database, which includes links to curated base pair and amino acid sequence data, relevant literature, and related databases. Use of the database is illustrated by a search for water channel proteins, aquaporins, and by a subsequent search for vasopressin receptors. Links are shown to the literature and to sequence data for human, rat, and mouse, as well as to relevant web-based resources. Extension of the database is dynamic and is done through a maintenance interface. This permits creation of new categories, updating of existing entries, and addition of new ones.

METHODS

We developed a database that organizes lists of genes found in collecting duct tissues from three mammalian species: human, rat, and mouse. Proteins are divided into categories by family relationships and functional classification, and each category is assigned a section in the database. Over 25 sections constitute the database, accessible at http://mrb.niddk.nih.gov/cddb/, and the section headings appear in a table of contents shown in Fig. 1. Each section includes links to the literature and to sequence information for genes, proteins, expressed sequence tags, and related information. The user can peruse a section or use a search engine at the bottom of the web page to search the database for a name or abbreviation or for a link to a sequence. Each entry in the database includes links to relevant papers in the kidney and collecting duct literature. We use links to PubMed (18) to generate MEDLINE (15) searches for retrieval of references. In addition, each entry includes links to curated sequence data available in LocusLink (17). Individual links are made to sequence and protein data for human, rat, and mouse. Links are then added as curated sequences become available for proteins.
Collecting Duct Database (cddb)

The database is aimed to serve as a learning tool to obtain curated information for the design of macromolecular targets to scan collecting duct tissue (human, rat, mouse).

The database focuses on regulatory and transport proteins expressed in the collecting duct, but when collecting duct proteins are a member of a larger family of proteins, common additional members of the family are included even if they have not been demonstrated to be expressed in the collecting duct.

Table of Contents

- Aquaporin Cyclers
- Adhesion Proteins and Collagens
- Apical Scaffold Proteins
- Cyclic Nucleotide Phosphodiesterases
- Cytoskeletal Proteins and Linkers/Molecular Motors
- G-Protein Coupled Receptors
- Nuclear/Cytoplasmic and Mitochondrial Synthase
- Hydrolases or Cytoskeleton
- Ion Transporters
- Metabolism Proteins
- Membrane Transporters
- Not Yet Classified
- Nuclear Receptors
- Protein Degradation Pathways
- Protein Kinases, Serine/Threonine
- Protein Kinases, Tyrosine Kinase Receptor
- Protein Phosphatases
- Proteoglycans
- SNAREs and Related Proteins
- Signal Proteins
- Small GTP Binding Proteins and Related Proteins
- Steroid Metabolism Enzymes
- Stress Proteins
- Tight Junction Proteins
- Transcription Factors
- Transporters and Channels

identified in the renal collecting duct and for proteins identified in kidney and similar in function or homologous to proteins identified in the collecting duct.

The database has been created using PostgreSQL (http://www.postgresql.org), an open-source relational database management system that implements most SQL constructs (16), and resides on the Mathematical Research Branch web server (http://mrb.niddk.nih.gov). The scripts that connect the database to the web pages are written in PHP (10; see also http://www.php.net), an open-source, server-side, cross-platform, HTML-embedded scripting language that is used to generate web pages dynamically. The web page scripts themselves are maintained under the control of the Concurrent Versions System (CVS) (7; see also http://www.cvshome.org), an open-source concurrent version system.

RESULTS

Consider a search of all search fields in the database for the search string “AQP”, as shown in Fig. 2. The search yields the aquaporins shown in Fig. 3. Although all the aquaporins shown are not expressed in collecting duct, we include proteins that belong to the same family and are expressed in kidney. Note that all search fields have been selected, and the search has not been restricted (filtered) by requiring an entry in any of the linked databases. The links are to search MEDLINE entries that refer to collecting duct (C), to MEDLINE entries that refer to kidney (P), to LocusLink entries for human, rat, and mouse (L), to the Online Mendelian Inheritance in Man (O) (8), to GeneCards (G) (19), to the Rat Genome Database (R) (24), and to the Mouse Genome Informatics database (M) (1).

After use of the links shown in Fig. 3, one may do another search as shown in Fig. 3 or return to the Table of Contents. A search for “avp” retrieves data links for the vasopressin receptors shown in Fig. 4.

Fig. 1. Table of Contents. Contains links to the individual sections of the database. See http://mrb.niddk.nih.gov/cddb/.

Fig. 2. Search Page. Describes how to search and shows a search of the database for the string “AQP”.

Fig. 3. Result of the search shown in Fig. 2. Includes links to data for aquaporin-1 through aquaporin-9.

Transporters and Channels
Additional external resources (see Fig. 2) include I.M.A.G.E. (14), the Mammalian Gene Collection (21), and UCSC Genome Bioinformatics (11, 23), where a search for additional information can be done. Links to “Review Board” and to frequently asked questions “FAQ” are also shown.

DISCUSSION

Advances in genomics have made possible the study of transcriptional changes that underlie differences among organisms (22), and microarray technology serves to facilitate gene identification (4). SAGE has been described for quantitative mRNA profiling in the mouse kidney (25) and for study of the transcriptome in a cortical collecting duct cell line (20).

Here we describe the implementation and use of a database to facilitate the design of experiments for the determination of gene expression in kidney collecting duct. Hence, we seek to retrieve references pertinent to each gene product and to retrieve the most reliable sequence information available. As data becomes available in the literature and databanks, the database is updated, and links are added. A web interface is available for maintenance (http://mrb.niddk.nih.gov/cddh/admin), including creation of new sections and additions and extensions of the database.


Each section of the database was reviewed by experts in the specific area. We express our appreciation to the reviewers, who are listed under “Review Board” at http://mrb.niddk.nih.gov/cddh/misc/review.html.

REFERENCES

23. UCSC Genome Browser. Genome Bioinformatics Group, Univ. of California Santa Cruz, The Regents of the University of California [Online]. http://genome.ucsc.edu [2001].