

Database for renal collecting duct regulatory and transporter proteins

John Legato,¹ Mark A. Knepper,¹ Robert A. Star,² and Raymond Mejia³

¹Laboratory of Kidney and Electrolyte Metabolism, National Heart, Lung, and Blood Institute, Bethesda 20892-2690; ²Renal Diagnostics and Therapeutics, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda 20892-1268; and ³Mathematical 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

Legato, John, Mark A. Knepper, Robert A. Star, and Raymond Mejia. Database for renal collecting duct regulatory and transporter proteins. *Physiol Genomics* 13: 179–181, 2003. First published March 19, 2003; 10.1152/physiolgenomics.00021.2003.—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 quantitate 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.

regulatory protein; transporter protein; microarray

THE MAMMALIAN KIDNEY COLLECTING DUCT plays an important role in the production of both concentrated and dilute urine, as well as regulation and balance of water, Na, K, acid-base, and divalent cations (6, 12, 13). As such, its physiology and pathophysiology is a subject of considerable interest (2). To this end, an area of

current investigation is gene expression and the regulation and abundance of collecting duct proteins (3). To provide a reference resource for such studies, and as a general tool for study of collecting duct tissues, we have developed a tool to access mRNA and protein sequences for genes expressed in the collecting duct, and in particular for regulatory and transporter proteins.

An important application is in the design of cDNA microarrays. DNA microarrays are a potential tool for evaluation of renal function as described by Hsiao et al. (9). Use of DNA microarrays to measure RNA expression under various physiological conditions requires selection of DNA or gene products for the target array, as well as isolation and labeling of RNA from tissue samples. Protein arrays and antibody arrays are also beginning to be developed, and these are a potentially powerful means to evaluate the proteome. The Internet-based database that we describe should be useful in the design of the target array to test specific hypotheses; for example, to compare gene expression and protein regulation under varying physiological conditions (3). The goal is to facilitate development of arrays that contain gene products of interest against which mRNA and protein expression levels can be quantitatively assessed.

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://mrbl.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

Article published online before print. See web site for date of publication (<http://physiolgenomics.physiology.org>).

Address for reprint requests and other correspondence: R. Mejia, 12 South Drive, Rm. 4007, MSC 5621, National Institutes of Health, Bethesda, MD 20892-5621 (E-mail: ray@helix.nih.gov).

Collecting Duct Database(cddb)

This database is intended to serve as a learning tool to obtain curated information for the design of microarray targets to scan collecting duct tissues (human, rat, mouse).

The database focuses on regulatory and transporter 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

- [Adenyl Cyclases](#)
- [Adhesion Proteins and Collagens](#)
- [Anchor Scaffold Proteins](#)
- [Cyclic Nucleotide Phosphodiesterases](#)
- [Cytoskeletal Proteins and Linkers/Molecular Motors](#)
- [G-protein Coupled Receptors](#)
- [Guanylyl Cyclases and Nitric Oxide Synthases](#)
- [Heterotrimeric G-proteins](#)
- [Hormones and Autocoids](#)
- [Ion Pumps](#)
- [Metabolism Proteins](#)
- [Miscellaneous Proteins](#)
- [Not Yet Classified](#)
- [Nuclear Receptors](#)
- [Protein Degradation Pathways](#)
- [Protein Kinases, Serine/Threonine](#)
- [Protein Kinases, Tyrosine Kinase Receptors](#)
- [Protein Phosphatases](#)
- [Proteoglycans](#)
- [SNAREs and Related Proteins](#)
- [Signal Proteins](#)
- [Small GTP Binding Proteins and Related Proteins](#)
- [Steroid Metabolizing Enzymes](#)
- [Stress Proteins](#)
- [Tight Junction Proteins](#)
- [Transcription Factors](#)
- [Transporters and Channels](#)

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

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).

The database is searchable. Enter your term(s) in the box. The only boolean operator currently supported is AND. By default, all fields are searched. To omit a field from the search, deselect it from the search field box. Search results can be filtered. To exclude records without external links to a given resource, select it in the filter box.

Search Field

Abbreviations
Name
Alternate names

Filter

LocusLink(Human)
LocusLink(Rat)
LocusLink(Mouse)
Online Mendelian Inheritance in Man
GeneCards: encyclopedia for genes, proteins and diseases
Rat Genome Database
Mouse Genome Informatics

Search String

AQP

Search Clear

Review Board

Additional Resources

FAQ

We appreciate your feedback. If you have any comments or questions, please leave them using our [comment form](#). [Please read the disclaimer](#).

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

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.

Transporters and Channels

Abbreviations	Name	Alternate names	Links
AQP1	aquaporin-1	-	C P L I O G R M
AQP2	aquaporin-2	-	C P L I O G R M
AQP3	aquaporin-3	-	C P L I O G R M
AQP4	aquaporin-4	-	C P L I O G R M
AQP6	aquaporin-6	-	C P L I O G R M
AQP7	aquaporin-7	-	C P L I O G R M
AQP8	aquaporin-8	-	C P L I O G R M
AQP9	aquaporin-9	-	C P L I O G R M

Search Again:

Enter your term(s) in the box. The only boolean operator currently supported is AND. By default, all fields are searched. To omit a field from the search, deselect it from the search field box. Search results can be filtered. To exclude records without external links to a given resource, select it in the filter box.

Search Field

Abbreviations
Name
Alternate names

Filter

LocusLink(Human)
LocusLink(Rat)
LocusLink(Mouse)
Online Mendelian Inheritance in Man
GeneCards: encyclopedia for genes, proteins and diseases
Rat Genome Database
Mouse Genome Informatics

Search String

avp

Search Clear

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

G-protein Coupled Receptors

Abbreviations	Name	Alternate names	Links
AVPR1A	vasopressin receptor type 1a	-	C P L I O G R M
AVPR1B	vasopressin receptor type 1b	-	C P L I O G R M
AVPR2	vasopressin receptor type 2	-	C P L I O G R M

Search Again:

Enter your term(s) in the box. The only boolean operator currently supported is AND. By default, all fields are searched. To omit a field from the search, deselect it from the search field box. Search results can be filtered. To exclude records without external links to a given resource, select it in the filter box.

Search Field

Abbreviations
 Name
 Alternatenames

Filter

LocusLink(human)
 LocusLink(Rat)
 LocusLink(Mouse)
 Online Mendelian Inheritance in Man
 GeneCards: encyclopedia for genes, proteins and diseases
 Rat Genome Database
 Mouse Genome Informatics

Search String

Fig. 4. Result of a search for “avp”, shown in Fig. 3, retrieves data for vasopressin receptors.

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/cddb/admin>), including creation of new sections and additions and extensions of the database.

A preliminary report of the database was presented at the SIAM Annual Meeting in Rio Grande, PR, July 10–14, 2000.

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/cddb/misc/review.html>.

REFERENCES

1. Blake JA, Richardson JE, Bult CJ, Kadin JA, Eppig JT, and the Mouse Genome Database Group. The Mouse Ge-

nome Database (MGD): the model organism database for the laboratory mouse. *Nucleic Acids Res* 30: 113–115, 2002.

2. Brenner BM. *Brenner and Rector's The Kidney* (6th ed.). Philadelphia: Saunders, 2000.

3. Brooks HL, Ageloff S, Kwon TH, Brandt W, Terris JM, Seth A, Michea L, Nielsen S, Fenton R, and Knepper MA. cDNA array identification of genes regulated in rat renal medulla in response to vasopressin infusion. *Am J Physiol Renal Physiol* 284: F218–F228, 2003.

4. Burge CB. Chipping away at the transcriptome. *Nat Genet* 27: 232–234, 2001.

5. Eppig JT, Blake JA, Burkhardt DL, Goldsmith CW, Lutz CM, and Smith CL. Corraling conditional mutations: a unified resource for mouse phenotypes. *Genesis* 32: 63–65, 2002.

6. Flessner MF and Knepper MA. Ammonium transport in collecting ducts. *Miner Electrolyte Metab* 16: 299–307, 1990.

7. Fogel KF. *Open Source Development with CVS*. Scottsdale, AZ: The Coriolis Group, 1999.

8. Hamosh A, Scott AF, Amberger J, Valle D, and McKusick VA. Online Mendelian Inheritance in Man (OMIM). *Hum Mutat* 15: 57–61, 2000.

9. Hsiao LL, Stears RL, Hong RL, and Gullans SR. Prospective use of DNA microarrays for evaluating renal function and disease. *Curr Opin Nephrol Hypertens* 9: 253–258, 2000.

10. Hughes S and Zmievski A. *PHP Developer's Cookbook*. Indianapolis, IN: Sams, 2001.

11. Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu YT, Roskin KM, Schwartz M, Sugnet CW, Thomas DJ, Weber RJ, Haussler D, and Kent WJ. The UCSC Genome Browser Database. *Nucleic Acids Res* 31: 51–54, 2003.

12. Knepper MA. Long-term regulation of urinary concentrating capacity. *Am J Physiol Renal Physiol* 275: F332–F333, 1998.

13. Knepper MA and Star RA. The vasopressin-regulated urea transporter in renal inner medullary collecting duct. *Am J Physiol Renal Fluid Electrolyte Physiol* 259: F393–F401, 1990.

14. Lennon G, Auffray C, Polymeropoulos M, and Soares MB. I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. *Genomics* 33: 151–152, 1996.

15. MEDLINE. US National Library of Medicine, National Institutes of Health, Bethesda, MD [Online]. http://www.nlm.nih.gov/databases/databases_medline.html [22 March 2002].

16. Momjian B. *PostgreSQL: Introduction and Concepts*. Boston: Addison-Wesley, 2001.

17. Pruitt KD and Maglott DR. RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Res* 29: 137–140, 2001.

18. PubMed. National Center for Biotechnology Information, US National Library of Medicine, National Institutes of Health, Bethesda, MD [Online]. <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>.

19. Rebhan M, Chalifa-Caspi V, Prilusky J, and Lancet D. GeneCards: a novel functional genomics compendium with automated data mining and query reformulation support. *Bioinformatics* 14: 656–664, 1998.

20. Robert-Nicoud M, Flahaut M, Elalouf JM, Nicod M, Salinas M, Bens M, Doucet A, Wincker P, Artiguenave F, Horisberger JD, Vandewalle A, Rossier BC, and Firsov D. Transcriptome of a mouse kidney cortical collecting duct cell line: effects of aldosterone and vasopressin. *Proc Natl Acad Sci USA* 98: 2712–2716, 2001.

21. Strausberg RL, Feingold EA, Klausner RD, and Collins FS. The mammalian gene collection. *Science* 286: 455–457, 1999.

22. Streelman JT and Kocher TD. From phenotype to genotype. *Evol Dev* 2: 166–173, 2000.

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].

24. Twigger S, Lu J, Shimoyama M, Chen D, Pasko D, Long H, Ginster J, Chen CF, Nigam R, Kwitek A, Eppig J, Maltais L, Maglott D, Schuler G, Jacob H, and Tonellato PJ. Rat Genome Database (RGD): mapping disease onto the genome. *Nucleic Acids Res* 30: 125–128, 2002.

25. Virlon B, Cheval L, Buhler JM, Billon E, Doucet A, and Elalouf JM. Serial microanalysis of renal transcriptomes. *Proc Natl Acad Sci USA* 96: 15286–15289, 1999.