

Database

Open Access

VKCDB: Voltage-gated potassium channel database

Bin Li and Warren J Gallin*

Address: Department of Biological Sciences, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

Email: Bin Li - bli@ualberta.ca; Warren J Gallin* - wgallin@ualberta.ca

* Corresponding author

Published: 09 January 2004

Received: 01 September 2003

BMC Bioinformatics 2004, 5:3

Accepted: 09 January 2004

This article is available from: <http://www.biomedcentral.com/1471-2105/5/3>

© 2004 Li and Gallin; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: The family of voltage-gated potassium channels comprises a functionally diverse group of membrane proteins. They help maintain and regulate the potassium ion-based component of the membrane potential and are thus central to many critical physiological processes. VKCDB (Voltage-gated potassium [K] Channel DataBase) is a database of structural and functional data on these channels. It is designed as a resource for research on the molecular basis of voltage-gated potassium channel function.

Description: Voltage-gated potassium channel sequences were identified by using BLASTP to search GENBANK and SWISSPROT. Annotations for all voltage-gated potassium channels were selectively parsed and integrated into VKCDB. Electrophysiological and pharmacological data for the channels were collected from published journal articles. Transmembrane domain predictions by TMHMM and PHD are included for each VKCDB entry. Multiple sequence alignments of conserved domains of channels of the four Kv families and the KCNQ family are also included. Currently VKCDB contains 346 channel entries. It can be browsed and searched using a set of functionally relevant categories. Protein sequences can also be searched using a local BLAST engine.

Conclusions: VKCDB is a resource for comparative studies of voltage-gated potassium channels. The methods used to construct VKCDB are general; they can be used to create specialized databases for other protein families. VKCDB is accessible at <http://vkcdb.biology.ualberta.ca>.

Background

Many biologists focus their research on one or a few specific protein families. While comprehensive sequence databases, such as GENBANK [1], and generally annotated databases, such as SWISSPROT [2], are freely available, intensive studies on protein families are better supported by relatively small, focused databases that are developed for specific research needs. Our experience with voltage-gated potassium channel database (VKCDB) provides an example of such a small, targeted protein family database. The approaches that we used to create this database are generally applicable to building databases for functional studies of other protein families.

Voltage-gated potassium channels (VKCs) are intrinsic membrane proteins that respond to changes in the transmembrane electrical field by altering conformation and selectively allowing potassium ions to pass through the membrane [3]. This property is the basis for VKCs' roles in shaping action potentials in neurons and modulating the electrical activity of excitable membranes. Mutations in VKC genes can lead to severe diseases, such as long QT syndrome and epilepsy [4,5]. Thus, VKCs have been considered as possible targets for drug design [6].

VKCs constitute a structurally and functionally diverse protein family. At this writing, there are over two hundred

described members of this family from more than 35 organisms. VKC-related structural and functional data, particularly electrophysiological and pharmacological parameters, are distributed in dozens of databases and hundreds of journal articles. No single database contains structural and functional data for the various members of this large protein family. The application of comparative methods to the study of VKCs and other protein families depends on ready availability of both structural and functional data in an easily accessed database. Here we report a customized database of VKC-related data that was created using semi-automated collection and management. This relational database currently holds 346 VKC entries. Each entry contains sequences, motifs, references, hyperlinks to other databases, and other available structural information. We have also collected available electrophysiological and pharmacological parameters for VKCs from several hundred published articles. These types of data are not properties of most proteins, and are not contained in general protein databases.

Construction and Content

VKCDB was initially populated by performing a redundant set of searches of GENBANK for family members (Figure 1). GENBANK was first searched for protein sequences similar to the human Kv1.2 protein sequence [7] using BLASTP [8]. The top 200 hits were used to perform BLASTP searches against GENBANK and SWISS-PROT, yielding a comprehensive collection of VKCs. After collapsing all redundant BLASTP results, the top 319 non-redundant hits were collected; sequences with lower scores were not VKCs.

A Perl script was used to retrieve information on the 319 VKCs from GENBANK and SWISS-PROT and store it in a MySQL relational database. A schematic diagram of the ER model of VKCDB can be found at VKCDB website <http://vkcdb.biology.ualberta.ca/images/ermodel.gif>.

Data from redundant records in GENBANK and SWISS-PROT were combined into a single entry. VKCDB entries with very similar sequences were manually checked, and their sequences were compared and annotated as "possible isoforms" or "sequence conflicts". Records for splicing variants were cross-referenced. Conflicting sequences that were submitted by different authors were cross-indexed as sequence conflicts, unless sequence errors were indicated in the literature, in which case the most recently updated sequence was kept. Entries labelled as "unknown products" from large sequencing projects that had the characteristic sequence pattern of the voltage sensor (a lysine or arginine residue at every third position of the fourth transmembrane domain) were used as BLASTP queries and annotated as members of a specific family of VKC based on the annotation of most similar BLAST results.

Using literature citations from GENBANK and SWISS-PROT, we manually collected available electrophysiological and pharmacological data from published articles for each VKC entry. Conflicting data were all kept and hyperlinked to the references in PUBMED [9].

All sequences were submitted to the TMHMM [10] and PHD [11] servers for secondary structure prediction. Results from both analyses were parsed and combined into a single annotated sequence figure. This information is currently stored as a graphic because both of the prediction programs are not sufficiently accurate to be taken as a definitive result.

Sequences belonging to each of the four Kv families and the KCNQ family [12] were extracted and a multiple alignment for each family was generated with ClustalW [13]. Alignments of the highly conserved regions (the T1 domain and the six transmembrane domains) were manually adjusted and included in VKCDB. Subsets of the aligned sequences can be selected and exported in FASTA format.

VKCDB is updated monthly. The "last modified" date of each VKCDB entry is compared to the corresponding field in the cognate GENBANK and SWISS-PROT entries. Information from any of the archival entries that has been changed since the last VKCDB update is then parsed and used to update VKCDB. New VKC entries are collected by performing a BLAST search of GENBANK and SWISS-PROT with all entries in VKCDB. The hits are combined into a non-redundant list for each subfamily and the top twenty scores on that non-redundant list that are not already entered in VKCDB are manually checked to confirm that they are indeed VKCs before adding them to VKCDB. If all twenty hits are VKCs, the next twenty hits are also manually evaluated; this process is repeated until non-VKC entries are found. Current entries in VKCDB were updated with SWISS-PROT Release 42.1 (October 2003) and GENBANK entries as of October 2003.

Utility and Discussion

At present VKCDB contains 346 VKC entries from 35 organisms. 39 VKCs are annotated as having between two and nine different isoforms, although some of these might be due to cloning or annotation artefacts. VKCDB can be browsed and searched through a web interface by several criteria, including VKC Kv subfamilies [12], organism names, GENBANK Protein ID, protein description, reference information, and electrophysiological parameters. The VKC entries in all search and browse results can be individually selected on the web page to produce a batch sequence file in FASTA format for use in other applications. This site also implements a local BLAST server for similarity searches against VKCDB entries.

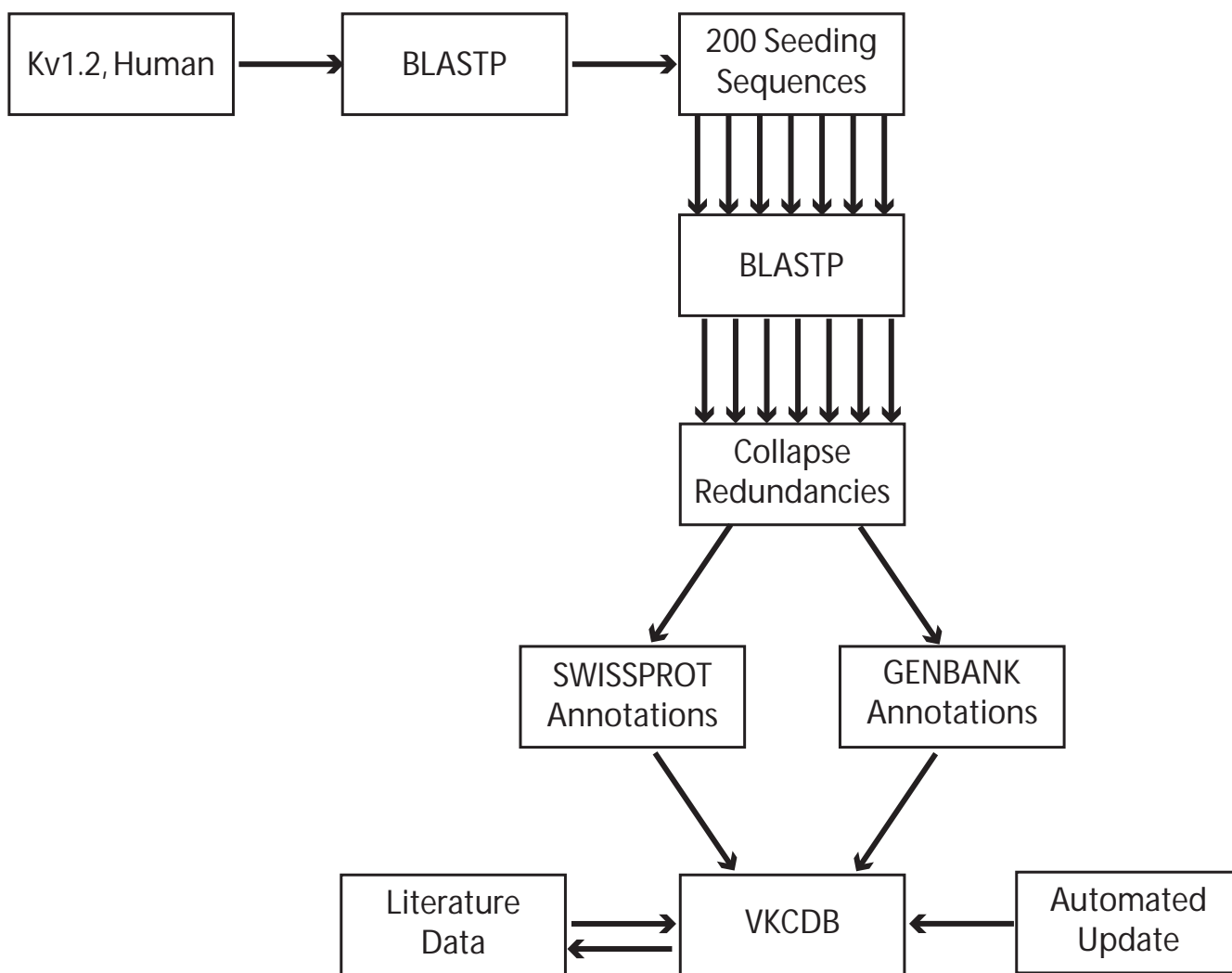


Figure 1
Populating VKCDB. VKCDB was populated by searching against GENBANK and SWISSPROT databases with 200 seeding VKC protein sequences. The redundancies in these results were collapsed, then structural and functional information was extracted from different databases and published articles using a combination of automated scripts and manual selection.

Each VKC entry page contains information such as protein accession number, protein name, protein sequence, coding gene name and accession number, SWISSPROT function description, references and hyperlinks to other biological databases. On each entry page, a button labelled "Electrophysiology" opens a pop-up window containing electrophysiological parameters, pharmacological data, and related references hyperlinked to PUBMED (Figure 2). We will be adding data on synthetic VKC mutants to VKCDB in the near future, including links to the cognate wild type protein and the electrophysiological and pharmacological data. There is also a link to

transmembrane helix predictions by TMHMM [10] and PHD [11] on each entry page.

Multiple alignments of conserved regions of the four Kv families and of the KCNQ family are available on the VKCDB web site [12], on the tools page <http://vkcdb.biol.ualberta.ca/alignment.html>. The individual sequences can be selected and downloaded, with gaps in place, in FASTA format for use in other applications.

The VKCDB web site includes a "submit" page that allows us to communicate with users on annotation errors, missing entries, and other information so that we can

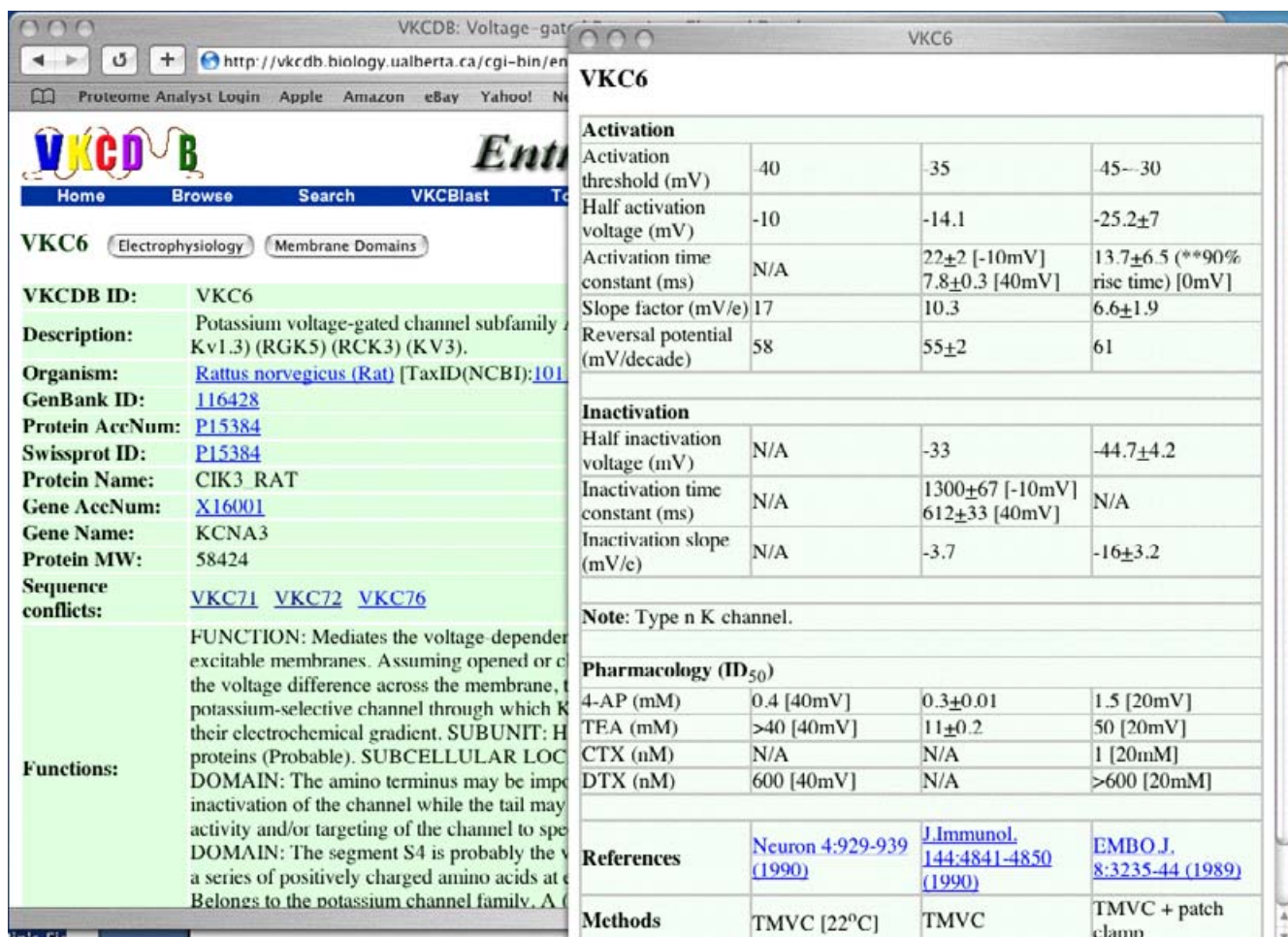


Figure 2
Screen dump of the entry page of a VKCDB entry. The popup window contains the electrophysiological parameters of this entry. The transmembrane helix prediction by TMHMM and PHD can also be displayed for each VKCDB entry. Content of the entry page is extensively hyperlinked to various databases.

maintain accurate and updated VKC information in our database.

Conclusions

VKCDB contains structural and functional data and related multiple alignments for voltage-gated potassium channels in a single database. The VKCDB web page is designed to provide easy access and searching through a user-friendly interface. It is also designed to interact easily with tools that we are developing to study the structure-functional relationship in VKCs using machine-learning approaches. The database information is also available as an XML file for users who wish to implement customized configurations.

Similar approaches can be taken to construct specific, small-to-medium-sized protein family databases, with minimum knowledge of Perl and MySQL database management. As a small, customized protein family database, VKCDB is a useful and convenient resource for research on VKCs. As our understanding of VKCs increases, more annotations and applications will be added to enrich VKCDB so that it can continue to serve as a main resource for structural and functional studies of VKCs.

Availability and Requirements

VKCDB is freely accessible at <http://vkcdb.biology.ualberta.ca>. A snapshot of VKCDB (as of November 2, 2003 at this writing) in XML format can be freely downloaded from the website of VKCDB.

Authors' contributions

BL carried out the design of the database schema and the building of the MySQL database. He built the web interface and related CGIs, and wrote code to populate, implement and update VKCDB, manually checked each VKC entry and collected the electrophysiological and other functional data from journal articles. WJG conceived of the study, and participated in the design, building, data checking and analyses.

Acknowledgments

Authors would like to thank Haiyan Zhang and Dr. David Wishart for their technical support and system maintenance during the early stages of VKCDB development. Development of VKCDB was supported by a grant from CIHR (MOP-62685) to Andrew N. Spencer and WJG.

References

1. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL: **GenBank**. *Nucleic Acids Res* 2003, **31**:23-27.
2. Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M: **The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003**. *Nucleic Acids Res* 2003, **31**:365-370.
3. Yellen G: **The voltage-gated potassium channels and their relatives**. *Nature* 2002, **419**:35-42.
4. Kaneko S, Okada M, Iwasa H, Yamakawa K, Hirose S: **Genetics of epilepsy: current status and perspectives**. *Neurosci Res* 2002, **44**:11-30.
5. Towbin JA, Vatta M: **Molecular biology and the prolonged QT syndromes**. *Am J Med* 2001, **110**:385-398.
6. Cooper EC: **Potassium channels: how genetic studies of epileptic syndromes open paths to new therapeutic targets and drugs**. *Epilepsia* 2001, **42 Suppl 5**:49-54.
7. Ramaswami M, Gautam M, Kamb A, Rudy B, Tanouye MA, Mathew MK: **Human potassium channel genes: molecular cloning and functional expression**. *Mol. Cell. Neurosci.* 1990, **1**:214-223.
8. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs**. *Nucleic Acids Res* 1997, **25**:3389-3402.
9. McEntyre J, Lipman D: **PubMed: bridging the information gap**. *CMAJ* 2001, **164**:1317-1319.
10. Krogh A, Larsson B, von Heijne G, Sonnhammer EL: **Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes**. *J Mol Biol* 2001, **305**:567-580.
11. Rost B, Sander C: **Combining evolutionary information and neural networks to predict protein secondary structure**. *Proteins* 1994, **19**:55-72.
12. Chandy KG: **Simplified gene nomenclature**. *Nature* 1991, **352**:26.
13. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice**. *Nucleic Acids Res* 1994, **22**:4673-4680.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

