Developing Rigorous Sequence Profiles of Voltage-Gated Potassium Channels; Applications in Functional Characterisation, Deleterious Variant Prediction, Domain Allocation and Comparative Modeling

Lucy F. Stead, Ian C. Wood and David R. Westhead

University of Leeds.
Faculty of Biological Sciences

Abstract. Mutations within voltage-gated potassium (Kv) channels are responsible for several inherited disorders including cardiac arrhythmogenesis and some forms of epilepsy. Early detection and treatment of these diseases will be aided by further characterisation of Kv channels at the level of both gene and protein. The latter requires identification of the key residues involved in protein function. We have created robust sequence profiles, in hidden Markov model format, of the key structural elements of Kv channels using multiple sequence alignments of 937 proteins from 21 mammals. There are several applications for these profiles. Here we report on the application of information theoretic analysis to highlight key positions within each structural element. This analysis highlights positions that are universally informative within the Kv channel family; a dataset that includes all known functionally important residues plus some additional sites that we propose are worthy of further experimental investigation. It also highlights residues that we hypothesise could be important for the electrophysiological differences observed between Kv channel subfamilies and are, thus, also worth further experimentation. Further applications of the rigourous statistical sequence profiles that we have created include use in deleterious SNP (single nucleotide polymorphism) prediction methods, prediction of structural segment location within new Kv sequences and improved alignments for comparative modeling of unknown Kv structures.

1 Introduction

Voltage-gated potassium (Kv) channels are involved in a number of important physiological roles including the generation of the heartbeat and transmission of signals through the central nervous system. Mutations in genes encoding Kv channels are associated with diseases such as epilepsy [1] and cardiac arrhythmogenesis, the latter of which sometimes occurs in apparently healthy patients only upon administration of certain drugs[2].

Kv channel proteins are clustered into 13 homologous subfamilies. Channels form via homo (all subunits the same) or hetero (subunits of different types) tetramerisation of individual protein subunits. This diversity allows a large range of channels with different electrophysiological properties to form. The common topology for all Kv channel subunits is 6 transmembrane (TM) helices; S1 to S6 (Fig. 1). S1 to S4 constitute the voltage sensing domain and S5 and S6 constitute the pore domain[3].

The way that a change in membrane voltage is transmitted from the voltage sensing domain to the pore to regulate the flow of K^+ is currently the subject of much debate[4][5]. The role of each TM helix, and the association of different subunits into channels with a range of properties, is also still unclear. Furthering our understanding of these channels is important because of the delicate balance of electrical signals required for action potential generation and propagation; disruption of which can lead to disease phenotypes with potentially fatal consequences.

Kv channels have been the subject of much experimental study both of the functional effects of artificial mutants and the disease relationships of natural variants (for instance we have gathered a set of more than 1000 known functionally characterised sequence variants [6]). We desired to build robust descriptors of the evolution of Kv channels in general and within specific subfamilies to enable understanding of this large data set and the development of methods able to predict functional effects of further natural and artificial variants. Here we present the development of sequence patterns of evolutionary conservation in hidden Markov model format and present a preliminary information theoretic analysis that strongly supports this planned future application and interactions with experimental workers to test predictions. Our analysis highlights known functional residues and also indicates additional residues that may be functionally important in all Kv channels, or responsible for subtle differences in characteristics between subfamilies. Although mutations that severely compromise function have been relatively easy to identify and assess, mutations that produce only subtle alterations in channel activity are more problematic to identify. Such mutations are nonetheless likely to contribute to physiological differences between individuals within the population and potentially predispose individuals to particular diseases. Being able to predict such mutations would be a major step forward in understanding ion channel physiology.

Other applications of our sequence patterns are in the analysis of new Kv channel sequences (for instance to predict TM segments where they significantly outperform generic predictors) and in improvements to alignments for comparative modeling studies.

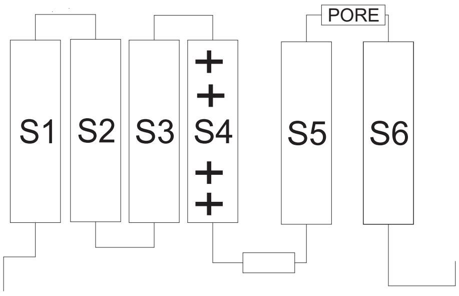


Fig. 1. Transmembrane (TM) Segment Topology in Kv Channels. Six TM helices, denoted S1 to S6, traverse the lipid bilayer. S4 has several positive residues dotted along its length which are responsible for sensing voltage changes across the membrane.

2 Methods

2.1 Multiple Sequence Alignment Generation

A non-redundant set of 937 Kv sequences from 21 different mammals were extracted from Ensembl 47[7] and Interpro 16.1[8] using the gene ontology term GO:0008076 (voltage-gated potassium channel complex). This was checked against both the HUGO[9] and IUPHAR[10] nomenclature. Multiple sequence alignments (MSAs) of each subfamily were produced using MUSCLE[11], and manually curated in Jalview[12] with guidance from previously published alignments and known structures[14][15][16]. The location of the 6 TM helices were assigned by comparative modeling the consensus sequence of each MSA with a template structure (PDB code: 2r9r [16]) of a rat KCNA family chimera, and by making TM predictions for each sequence using HMMTOP2[17][18]. Once assigned, the alignment within each TM segment was re-checked before being extracted; creating 6 TM alignments per subfamily. All the sequences for each TM were also amalgamated and re-aligned to form a final alignment per TM that included all the Kv sequences.

2.2 Emission Probability Distribution Extraction

Application of information theory to MSAs requires the formation of amino acid probability distributions at each position. These can be created using count data extracted from the MSAs themselves, but this will not account for bias created by entries with high sequence identity or in MSAs where there are fewer sequences with which to create the distribution. To overcome these biases we created, and calibrated, hidden Markov models (HMMs) of our MSAs using the HMMER software suite[19]. We then extracted the emission probability of each amino acid at each position. These probabilities are weighted using a method described in [20] to account for highly similar sequences, and using Dirichlet priors to account for alignments containing fewer sequences [21]. The HMM output is actually scores for each amino acid in each position but these can be converted into emission probabilities as described in [22]. This was done using perl scripts.

2.3 Shannon Entropy and Kullback-Leibler Divergence Calculations

Shannon entropy (H)[23] relates to the amount of uncertainty associated with a random variable or, in this case, associated with the amino acid, x$_i$, at position, X, within a TM segment.

$$H(X) = -\sum_{i=1}^{20} p(x_i)\ log_2 p(x_i)\ . \tag{1}$$

H quantifies the degree of variability in each emission distribution: uniform distributions have the highest entropy, while distributions in which all amino acid probabilities are zero except for one have zero entropy. The entropy is the information gain associated with a single message from the distribution.

We calculated the Shannon entropy for every position in each MSA using emission probabilities extracted from our HMMs. We also wished to compare subfamily MSAs with MSAs of all Kv sequences. To do this we calculated the Kullback-Leibler divergence (D)[24] at each position, X. This quantifies the difference between a subfamily's probability distribution (SS) and that of all Kv sequences (AS), and is calculated using equation 2. D can be interpreted as the amount of additional information gained when using SS instead of AS. Entropy and divergence calculations were performed using the R statistical software package[31].

$$D(SS||AS) = \sum_{i=1}^{20} SSp(x_i)\ log_2 \frac{SSp(x_i)}{ASp(x_i)}\ . \tag{2}$$

2.4 Information Gain Comparisons

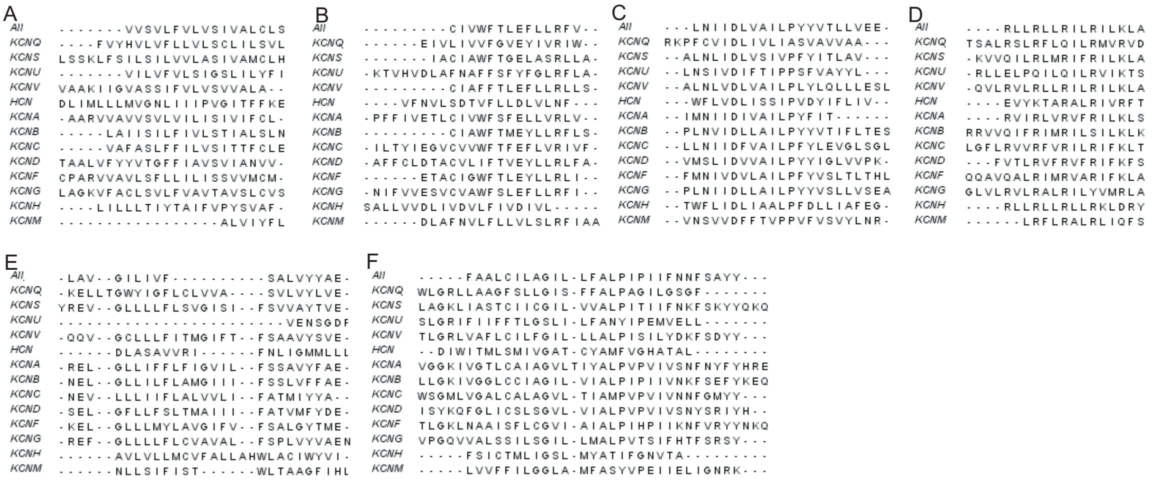
The MSAs for each TM varied in length, hindering direct comparison of entropy and divergence per position. Hence, the consensus sequence for each HMM (acquired via a HMMER tool; hmmemit[19]) were aligned using MUSCLE and the entropy and divergence were compared across the alignment. Positions with low entropy and low divergence are hypothesised to be universally important for Kv channel function; these are highly conserved across all Kv channels and invariant between subfamilies. We thus we collated data on TM positions with D<0.5 and

H<2.5. Subfamily-specific positions are hypothesised to have low entropy and high divergence; these are conserved within sub-families but differ between them. We thus collated data on TM positions with D>1.5 and H<2.5. Conversely it is hypothesised that positions that are neither Kv-channel nor subfamily specific will have high entropy and low divergence. We thus collated data on TM positions with D<0.5 and H>2.5. We deemed a position truly subfamily-specific if it appeared for a single subfamily in the former dataset and 3 or more subfamilies in the latter; indicating that the information gain is exclusive to a single subfamily. Threshold values of D and H were chosen arbitrarily upon inspecting plots of these values for each TM for all subfamilies (data not shown).

3 Results

3.1 Multiple Sequence Alignments

84 multiple sequence alignments (MSAs) were created initially; one for each of 6 transmembrane (TM) helices per 13 subfamilies and one containing all Kv channel sequences for that TM. These were used to build hidden Markov models (HMMs) from which consensus sequences were emitted and used to create a further 6 alignments; 1 per TM helix. These are shown in Fig. 2; revealing the positions that were aligned and, thus, had their Shannon entropy (H) and Kullback-Leibler divergence (D) calculated.



```
A
All   -------VVSVLFVLVSIVALCLS
KCNQ  ----FVYHVLVFLLVLSCLILSVL
KCNS  LSSKLFSILSILVVLASIVAMCLH
KCNU  -------VILVFVLSIGSLILYFI
KCNV  VAAKIIGVASSIFVLVSVVALA--
HCN   DLIMLLLMVGNLIIIPVGITFFKE
KCNA  -AARVVAVVSVLVILISIVIFCL-
KCNB  -----LAIISILFIVLSTIALSLN
KCNC  -----VAFASLFFILVSITTFCLE
KCND  TAALVFYYVTGFFIAVSVIANVV-
KCNF  CPARVVAVLSFLLILISSVVMCM-
KCNG  LAGKVFACLSVLFVAVTAVSLCVS
KCNH  ----LILLLTIYTAIFVPYSVAF-
KCNM  -----------------ALVIYFL
```

```
B
All   ---------CIVWFTLEFLLRFV--
KCNQ  ------EIVLIVVFGVEYIVRIW--
KCNS  --------IACIAWFTGELASRLLA-
KCNU  -KTVHVDLAFNAFFSFYFGLRFLA-
KCNV  ---------CIAFFTLEFLLRLLS-
HCN   ----VFNVLSDTVFLLDLVLNF---
KCNA  -PFFIVETLCIVWFSFELLVRLV--
KCNB  ---------CIAWFTMEYLLRFLS-
KCNC  -ILTYIEGVCVVWFTFEFLVRIVF-
KCND  -AFFCLDTACVLIFTVEYLLRLFA-
KCNF  ------ETACIGWFTLEYLLRLI--
KCNG  -NIFVVESVCVAWFSLEFLLRFI--
KCNH  SALLVVDLIVDVLFIVDIVL-----
KCNM  ------DLAFNVLFLLVLSLRFIAA
```

```
C
All   ---LNIIDLVAILPYYVTLLVEE-
KCNQ  RKPFCVIDLIVLIASVAVVAA---
KCNS  --ALNLIDLVSIVPFYITLAV---
KCNU  --LNSIVDIFTIPPSFVAYYL---
KCNV  --ALNLVDLVAILPLYLQLLLESL
HCN   ---WFLVDLISSIPVDYIFLIV--
KCNA  --IMNIIDIVAILPYFIT------
KCNB  --PLNVIDLLAILPYYVTIFLTES
KCNC  --LLNIIDFVAILPFYLEVGLSGL
KCND  --VMSLIDVVAILPYYIGLVVPK-
KCNF  --FMNIVDVLAILPFYVSLTLTHL
KCNG  --PLNIIDLLAILPYYVSLLVSEA
KCNH  --TWFLIDLIAALPFDLLIAFEG-
KCNM  --VNSVVDFFTVPPVFVSVYLNR-
```

```
D
All   ----RLLRLLRILRILKLA
KCNQ  TSALRSLRFLQILRMVRVD
KCNS  -KVVQILRLMRIFRILKLA
KCNU  -RLLELPQILQILRVIKTS
KCNV  -QVLRVLRLLRILRILKLA
HCN   ----EVYKTARALRIVRFT
KCNA  ----RVIRLVRVFRILKLS
KCNB  RRVVQIFRIMRILSILKLK
KCNC  LGFLRVVRFVRILRIFKLT
KCND  ---FVTLRVFRVFRIFKFS
KCNF  QQAVQALRIMRVARIFKLA
KCNG  GLVLRVLRALRILYVMRLA
KCNH  ----RLLRLLRLLRKLDRY
KCNM  ------LRFLRALRLIQFS
```

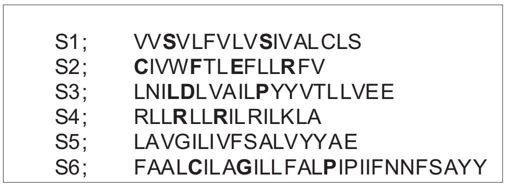
```
E
All   -LAV--GILIVF----------SALVYYAE-
KCNQ  -KELLTGWYIGFLCLVVA----SVLVYLVE-
KCNS  YREV--GLLLLFLSVGISI--FSVVAYTVE-
KCNU  ------------------------VENSGDF
KCNV  -QQV--GCLLLFITMGIFT--FSAAVYSVE-
HCN   ------DLASAVVRI------FNLIGMMLLL
KCNA  -REL--GLLIFFLFIGVIL--FSSAVYFAE-
KCNB  -NEL--GLLILFLAMGIII--FSSLVFFAE-
KCNC  -NEV--LLLIIFLALVVLI--FATMIYYA--
KCND  -SEL--GFLLFSLTMAIII--FATVMFYDE-
KCNF  -KEL--GLLLMYLAVGIFV--FSALGYTME-
KCNG  -REF--GLLLLFLCVAVAL--FSPLVYVAEN
KCNH  ------AVLVLLMCVFALLAHWLACIWYVI-
KCNM  ------NLLSIFIST------WLTAAGFIHL
```

```
F
All   -----FAALCILAGIL-LFALPIPIIFNNFSAYY---
KCNQ  WLGRLLAAGFSLLGIS-FFALPAGILGSGF-------
KCNS  LAGKLIASTCIICGIL-VVALPITIIFNKFSKYYQKQ
KCNU  SLGRIFIIFFTLGSLI-LFANYIPEMVELL-------
KCNV  TLGRLVAFLCILFGIL-LLALPISILYDKFSDYY---
HCN   --DIWITMLSMIVGAT-CYAMFVGHATAL--------
KCNA  VGGKIVGTLCAIAGVLTIYALPVPVIVSNFNYFYHRE
KCNB  LLGKIVGGLCCIAGIL-VIALPIPIIVNKFSEFYKEQ
KCNC  WSGMLVGALCALAGVL-TIAMPVPVIVNNFGMYY---
KCND  ISYKQFGLICSLSGVL-VIALPVPVIVSNYSRIYH--
KCNF  TLGKLNAAISFLCGVI-AIALPIHPIIKNFVRYYNKQ
KCNG  VPGQVVALSSILSGIL-LMALPVTSIFHTFSRSY---
KCNH  -----FSICTMLIGSL-MYATIFGNVTA---------
KCNM  -----LVVFFILGGLA-MFASYVPEIIELIGNRK---
```

Fig. 2. Kv Channel and Subfamily-Specific Consensus Sequence Alignments. Consensus sequences for Kv channels (labelled 'All') and per subfamily for S1 through to S6 (parts A to F respectively), which have been automatically aligned and manually curated.

3.2 Universally Informative TM Positions

TM helix positions that had entropy, H<2.5 and divergence, D<0.5 are shown in Fig. 3, where they are highlighted on the Kv channel consensus sequence for each TM helix.

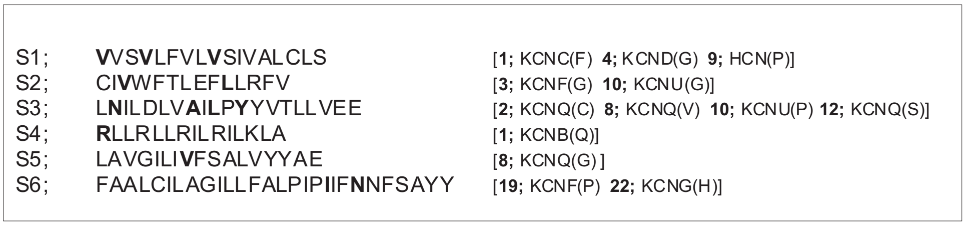


S1;	VV**S**VLFVLV**S**IVALCLS
S2;	**C**IVW**F**TL**E**FLL**R**FV
S3;	LNI**LD**LVAIL**P**YYVTLLVEE
S4;	RLL**R**LL**R**ILRILKLA
S5;	LAVGILIVFSALVYYAE
S6;	FAAL**C**ILA**G**ILLFAL**P**IPIIFNNFSAYY

Fig. 3. Universally Informative Transmembrane (TM) Positions. Positions within each TM segment (S1 to S6) that have low entropy (H) and low divergence (D). These positions are highlighted in bold on the Kv channel consensus sequence for each TM helix.

3.3 TM Positions Informative for Specific Subfamilies

TM positions with H<2.5 and D>1.5 in one subfamily but H>2.5 and D<0.5 in at least 3 others are shown in Fig. 4; highlighted on the Kv channel consensus sequence for each TM helix. The subfamily that is specific to each position is indicated adjacent to the consensus sequence with the corresponding residue that appears in that subfamilys consensus in parentheses (Fig. 4).



S1;	**V**VS**V**LFVL**V**SIVALCLS	[**1;** KCNC(F) **4;** KCND(G) **9;** HCN(P)]
S2;	CI**V**WFTLEF**L**LRFV	[**3;** KCNF(G) **10;** KCNU(G)]
S3;	L**N**ILDLV**A**I**L**P**Y**YVTLLVEE	[**2;** KCNQ(C) **8;** KCNQ(V) **10;** KCNU(P) **12;** KCNQ(S)]
S4;	**R**LLRLLRILRILKLA	[**1;** KCNB(Q)]
S5;	LAVGILI**V**FSALVYYAE	[**8;** KCNQ(G)]
S6;	FAALCILAGILLFALPIP**I**IF**N**NFSAYY	[**19;** KCNF(P) **22;** KCNG(H)]

Fig. 4. Subfamily-Specific Transmembrane (TM) Positions. Positions that have low entropy (H) and high divergence (D) in one subfamily but high entropy and low divergence in others are highlighted in bold on the Kv channel consensus sequence for each TM. The subfamily for which these residues are specifically informative is noted, per residue, on the right hand side with the corresponding residue in said subfamily's consensus sequence in parentheses.

4 Discussion

We have produced high quality multiple sequence alignments (MSA) containing evolutionary information for Kv channel proteins in 21 mammals. These have been used to build statistically rigorous sequence profiles for the main structural elements common to all Kv proteins using hidden Markov models (HMMs). A number of applications exist for these robust profiles;

- Prediction, and allocation, of each domain within new Kv sequences to a higher degree of accuracy than generic predictors.
- Features with which to train machine learning models designed to predict deleterious variants[25].
- Improved alignments for comparative modeling.
- Information theoretic analysis to determine positions within each profile from which most information can be gained.

This paper reports preliminary findings in the latter application in an attempt to identify functionally important residues for Kv sequences in general, and within specific subfamilies. Previous attempts to functionally characterise Kv channels using *in silico* methods include application of machine learning to identify the residues responsible for activation voltage [26], and statistical analysis of MSAs to identify residues that are co-evolved between the voltage sensor and pore[27]. Ours is the first approach, however, to focus on the differences between the Kv subfamilies. These differences will be more subtle and, thus, harder to identify. This is compounded by the reduced number of sequences of a particular subfamily type compared with those within the Kv superfamily in general.

The majority of residues highlighted in our method as being universally informative (Fig. 3) are known functional residues:

- In S2; the Phe is implicated in stabilisation of gating charges as they traverse the membrane. The Glu and Arg are part of a salt bridge network that stabilise charged residues within the lipid environment[16].
- In S3; the Asp is part of the aforementioned salt bridge network. The Pro effectively breaks S3 into 2 parts, allowing rotation of one with respect to the other[16].
- In S4; the R residues sense a change in membrane voltage via electrostatic interactions[28][29].
- In S6; the Gly hinge is required for channel gating. The Pro is part of a PX[P or G] motif that produces a bend at the point where all the S6 helices cross[3].

The only functionally important residues not highlighted in our analysis are additional positively charged residues in S4. However, the gating charge is formed by motion of 3-3.5 positive residues per subunit[30] and there are 5 potential positions where these can reside within S4 (Fig. 2D). Hence, it may be that the 2 positions highlighted in our analysis are universal and the additional charge can be sited at any of the other potential positions. It is interesting to note that our analysis highlights several residues that have not yet been implicated as functionally important and, thus, worth further validation:

- In S1; the 2 Ser residues may be an important part of the recently suggested interface between S1 and the pore helix[27], or may further aid charge stabilisation within the voltage sensor domain as indicated by Jiang et al[13].
- In S2; a Cys is implicated that may be important for structural stabilisation.
- In S3; a Leu that appears to be the key part of a leucine zipper type motif
- In S6; a Cys which, again, could be structurally relevant.

Our analysis has further highlighted residues that may be functionally, or structurally, relevant in a specific subfamily. This is based upon the observation that Kv channels form with a range of electrophysiological characteristics, such that each subunit must contribute to the overall channel properties. The residues that were highlighted are indicated in Fig. 4. It is worth noting that the residue most often present in the subfamily for which each position is deemed important (annotated in parentheses in Fig. 4), is often functionally interesting. That is, apart from position 8 in S3 where in KCNQ proteins a Val replaces an Ala, each substitution indicates a role for the new amino acid; glycines offer greater flexibility, prolines break helices, charges are conserved but are more or less delocalised. Unfortunately a search of over 1000 artificially-created, functionally-characterised mutations did not yield any validation data for our highlighted residues, though it also did not provide any to disprove our method. This is due to a bias towards functional characterisation of specific subfamilies where members are known to be heavily disease-associated. Design and execution of further artificial mutagenesis experiments is required to validate our findings.

5 Conclusion

We have created statistically rigorous profiles of voltage-gated potassium (Kv) channel transmembrane (TM) segments, in hidden Markov format, using multiple sequence alignments which have been carefully manually curated. These profiles constitute evolutionary descriptors of Kv channel sequences in general as well as those of each Kv subfamily. There are several important applications for these profiles, including use as a feature with which to train methods to predict disease-causing natural sequence variants (the outcome of single nucleotide polymorphisms; SNPs) to aid genotype-phenotype characterisation of Kv channels. The application reported here, however, is use of information theoretic analysis to identify positions from which most information is gained per profile. This analysis highlights universally informative TM positions within all Kv channels; a set that includes all known functional residues and some further residues that we suggest are now worthy of further experimental characterisation. This analysis also highlights TM positions that are only informative for a specific subfamily. We hypothesise that these positions may be functionally important for defining the electrophysiological properties for the subfamily in question but have been able neither to validate nor refute this hypothesis based on the body of experimental data we have extracted from the literature. We, therefore, suggest that new experiments be designed and executed with the focus of probing these potentially subfamily-specific residues.

References

1. Noebels, J. L.: The Biology of Epilepsy Genes. Annual Review of Neuroscience. 26, 599-625 (2003)
2. Anantharam, A., Markowitz, S. M., and Abbott, G. W.: Pharmacogenetic Considerations in Diseases of Cardiac Ion Channels. J Pharmacol Exp Ther. 307, 831-838 (2003)
3. Doyle, D. A., Cabral, J. M., Pfuetzner, R. A., Kuo, A. L., Gulbis, J. M., Cohen, S. L., Chait, B. T., and MacKinnon, R.: The structure of the potassium channel: Molecular basis of K^+ conduction and selectivity. Science. 280, 69-77 (1998)
4. Ahern, C. A., and Horn, R.: Stirring up controversy with a voltage sensor paddle. Trends In Neurosciences. 27, 303-307 (2004)
5. Tombola, F., Pathak, M. M., and Isacoff, E. Y.: How does voltage open an ion channel? Annual Review Of Cell And Developmental Biology. 22, 23-52 (2006)
6. Stead, L.F, Wood, I.C, Westhead, D.R.: Identifying Disease-Causing Variation in Voltage-Gated Potassium Channel Genes. Manuscript in preparation.
7. Hubbard, T. J. P., Aken, B. L., Beal, K., Ballester, B., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cunningham, F., Cutts, T., Down, T., Dyer, S. C., Fitzgerald, S., Fernandez-Banet, J., Graf, S., Haider, S., Hammond, M., Herrero, J., Holland, R., Howe, K., Howe, K., Johnson, N., Kahari, A., Keefe, D., Kokocinski, F., Kulesha, E., Lawson, D., Longden, I., Melsopp, C., Megy, K., Meidl, P., Ouverdin, B., Parker, A., Prlic, A., Rice, S., Rios, D., Schuster, M., Sealy, I., Severin, J., Slater, G., Smedley, D., Spudich, G., Trevanion, S., Vilella, A., Vogel, J., White, S., Wood, M., Cox, T., Curwen, V., Durbin, R., Fernandez-Suarez, X. M., Flicek, P., Kasprzyk, A., Proctor, G., Searle, S., Smith, J., Ureta-Vidal, A., and Birney, E.: Ensembl 2007. Nucl. Acids Res. 35, D610-617 (2007)
8. Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Buillard, V., Cerutti, L., Copley, R., Courcelle, E., Das, U., Daugherty, L., Dibley, M., Finn, R., Fleischmann, W., Gough, J., Haft, D., Hulo, N., Hunter, S., Kahn, D., Kanapin, A., Kejariwal, A., Labarga, A., Langendijk-Genevaux, P. S., Lonsdale, D., Lopez, R., Letunic, I., Madera, M., Maslen, J., McAnulla, C., McDowall, J., Mistry, J., Mitchell, A., Nikolskaya, A. N., Orchard, S., Orengo, C., Petryszak, R., Selengut, J. D., Sigrist, C. J. A., Thomas, P. D., Valentin, F., Wilson, D., Wu, C. H., and Yeats, C.: New developments in the InterPro database. Nucl. Acids Res. 35, D224-228 (2007)
9. Human Genome Organisation `http://www.hugo-international.org/`
10. Wei, A. D., Gutman, G. A., Aldrich, R., Chandy, K. G., Grissmer, S., and Wulff, H.: International Union of Pharmacology. LII. Nomenclature and Molecular Relationships of Calcium-Activated Potassium Channels. Pharmacol Rev. 57, 463-472 (2005)
11. Edgar, R. C.: MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res. 32, 1792-1797 (2004)
12. Clamp, M., Cuff, J., Searle, S. M., and Barton, G. J.: The Jalview Java alignment editor. Bioinformatics. 20, 426-427 (2004)

13. Jiang, Y. X., Ruta, V., Chen, J. Y., Lee, A., and MacKinnon, R.: The principle of gating charge movement in a voltage-dependent K^+ channel. Nature. 423, 42-48. (2003)
14. Jiang, Y. X., Lee, A., Chen, J. Y., Ruta, V., Cadene, M., Chait, B. T., and MacKinnon, R.: X-ray structure of a voltage-dependent K^+ channel. Nature. 423, 33-41 (2003)
15. Lee, S.-Y., Lee, A., Chen, J., and MacKinnon, R.: Structure of the KvAP voltage-dependent K^+ channel and its dependence on the lipid membrane. Proc Natl Acad Sci U S A. 102, 15441-15446 (2005)
16. Long, S. B., Tao, X., Campbell, E. B., and MacKinnon, R.: Atomic structure of a voltage-dependent K^+ channel in a lipid membrane-like environment. Nature. 450, 376-382 (2007)
17. Tusnady, G. E., and Simon, I.: The HMMTOP transmembrane topology prediction server. Bioinformatics. 17, 849-850 (2001)
18. Punta, M., Forrest, L. R., Bigelow, H., Kernytsky, A., Liu, J. F., and Rost, B.: Membrane protein prediction methods. Methods. 41, 460-474 (2007)
19. Eddy, S. R.: Profile hidden Markov models. Bioinformatics. 14, 755-763 (1998)
20. Gerstein, M., Sonnhammer, E. L. L., and Chothia, C.: Volume changes in protein evolution, J.Mol.Biol. 236, 1067-78 (1994)
21. Brown, M., Hughey, R., Krogh, A., Mian, I. S., Sjlander, K., and Haussler, D.: Using Dirichlet Mixture Priors to Derive Hidden Markov Models for Protein Families. In: Hunter, L., Searls, D. B., and Shavlik, J. W., eds. 1st international Conference on intelligent Systems For Molecular Biology. AAAI Press (1993)
22. HMMER User's Guide `http://www.csb.yale.edu/userguides/seq/hmmer/docs/index.html`
23. Shannon, C. E.: Prediction and entropy of printed english. Bell Systems Technical Journal. 30, 50-64 (1951)
24. Kullback, S., Leibler, A.: On the information and sufficiency. Annals of Mathematical Statistics. 22, 79-86. (1951)
25. Clifford, R. J., Edmonson, M. N., Nguyen, C., and Buetow, K. H.: Large-scale analysis of non-synonymous coding region single nucleotide polymorphisms. Bioinformatics. 20, 1006-1014 (2004)
26. Li, B., Gallin, W. J.: Computational identification of residues that modulate voltage sensitivity of voltage-gated potassium channels. BMC Struct Biol. 5, 16-21 (2005)
27. Lee, S.Y., Banerjee, A., and MacKinnon, R.: Two Separate Interfaces between the Voltage Sensor and Pore Are Required for the Function of Voltage-Dependent K^+ Channels. PLoS Biology. 7, e47 (2009)
28. Seoh, S.A., Sigg, D., Papazian, D. M., and Bezanilla, F.: Voltage-Sensing Residues in the S2 and S4 Segments of the Shaker K^+ Channel. Neuron. 16, 1159-1167 (1996)
29. Aggarwal, S. K., and MacKinnon, R.: Contribution of the S4 Segment to Gating Charge in the Shaker K^+ Channel. Neuron. 16, 1169-77 (1996)
30. Schoppa, N. E., and Sigworth, F. J.: Activation of Shaker potassium channels I. Characterization of voltage-dependent transitions. J. Gen. Physiol. 111, 271-294 (1998)
31. R Development Core Team. R: A language and environment for statistical computing. R Foundation for statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. `http://www.R-project.org`