doi:10.1016/j.jmb.2004.04.064 J. Mol. Biol. (2004) 340, 307–318

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

An Evolutionarily Conserved Network of Amino Acids Mediates Gating in Voltage-dependent Potassium Channels

Sarel J. Fleishman[1], Ofer Yifrach[2] and Nir Ben-Tal[1]*

[1]*Department of Biochemistry George S. Wise Faculty of Life Sciences, Tel-Aviv University Ramat Aviv 69978, Israel*

[2]*Department of Life Sciences and the Zlotowski Center for Neurosciences, Ben-Gurion University of the Negev, Beer Sheva 84105, Israel*

**Corresponding author*

A novel sequence-analysis technique for detecting correlated amino acid positions in intermediate-size protein families (50–100 sequences) was developed, and applied to study voltage-dependent gating of potassium channels. Most contemporary methods for detecting amino acid correlations within proteins use very large sets of data, typically comprising hundreds or thousands of evolutionarily related sequences, to overcome the relatively low signal-to-noise ratio in the analysis of co-variations between pairs of amino acid positions. Such methods are impractical for voltage-gated potassium (Kv) channels and for many other protein families that have not yet been sequenced to that extent. Here, we used a phylogenetic reconstruction of paralogous Kv channels to follow the evolutionary history of every pair of amino acid positions within this family, thus increasing detection accuracy of correlated amino acids relative to contemporary methods. In addition, we used a bootstrapping procedure to eliminate correlations that were statistically insignificant. These and other measures allowed us to increase the method's sensitivity, and opened the way to reliable identification of correlated positions even in intermediate-size protein families. Principal-component analysis applied to the set of correlated amino acid positions in Kv channels detected a network of inter-correlated residues, a large fraction of which were identified as gating-sensitive upon mutation. Mapping the network of correlated residues onto the 3D structure of the Kv channel from *Aeropyrum pernix* disclosed correlations between residues in the voltage-sensor paddle and the pore region, including regions that are involved in the gating transition. We discuss these findings with respect to the evolutionary constraints acting on the channel's various domains. The software is available on our website http://ashtoret.tau.ac.il/~sarel/CorrMut.html

Keywords: correlated mutations; phylogenetic analysis; maximum likelihood; voltage-gated potassium channel; structural biology

Introduction

Many potassium channels are gated in response to changes in transmembrane voltage.[1–3] This form of gating underlies the production of action potentials: electrical impulses that run across the cell membrane, allowing neurons, for example, to transmit signals over their lengths.[4] Voltage-gated potassium (Kv) channels are tetramers,[5] where each monomer consists of six hydrophobic stretches (S1–S6).[6,7] The S1–S4 region comprises a voltage-sensing domain, in which the S4 segment is thought to be the voltage-sensor element,[8,9] whereas the S5–S6 regions from the four channel subunits form a central pore. This pore domain contains, in addition to the outer (S5) and inner (S6) helices, the pore helix, and the selectivity filter, which are responsible for the channel's high potassium selectivity and throughput[10] (Figure 1(a)).

Comparison of the three-dimensional pore structures of K^+ channels in the closed[10] and open[11] states revealed significant structural rearrangement

Abbreviation used: Kv, voltage-gated potassium.

E-mail address of the corresponding author: bental@ashtoret.tau.ac.il

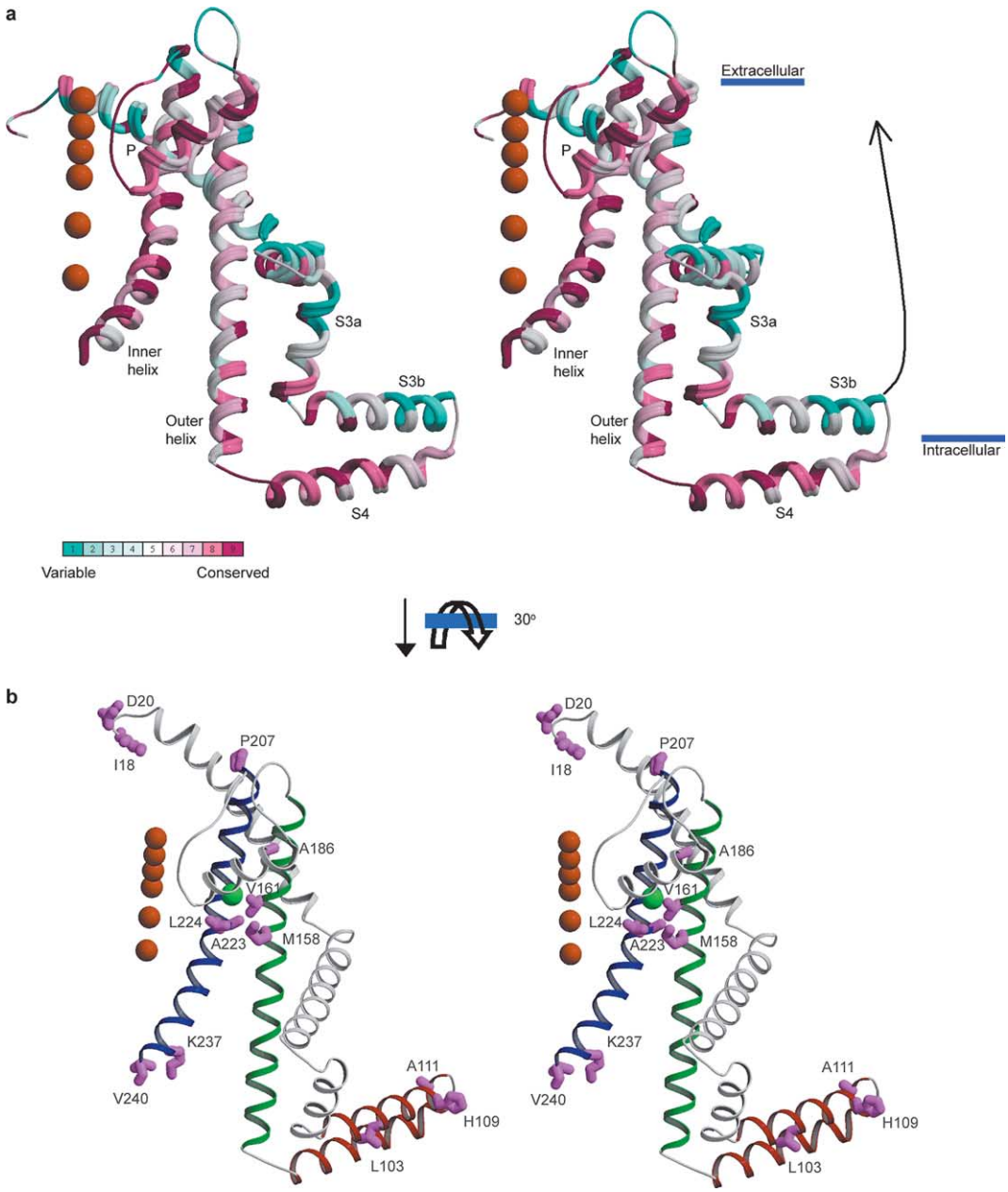

Figure 1. Stereo images of a monomer of the voltage-dependent potassium channel from *Aeropyrum pernix* (KvAP).[13] a, The trace model is color-coded according to evolutionary conservation,[61] with burgundy through turquoise, indicating conserved through variable residues (see color bar). Potassium ions are shown as orange spheres. The selectivity filter, S4 and parts of the inner helix are highly conserved, whereas the outer helix and S3b are more variable. The arrow indicates the direction of motion of the voltage sensor during the opening transition from a membrane-exposed to an extracellular-exposed orientation according to MacKinnon and co-workers' model.[14] b, The S3b–S4 segment is colored red, the outer helix green, and the inner helix blue. The position of the Gly220 gating hinge[11] is marked with a green sphere. A cluster of highly inter-correlated positions is indicated by magenta stick models. The cluster includes the following residues, where the pairwise alignment of the positions with the sequence of the *Shaker* channel[13] is indicated in parentheses: Ile18 (Tyr219) and Asp20 (Glu221) on S1; Leu103 (Thr326), His109 (Ala332) and Ala111 (Glu334) on S3b; Met158 (Ile405), Val161 (Val408) on the outer helix; Ala186 (Ala432) on the pore helix; Pro207 (Val453), Ala223 (Thr469), Leu224 (Ile470), Lys237 (Tyr483), Val240 (His486) and Glu242 (Glu488) on the inner helix. Glu242 (not shown) is missing from the KvAP structure. Table 1 lists some of the correlations connecting this cluster. The Figure was generated using MOLSCRIPT[62] and rendered with Raster3D.[63]

at the pore upon opening. The opening transition involves a large kink in the inner helix around a highly conserved Gly residue, which serves as a gating hinge. This bending allows the movement of the inner helices that leads to the disassembly of the activation gate (Figure 1b).[11] These conformational changes are mostly restricted to the intracellular portion of the channel (Figure 1a). The

region spanning the highly conserved selectivity filter remains, for the most part, rigid during gating.[11,12]

Studies by MacKinnon and co-workers of the voltage-dependent potassium channel from *Aeropyrum pernix* (KvAP) revealed several unexpected findings.[13,14] In contrast to earlier models that identified S4 as the major voltage-sensing element, they found that the S3 segment contains two helices (S3a and S3b), where the S3b helix and the N-terminal portion of the S4 helix form a tight helix hairpin, which they termed the voltage-sensor paddle.[13] Secondly, their results indicated that the paddle moves approximately 20 Å across much of the membrane span in response to the changes in transmembrane voltage[14] (Figure 1a). On the basis of these findings, MacKinnon and co-workers proposed that channel opening occurs *via* coupling of the voltage-sensor paddle's movement to that of the outer helix.[14] According to this proposition (Figure 1a), this movement, in turn, induces conformational changes in the inner helix[11] that open the energetically more stable closed structure.[15]

During the gating transition, at least three charged arginine residues per subunit in the tetramer cross the membrane[16–18] (Figure 1a). Contrary to previous models of activation (summarized by Bezanilla,[19]) the KvAP model argues that these charges move mostly in an unshielded manner through the hydrophobic membrane environment.[7,14] This conclusion[14] is astonishing from a thermodynamic point of view because of the prohibitive cost in desolvation free energy[20] associated with the transfer from water to lipid of at least a dozen charged arginine residues per channel.[14]

Following MacKinnon and co-workers' new view of voltage dependence, several studies have been devoted to test its validity.[21–23] The model has been criticized[24] for its reliance on a structure that may not be physiologically relevant owing to possible artefacts originating from co-crystallization with Fab fragments that may have distorted its conformation.[13,23] Studies on the *Shaker* homologue of KvAP provided evidence that residues within S4 are in close proximity to residues at the extracellular part of the outer helix, in apparent contradiction to the KvAP model.[21,25] In addition, it was shown that S3b does not move significantly in response to changes in the transmembrane voltage,[24] and based on accessibility studies using the homologous Kv2.1, it was suggested[22] that the motion of the voltage sensor is not as large as that implied by MacKinnon and co-workers.[14] On the basis of these results, an alternative model of the gating transition has been proposed[24] that is coherent with the previous view, in which the voltage sensor, which is comprised solely of S4, is encapsulated within a proteinaceous environment. This model further argues that the channel's conformational changes upon gating are of smaller magnitude, when compared to that suggested on the basis of experiments on KvAP.[14]

Nevertheless, relatively large conformational changes are anticipated in both gating models.[14,24] Such large changes make it exceedingly difficult to plan and interpret mutation and accessibility studies aimed at uncovering conformational substates.[14,21,22,24,26] For instance, it is difficult to control whether the modifications introduced in these studies trap the molecule in physiologically relevant states. The fact that some of these recent studies were performed on the *Shaker* homologue of KvAP,[21,22,24,25] which contains a long segment between S3 and S4 that is missing in the KvAP structure[13] (Figure 2), adds another layer of complexity.

Here, we study the inter-domain relationships in Kv channels from an evolutionary perspective. We found a network of inter-correlated amino acid positions, which cluster in functionally important regions when mapped on the KvAP structure. Specifically we show that residues on S3b, which forms part of the voltage sensor according to the KvAP structure[13,14] (Figure 1), but not according to alternative models,[24] are coupled to pore residues distributed in the vicinity of the activation gate and the gating-hinge position. These regions experience major structural rearrangements upon pore opening.[11,15]

Phylogeny-based Detection of Correlations

In silico analysis of correlated mutations has been used to identify positions that are implicated in contact formation or allosteric regulation and conformational changes in large protein families.[27–33] The underlying assumption in these studies was that functional or structural associations between a pair of positions force a coherent change in their amino acid identities during evolution. In other words, substitution of one position would induce

Figure 2. A multiple-sequence alignment showing the S3b–S4 segment of a few divergent sequences of Kv channels. Residues that were identified as part of the cluster of inter-correlated positions are shaded. The S4 segment is relatively conserved, whereas S3b, which contains the three correlated positions, is highly variable. The two helices are connected by a linker of variable length.

the other to undergo a compensatory change in order to maintain the structural or functional relationships between the two positions.

Detection of co-variation in amino acid positions within proteins, when combined with experimental data, may indicate what differences are necessary for modifying function. For instance, all isoforms of Kv channels are known to have the same ion selectivity and permeation characteristics, yet they show differences in terms of voltage sensitivity and closing and opening kinetics. Such changes might be reflected in variations in the amino acid sequences of the family. Since multiple positions are involved in determining these traits, such sequence variations should occur concomitantly in the relevant locations.

A key problem in identifying correlations in amino acid positions along multiply aligned sequences of a protein family is the difficulty in distinguishing co-variation (signal) from noise. Therefore, contemporary methods for identifying correlations often rely on very large multiple-sequence alignments of homologous proteins (typically hundreds or thousands of sequences) in order to obtain good signal-to-noise ratios.[30–32] In the case of the Kv channel family, however, only a few tens of protein sequences have been discovered. Such paucity of homologous protein sequences is typical for many protein families. Nevertheless, a collection of sequences that is sufficiently heterogeneous in terms of functions and sequences can be constructed by the inclusion of various Kv paralogues (see Materials and Methods). We present a novel method for detecting co-varying amino acid positions that is applicable for the analysis of intermediate-size protein families (50–100 sequences) that are sufficiently heterogeneous. The method is similar to that of Shindyalov *et al*.[28] in that it is based on phylogenetic reconstruction rather than on multiple-sequence alignment alone.

Generally speaking, by tracing the evolutionary pathway for every pair of amino acid positions within the protein, it is possible to substantially increase detection accuracy. As a first step in the analysis, we reconstruct the evolutionary history of the protein family by inferring the sequences of hypothetical (now-extinct) ancestral proteins of the family.[34] The phylogenetic tree (Figure 3) together with the set of reconstructed and contemporary sequences specify the evolutionary pathway that has generated the protein family as we observe it today, where each branch connects evolutionarily close sequences. By following the reconstructed pathway, we trace the changes that occurred at each evolutionary step for every position, thus reducing the errors that arise when comparing sequences that are phylogenetically distant.

Many contemporary methods for detecting correlations employ a simplistic amino acid substitution scheme, whereby all changes are treated equally.[28–30,32] Since we consider the changes that occurred at each position in subsequent evolutionary steps, we can employ a substitution matrix that reflects the subtleties of amino acid replacements in proteins more realistically, e.g. a Val for Ile change would be considered of smaller magnitude than a Gly for Trp substitution. That is, in each evolutionary step, represented by a branch on the phylogenetic tree, the changes in amino acid identities are measured. The correlations between changes in different positions of the alignment can then be calculated in a straightforward manner. We note that the method does not consider back mutations or multiple mutations in a single branch.

Here, we used the Miyata matrix,[35] which provides a measure for the physicochemical differences between amino acids. The advantages of using a phylogenetic tree are hence twofold: first, only changes that occurred at the same evolutionary interval are compared; and second, we may discriminate between small and large amino acid substitutions. Thus, the method not only detects the positions that change concurrently, but also identifies those that undergo changes of similar magnitude during evolution. We note that the Miyata[35] substitution matrix may be replaced by other substitution schemes, such as the Dayhoff matrix that was derived from the observed substitution frequencies in homologous proteins.[36]

The difficulty in detecting correlations in intermediate-size protein families is compounded by the uneven sampling or bias of homologues in sequence space. In many cases there is an over-representation of particular families of sequences, while others are under-represented. Thus, high correlations might simply be the result of a lack of variability in the given collection of sequences. To decrease bias in the set of sequences, we manually removed those that shared high homology in the S1–S6 segments with others.

The phylogenetic tree of the Kv family demonstrates that in the current selection of sequences, bias resulting from lack of variability is rather low (Figure 3). The majority of the sequences are from mammals; however, by including many paralogous sequences we were able to gain sequence variability. Following the computation of the phylogenetic tree and the reconstruction of ancestral sequences,[34] we eliminated from the alignment all positions showing relatively low entropy or information content,[37] which is a measure of the heterogeneity of amino acid identities in a given position of the alignment. This step is applied to avoid the detection of pairs of positions that changed a small number of times in the family's evolutionary history. We also eliminated positions exhibiting at least one gap in the multiple-sequence alignment because of the unreliability of ancestral-sequence reconstruction at such sites.[34]

To further reduce the possibility of errors due to bias, we derived confidence intervals for the correlation coefficients using bootstrap sampling.[38] Briefly, bootstrapping randomly generated samples

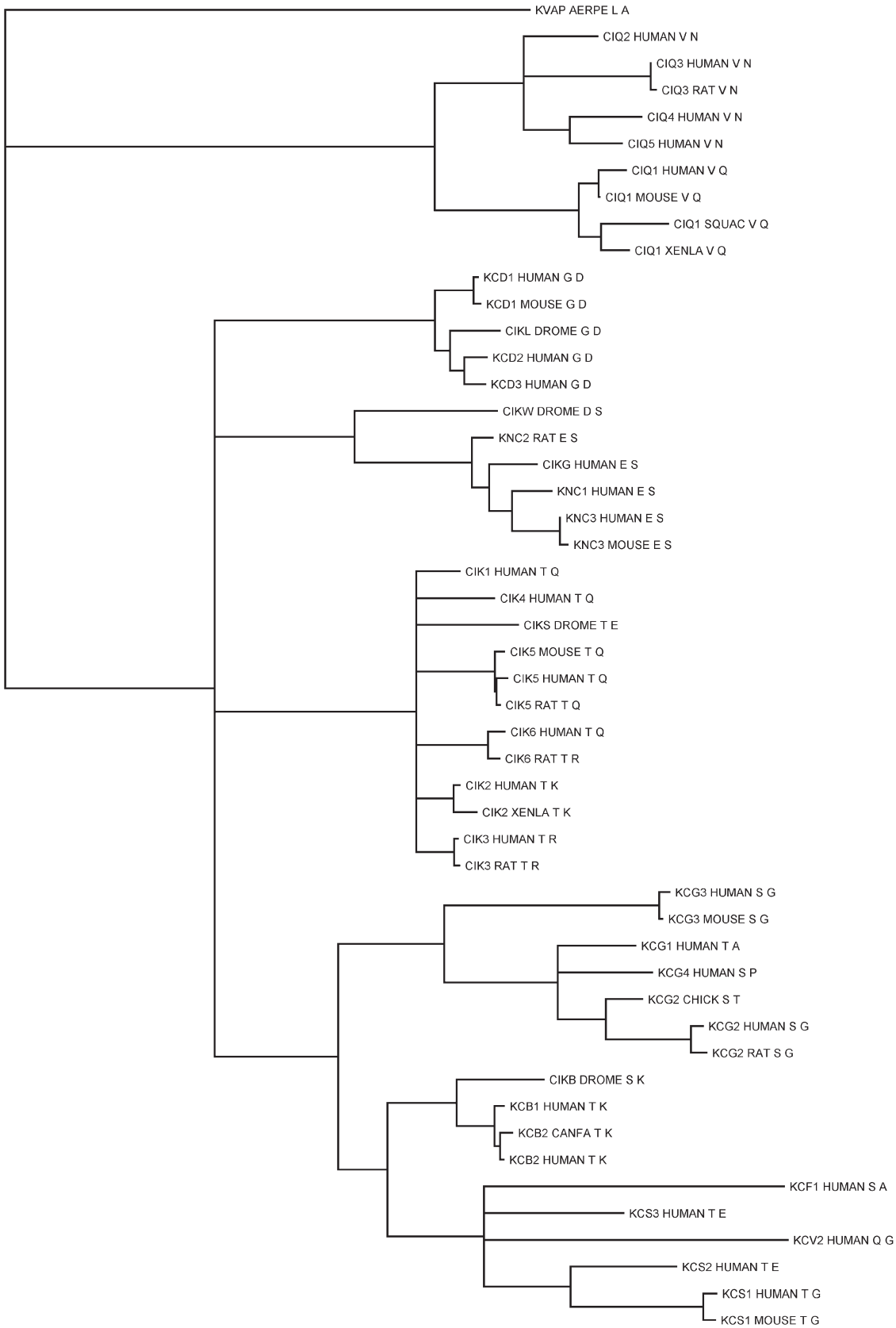

Figure 3. The phylogenetic tree used in this study, displaying at the terminal nodes one-letter codes for residues aligned with positions Leu103 (left) and Ala111 (right) on the KvAP sequence. The two correlated positions are located on the S3b helix, which forms part of the voltage sensor according to the KvAP structures[11] (Figure 1). The phylogenetic tree was computed[50] on the basis of the multiple-sequence alignment of 50 voltage-gated potassium channel sequences, and was used throughout the analysis (see Materials and Methods).

of phylogenetic branches. On each sample, the correlation coefficients of the changes occurring in each pair of positions were computed. Thus, we obtained a set of correlation coefficients for each pair of positions, on which we computed the average value as well as the confidence interval. We then eliminated correlations that were statistically insignificant. By applying the bootstrapping procedure on our data set we also rejected those correlations that were highly dependent on a particular subset of phylogenetic branches, thus reducing the possibility that the evolutionary pattern in certain parts of the tree would have a dominant effect.

We detected a large set of correlated positions, which we then subjected to principal-components analysis[39] in order to identify amino acid networks, in which all residues are highly inter-correlated. This filtering step also provided a means for reducing the effects of spurious correlations.

Results

A network of inter-correlated residues mediates channel opening

Overall, 158 correlations between pairs of amino acids were identified that met the requirements of high mean Pearson correlation coefficients ($r > 0.5$), and for which the lower confidence boundary, measured by bootstrapping, was judged to be statistically significant ($r_{low} > 0.15$). The list of correlations shows the amino acid positions to be heterogeneous, with some positions being linked to many, and others to just a few. To identify networks of highly inter-correlated positions within this list, we used principal-components analysis.[39] Several distinct sets of highly inter-correlated positions were identified. Figure 1b shows a mapping of one such set that was identified as the most significant cluster of correlated positions, on the KvAP structure. This cluster of 14 positions is linked by 50 significant correlations according to the above criteria; some of the positions, which were identified in this cluster, were associated with all of the others. Representative correlations are listed in Table 1. It may be seen that most of the inter-correlated residues are in the pore domain. However, several others were identified in the voltage-sensing paddle.

Ten out of 14 positions in the cluster of highly inter-correlated positions were previously tested in scanning-mutagenesis studies for their effects on voltage-dependent gating of the *Shaker* homologue of KvAP. Tryptophan-scanning mutagenesis showed that mutations of *Shaker* positions aligned with Leu103 and His109 on the S3b helix,[40] and Met158, Val161, Ala223 and Leu224 of the pore region[41] caused high-impact changes in gating transitions (Figure 1b). For these mutants, the voltage-activation relations were dramatically different compared to that of the wild-type channel (effects on the stability of the closed *versus* the open states). An alanine scan showed that Lys237 is another gating-sensitive residue, but mutations of three positions, including Ala186, Pro207 and Val240, that are part of the cluster on the inner helix did not alter the gating equilibrium.[15] Thus, seven out of the ten positions tested experimentally have large effects on channel-gating transitions, implying that this network of correlated amino acids may have a functional role in the voltage-induced conformational changes that lead to channel opening.

From a structural point of view, if indeed the cluster of inter-correlated residues comprises mostly gating-sensitive positions, we would expect these residues to occupy pore regions that are involved in the conformational changes during channel gating. The inter-correlated residues distributed at the intracellular end of the inner helix lie roughly two helical turns below the channel's activation gate, which opens to grant potassium ions entry into the channel during the gating transition[42,43] (Table 1; Figure 1b). The gate itself, which consists of a relatively conserved Pro-X-Pro sequence motif (*Shaker* positions 473–475), likely

Table 1. A list of representative pairs of correlated positions involving the S3b and S4 segment,[13] activation gate (C terminus of the inner helix[43]) and the gating-hinge region[11] (surrounding Gly220, shown as a green sphere in Figure 1b)

S3b-S4 inter-correlations		
Leu103	His109	0.53 (0.27, 0.74)
Leu103	Ala111	0.59 (0.34, 0.77)
Glu108	Gly114	0.50 (0.17, 0.74)
Leu113	Leu118	0.58 (0.23, 0.87)
S3b-outer helix		
Leu103	Val161	0.50 (0.24, 0.75)
His109	Val161	0.55 (0.20, 0.77)
Leu110	Ala140	0.58 (0.15, 0.83)
Leu110	Asp143	0.51 (0.19, 0.73)
Ala111	Val161	0.66 (0.34, 0.86)
S3b-gate		
Leu103	Glu242	0.69 (0.46, 0.84)
His109	Val240	0.57 (0.28, 0.79)
Ala111	Val240	0.56 (0.22, 0.83)
Ala111	Glu242	0.71 (0.42, 0.90)
Gate-gating hinge region		
Met158	Lys237	0.66 (0.31, 0.91)
Val161	Val240	0.58 (0.19, 0.83)
Ala186	Val240	0.53 (0.20, 0.79)
Ala223	Lys237	0.64 (0.35, 0.91)
Ala223	Val240	0.55 (0.18, 0.81)
Inter-correlations in the gating hinge region		
Met158	Ala223	0.59 (0.27, 0.90)
Met158	Leu224	0.71 (0.36, 0.91)
Val161	Ala223	0.58 (0.16, 0.89)
Val161	Leu224	0.79 (0.51, 0.91)
Ala186	Leu224	0.74 (0.49, 0.94)
Pro207	Ala223	0.56 (0.30, 0.80)
Pro207	Leu224	0.55 (0.26, 0.73)

The trimmed means in the 95% confidence interval of correlations (r), which were calculated from 400 bootstrapping samples, are indicated and the 95% confidence interval is parenthesized (see Materials and Methods).

forms a bend and adds flexibility to the intracellular part of the S6 segment, which is important for the opening transition.[44] The five positions in the extracellular region are all within 10 Å of a mostly conserved glycine (Gly220 in KvAP (green sphere in Figure 1b)).[10,13] This position serves as the gating hinge during pore opening, where a bend occurs in the inner helix.[11] This co-variance can be rationalized by assuming that substitutions in one region of the pore domain can be compensated for by mutations in the other. An alternative interpretation is that in order to modify existing function or to gain a new one, both regions have changed in evolution concomitantly. Since the set of sequences used in the current study consisted of many different paralogous sequences (Figure 3), where each may have slightly different characteristics of voltage sensing and gating kinetics, it is tempting to adopt the latter explanation.

The distribution pattern of inter-correlated residues on the pore domain, determined by the correlated-mutations analysis is in very good agreement with an energetics analysis of pore opening performed for the *Shaker* Kv channel.[15] In that study as well, gating-sensitive positions at the pore were found to cluster at the activation gate of the channel and at the region just extracellular to the glycine gating-hinge residue. The particular pairs that were identified in that study[15] were not highlighted in our analysis because of the fact that residues at some of the positions that were analyzed experimentally are highly conserved.

From a structural perspective, it appears unexpected that many of the positions identified in this cluster are distant in 3D space, and yet are inter-correlated. However, this result is in line with experimental data on the *Shaker* channel pore that demonstrated, by using double-mutant cycle analysis,[15] that gating-sensitive positions at the pore are energetically coupled to each other even at distances as large as 15 Å.[10,15] Such long-range energetic couplings betweens residue pairs are indicative of large tertiary or quaternary conformational changes[45] as was indeed verified upon comparison of the closed (KcsA[10]) and open (MthK[11]) pore channel structures.

The amino acid correlations detected here and their distribution pattern on the pore imply that during the evolutionary process, the activation gate and the gating hinge regions of the channel have accumulated substitutions in order to assume slightly different gating characteristics. Since the correlated positions occur mostly in regions that mediate inter-helical contacts, where extensive packing interactions occur, their substitution from one channel to another may increase or decrease the thermodynamic stability of the closed *versus* the open states of the channel. Such changes in the packing interfaces of regions that experience conformational changes during gating are expected to alter the gating kinetics.

Another interesting result in our analysis is the finding that several of the highly correlated residues occupy positions on the S3b helix, and co-vary with the pore domain residues that affect channel gating (Figure 1b; Table 1). This co-variation implies that S3b affects the opening transition, along with the activation gate and gating hinge regions. An important role for S3b in affecting the gating transition makes sense in the light of the KvAP structure, which shows that the helices S3b and S4 form one structural unit (a "paddle").[13] Moreover, electrophysiological assays demonstrated that the two helices move together between membrane and extracellular exposures in response to transmembrane voltage changes.[14] This result is also in agreement with alanine and tryptophan-scanning mutagenesis analyses, which showed that some positions on S3b are gating-sensitive.[40,46] On the other hand, a recent accessibility study showed that the *Shaker* channel's S3b segment is externally exposed in both open and closed channel states, thereby contradicting the notion that S3b and S4 move as one structural unit.[24]

Other residues in this cluster of correlated positions are less readily explained within the context of a network of positions that are involved in the gating transition. One of these positions is Pro207 at the N-terminal end of the inner helix. Mutation of the *Shaker* position that is aligned with Pro207 to either alanine[15] or tryptophan[41] did not alter channel gating significantly, thus making it unlikely that this position is involved in the gating transition. Two other positions in the cluster, Ile18 and Asp20, which are located N-terminal to S1, have not been tested experimentally. Indeed, it is difficult to imagine a role for these residues in gating according to the KvAP structure (Figure 1), but it is very likely that the structure does not represent the physiological position of S1, whose N-terminal part is intracellular.[13,47,48] It would be interesting to experimentally test whether and how this segment is coupled to the opening transition.

Sensitivity of the analysis to the phylogenetic inference method and to the amino acid substitution matrix

Using neighbor-joining

To gauge the results' sensitivity to the particular phylogenetic inference methods used here, we also computed the phylogenetic tree using the neighbor-joining algorithm.[49] This method, when compared to the maximum-likelihood program Tree-Puzzle[50] is computationally less intensive, but is less robust, in particular when analyzing a very divergent sequence set. Aside from this step, all others used to compute correlations and to derive the cluster of highly inter-correlated positions were the same as elaborated in Materials and Methods.

The list of pairs of positions showing high ($r \geq 0.5$) and significant ($r_{low} \geq 0.15$) correlations detected using the neighbor-joining tree was

approximately half the one obtained by using the phylogenetic tree of maximum-likelihood. The fact that a smaller number of correlations were found to be statistically significant indicates that the analysis based on the neighbor-joining tree was noisier than that of the maximum-likelihood tree. Based on the neighbor-joining tree, the most significant network of correlated positions included the six amino acid residues on the voltage sensor and the activation-gate region that were detected also using the tree of maximum-likelihood. However, the cluster of five positions surrounding the gating-hinge Gly220 was missing. This result indicates that while an analysis based on a neighbor-joining tree retrieves some of the most significant correlations, the results based on the maximum-likelihood tree are more sensitive, as is indeed expected.

The amino acid replacement matrix

Many existing approaches for the detection of correlated positions within protein families treat all amino acid substitutions in the same way, i.e. without differentiating among small and large changes.[28–30,32] In contrast to these methods, we have employed the Miyata substitution matrix,[35] which assigns small values to physicochemically moderate substitutions such as Val for Ile, and large values to drastic substitutions such as Gly for Trp. To test whether the use of the Miyata matrix increases the method's sensitivity, we replaced the Miyata matrix with a binary substitution matrix, in which every change in amino acid identity is given a value of 1, whereas no change is assigned a value of 0. Based on the phylogenetic tree and ancestral-sequence reconstruction of maximum likelihood, we computed the correlations among amino acids using this binary matrix.

The list of high ($r \geq 0.5$) and significant ($r_{low} \geq 0.15$) correlations was larger by more than 60% in the case of the binary matrix than when using the Miyata matrix.[35] Many of the pairs of positions showed very similar correlation coefficients and smaller confidence intervals, reflecting the lower discriminating strength of the binary substitution matrix. Importantly, whereas in the case of the Miyata matrix more than 50% of the correlations were deemed statistically insignificant using the bootstrap criterion $r_{low} \geq 0.15$, none of the correlations using the binary matrix was rejected on this basis. We conclude that, at least in cases in which relatively small sets of sequences are analyzed, a physicochemical substitution matrix is preferable.

Discussion

Recently, an alternative model for channel gating was suggested for voltage-dependent potassium channels based on biochemical, electrophysiological, and structural studies.[13,14] In addition, it has been shown that the pore opening transition involves coupled interactions in different regions of the pore domain.[15,51] In view of the alternative models of channel opening, we set out to investigate whether residues outside the pore are also coupled to regions that are important for the gating transition. We did so by examining inter-domain relationships from an evolutionary perspective.

We developed a novel method that identifies evolutionarily co-varying amino acid positions in intermediate-size protein families, and applied it to study voltage-dependent potassium channels. One of the method's strengths is its use of phylogenetic inference, allowing the algorithm to trace the evolutionary pathway for every pair of positions in the protein family. Various measures have been employed to limit the effects of bias in sequence space and of errors in ancestral-sequence reconstruction.

Despite the method's enhanced sensitivity, it cannot be used reliably to detect correlations within a family represented by a very small sequence set. The actual boundary, below which the method's dependability is too low, cannot be determined *a priori* as it is contingent on a variety of factors. However, a critical element to a reliable analysis is that the collection of homologues spans as much as possible of the function and sequence space of the protein family. This may be achieved by the inclusion of a variety of paralogues. The family of Kv channels comprises many paralogues (see Materials and Methods), providing the necessary functional and sequence variability.

The Kv channel family provided this study with a wealth of experimental data for validation and for advancing hypotheses that may not be available in families that are less well characterized. To interpret an analysis of correlations in such cases, it is possible to employ the standards used in the study of double-mutant cycles, where a coupling between positions that are proximal is deemed a consequence of physical contact, whereas the coupling of distant positions is a result of allostery (e.g. Yifrach & Mackinnon[15]).

Of the cluster of highly inter-correlated positions detected here, a large fraction (seven of ten) were also identified in mutagenesis studies as being gating-sensitive,[15,40,41] providing support for the method's capabilities in identifying functionally related positions. For comparison, Miller and co-workers found roughly 50% of the positions in the S1, S2, and S3 segments to be sensitive to substitution by tryptophan.[40,52] The fraction of highly inter-correlated positions that were found to be gating-sensitive should be considered an understatement, since two out of the three positions that were presumably identified erroneously as correlated were only tested using an alanine scan, which is relatively stringent.[15] Notably, as has been observed before, gating-sensitive positions do not necessarily map to evolutionarily conserved regions of the protein, e.g. the S3b segment[40]

(Figure 1b). It has therefore been difficult to identify such positions without the use of large-scale scanning mutagenesis experiments. Thus, an analysis of evolutionarily correlated mutations may provide a means to focus experimental efforts.

Analysis of double-mutant cycles provides a direct means to test the functional implications of evolutionarily detected couplings between positions.[32,53] Our results agree well with experimental findings on gating of Kv channels that show that the regions encompassing the glycine gating hinge on the inner helix and the activation gate in the intracellular part of the channel are energetically coupled in the context of the gating transition.[15] The correlations indicate that the energetic coupling between positions in the pore domain is also reflected by the positions' co-variation in the Kv family's evolution.

The correlations that we have identified extend this coupling, and include the S3b segment as well (Table 1, Figure 1b). The results imply that these three functional elements of the channel are evolutionarily coupled, suggesting that substitutions in S3b would have an effect on channel gating. It is interesting to note that the three positions identified on S3b are all evolutionarily variable (Figure 1). In fact, the segment's hypervariability contrasts with the relatively high conservation of S4. Moreover, the two segments are connected *via* a linker of variable length in different paralogues (Figure 2).

It has been suggested that the low conservation of the S3b helix implies lack of a functional constraint and is, therefore, an indication that its structural association with S4 is not universal.[23,24] Our analysis suggests that the functional constraint is manifested in this case through the pattern of substitutions of pairs of positions (Table 1), i.e. through the inter-correlated amino acids detected. This argument is strengthened by the observation that many positions at the C-terminal part of the S3b helix, where two of the highly inter-correlated positions are located (Figure 1b), are gating-sensitive in two different channel subtypes, despite the segment's sequence variability.[40,46,54]

The evolutionary advantage of residue substitutions in important functional regions such as the S3b part of the voltage-sensor paddle, the activation gate, and the region encompassing the conserved gating hinge is clear. Modifications in these regions would have significant effects on the gating characteristics of the channel. For instance, if S3b indeed forms part of the voltage sensor,[13] its modification might alter the sensitivity of the channel to changes in transmembrane voltage, whereas substitutions in the gate and the region surrounding the gating hinge might alter the thermodynamic stability of the open or closed states. Given that these effects are intertwined in the sense that they all modulate the gating transition, the changes that are observed in amino acid identities should be coupled as is evident through correlated-mutation analysis.

In view of these conclusions, an interesting experiment might be to identify what constitutes a minimal set of mutations that alter the function of a given Kv paralogue to obtain a channel with the characteristics of another. That the domains of Kv channels are at least grossly modular was exemplified by an experiment, in which the voltage-sensor segment (S1–S4) of the *Shaker* Kv channel was connected to the voltage-insensitive pore domain from KcSA to produce a voltage-sensitive chimera.[55] Also, substitutions of just three hydrophobic amino acid positions of the S4 segment of the *Shaw* channel to the corresponding ones of the *Shaker* member of the Kv family were enough to switch the gating characteristics of the former channel into those of the latter.[56] The fact that the S3b segment, the activation gate, and the region encompassing the gating hinge are evolutionarily coupled (Figure 1(b)) suggests that chimera that include subsequences of these regions and parts of the S4 segment from different channels might indeed switch channel characteristics in a manner that can be rationalized.

In summary, our results demonstrate that along with conservation analysis, a study of correlated mutations is an important part of phylogenetic investigation. Often in such analyses, positions that are not evolutionarily conserved are assumed to have little functional role. Here, we show, however, that certain positions that are evolutionarily variable or only show intermediate conservation (e.g. residues His109 and Lys237 of the S3b and activation-gate regions, respectively) are coupled in a functionally meaningful way (Table 1; Figure 1b). Hence, whereas the functional importance of the S4 helix as the principal carrier of the gating charge[17,18] is immediately apparent from its conservation profile (Figure 1a), the implied importance of S3b in modulating voltage sensitivity is evident only through its co-variation with other domains that are relevant to gating.

On the basis of our results, we argue that the architectural design of Kv channels, in terms of evolutionary conservation, is two-tiered. The selectivity filter and parts of the pore and inner helices are all evolutionarily constrained to maintain the channel's hallmark features of selectivity and high throughput[11,12] (Figure 1a). S4 is also highly conserved to maintain the nominal gating charge. In contrast, the activation gate and parts of the outer helix, the gating-hinge region, and the S3b helix, all of which do not directly control ion selectivity or carry the gating charge, are freer to accumulate substitutions in order to change certain gating characteristics. These substitutions are correlated, reflecting the concerted effect of these domains on the gating transition.

Kv channels show a large degree of modularity, with the activation gate, gating hinge, selectivity filter, and voltage sensor occupying different regions of the protein. This modularity contrasts with the apparently more "parsimonious" architecture of the ClC chloride channel, in which the gate

and selectivity filter inhabit the same region.[57] The evolutionary advantage of a modular architecture is that it is possible to introduce modifications to particular functions of the protein without undermining others. For instance, in Kv channels, changes in voltage sensitivity or gating kinetics upon the evolutionary pathway need not interfere with the channel's selectivity for potassium. With respect to the Kv channel family, whose members respond differently to a large spectrum of voltages, this separation of functionally important regions may have been a highly important evolutionary force shaping the voltage-gated channel structure.

Materials and Methods

Data

We constructed an initial multiple-sequence alignment of a few tens of sequences of Kv channels derived from the SWISS-PROT database.[58] On the basis of this alignment, a hidden Markov model was then constructed,[59] calibrated, and used to search for more Kv channel sequences. From the final list of homologues we removed sequences that showed very high homology to others in the data set, and retained 50 mostly mammalian sequences with a few fly, frog, fish, and chicken representatives. These consisted of Kv1.1-1.6, Kv2.1-2.2, Kv3.1-3.4, Kv4.1-4.3, Kv5.1, Kv6.1-6.3, Kv9.1-9.3, Kv11.1, the KQT members 1–5, the *Drosophila melanogaster* sequences *Shab*, *Shal*, *Shaker*, and *Shaw*, and the KvAP sequence.

All sequences except for KvAP were aligned using the CLUSTAL W algorithm using default parameters,[60] and KvAP was then added manually, based on its pairwise alignment with the *Shaker* Kv channel.[13] Because the N and C termini of the multiply aligned sequences contained many gaps, thus reducing the alignment's reliability, the sequences were trimmed to produce a core alignment consisting mostly of the transmembrane S1–S6 segments. The final alignment contained the positions corresponding to 138–505 of the *Shaker* Kv channel.

Phylogenetic reconstruction

An unrooted phylogenetic tree was computed using the maximum-likelihood method Tree-Puzzle,[50] using eight Gamma rate categories, the Muller–Vingron model of amino acid substitution, and default parameters (Figure 3). An alternative tree was constructed using the neighbor-joining method[49] based on Jukes–Cantor distances. The ancestral (now-extinct) sequences were reconstructed on the basis of both trees with the maximum-likelihood program PAML[34] using marginal reconstruction, eight Gamma rate categories, the JTT substitution matrix, and default parameters. Positions in the multiple-sequence alignment that exhibited one or more gaps were discarded owing to the uncertainty in sequence reconstruction at sites with insertion or deletion.

To gauge the extent of change at each position during the evolutionary process, we followed the phylogenetic tree, and for each amino acid position and branch, the differences in amino acid identities were converted to physicochemical distances according to the Miyata substitution matrix.[35] Alternatively, we used a binary matrix, where every change in amino acid identity was given an equal weight of 1 and no change a value of 0.

Calculating entropy

We calculated the entropy (or information content),[37] which is a measure of the heterogeneity of amino acid identities, at each position in the alignment of extinct and extant sequences according to $\Sigma_{i=1..20} - f_{p,i} \ln(f_{p,i})$, where $f_{p,i}$ is the frequency of amino acid i at position p. Positions showing entropy lower than 1.1 were eliminated. Such positions were judged to be too conserved, and therefore unlikely to contain enough information for computing correlations.

Calculating correlations among residues

Pearson correlation coefficients (r) between the physicochemical distances[35] of each pair of amino acid positions were calculated by taking into account the changes that occurred along all of the branches of the phylogenetic tree. Hence, high correlations are expected for pairs of positions, whose identities change at similar physicochemical magnitudes and at the same evolutionary time. Pairs of positions with $r < 0.5$ were assumed to be poorly correlated and were not further analyzed.

We used the bootstrap method[38] to obtain confidence intervals for the correlation coefficient of every pair of amino acid positions that were not rejected in previous steps. Bootstrapping redrew 400 samples of branches with replacements from the phylogenetic tree, i.e. pairs of evolutionarily related proteins. For each such sample, we calculated the Pearson correlation coefficient for all the pairs of positions. For each pair of positions, these 400 correlation coefficients were then sorted numerically, and the correlation coefficients at the lower (r_{low}) and upper (r_{high}) 2.5 percentiles were considered to be the lower and upper 95% confidence boundaries. For each pair of positions i, j the trimmed mean $r_{i,j}$ of the correlation coefficients in the 95% confidence interval was also computed. Pairs of positions, whose lower 2.5% confidence boundary was $r_{\mathrm{low}} < 0.15$ were rejected, as were positions with trimmed means of correlation coefficients $r_{i,j} < 0.5$.

As a further test of significance, we discarded pairs of positions if one or both of the residues i in each pair was found to have $\mathrm{cov}(i, i) = 0$ in at least 2.5% of the bootstrap samples. This eliminated correlations that were high simply because of homogeneous evolutionary conservation in large parts of the phylogenetic tree.

Principal components analysis

In searching for networks of highly inter-correlated amino acid positions, we subjected the complete set of correlated pairs of positions to principal-components analysis.[39] We constructed a symmetric matrix of all of the correlations identified in the study, where each element i, j of the matrix corresponded to the trimmed mean of correlations between positions i and j, $r_{i,j}$. Pairs showing low ($r_{i,j} < 0.5$) or insignificant ($r_{\mathrm{low}} < 0.15$) correlations were assigned a value of 0, and the diagonal elements a value of 1. The matrix was decomposed into eigenvalues and eigenvectors. The eigenvector associated with the eigenvalue of highest magnitude was regarded as the most significant correlated network. Position i was considered to be part of this cluster if

$|e_i| \geq 0.15$, where e_i corresponds to element i in the eigenvector.

Availability

The programs used in this analysis are available†.

Acknowledgements

The authors thank Christopher Miller, Dan Graur, and Meytal Landau for helpful discussions. This study was supported by a Research Career Development Award from the Israel Cancer Research Fund to N.B.T. and by a doctoral fellowship from the Clore Israel Foundation to S.J.F. Some computations were conducted using the facilities of the Bioinformatics Service Unit at Tel-Aviv University.

References

1. Sigworth, F. J. (1994). Voltage gating of ion channels. *Quart. Rev. Biophys.* **27**, 1–40.
2. Yellen, G. (1998). The moving parts of voltage-gated ion channels. *Quart. Rev. Biophys.* **31**, 239–295.
3. Bezanilla, F. (2000). The voltage sensor in voltage-dependent ion channels. *Physiol. Rev.* **80**, 555–592.
4. Hille, B. (2001). *Ion Channels of Excitable Membranes*, 3rd edit., Sinauer Associates, Sunderland, MA.
5. MacKinnon, R. (1991). Determination of the subunit stoichiometry of a voltage-activated potassium channel. *Nature*, **350**, 232–235.
6. Tempel, B. L., Papazian, D. M., Schwarz, T. L., Jan, Y. N. & Jan, L. Y. (1987). Sequence of a probable potassium channel component encoded at Shaker locus of Drosophila. *Science*, **237**, 770–775.
7. Miller, C. (2003). A charged view of voltage-gated ion channels. *Nature Struct. Biol.* **10**, 422–424.
8. Papazian, D. M., Timpe, L. C., Jan, Y. N. & Jan, L. Y. (1991). Alteration of voltage-dependence of Shaker potassium channel by mutations in the S4 sequence. *Nature*, **349**, 305–310.
9. Liman, E. R., Hess, P., Weaver, F. & Koren, G. (1991). Voltage-sensing residues in the S4 region of a mammalian K^+ channel. *Nature*, **353**, 752–756.
10. Doyle, D. A., Morais Cabral, J., Pfuetzner, R. A., Kuo, A., Gulbis, J. M., Cohen, S. L., Chait, B. T. & MacKinnon, R. (1998). The structure of the potassium channel: molecular basis of K^+ conduction and selectivity. *Science*, **280**, 69–77.
11. Jiang, Y., Lee, A., Chen, J., Cadene, M., Chait, B. T. & MacKinnon, R. (2002). The open pore conformation of potassium channels. *Nature*, **417**, 523–526.
12. Kelly, B. L. & Gross, A. (2003). Potassium channel gating observed with site-directed mass tagging. *Nature Struct. Biol.* **10**, 280–284.
13. Jiang, Y., Lee, A., Chen, J., Ruta, V., Cadene, M., Chait, B. T. & MacKinnon, R. (2003). X-ray structure of a voltage-dependent K^+ channel. *Nature*, **423**, 33–41.
14. Jiang, Y., Ruta, V., Chen, J., Lee, A. & MacKinnon, R. (2003). The principle of gating charge movement in a voltage-dependent K^+ channel. *Nature*, **423**, 42–48.
15. Yifrach, O. & MacKinnon, R. (2002). Energetics of pore opening in a voltage-gated K(+) channel. *Cell*, **111**, 231–239.
16. Schoppa, N. E., McCormack, K., Tanouye, M. A. & Sigworth, F. J. (1992). The size of gating charge in wild-type and mutant Shaker potassium channels. *Science*, **255**, 1712–1715.
17. Seoh, S. A., Sigg, D., Papazian, D. M. & Bezanilla, F. (1996). Voltage-sensing residues in the S2 and S4 segments of the Shaker K^+ channel. *Neuron*, **16**, 1159–1167.
18. Aggarwal, S. K. & MacKinnon, R. (1996). Contribution of the S4 segment to gating charge in the Shaker K^+ channel. *Neuron*, **16**, 1169–1177.
19. Bezanilla, F. (2002). Voltage sensor movements. *J. Gen. Physiol.* **120**, 465–473.
20. Honig, B. H. & Hubbell, W. L. (1984). Stability of "salt bridges" in membrane proteins. *Proc. Natl Acad. Sci. USA*, **81**, 5412–5416.
21. Laine, M., Lin, M. C., Bannister, J. P., Silverman, W. R., Mock, A. F., Roux, B. & Papazain, D. M. (2003). Atomic proximity between S4 segment and pore domain in Shaker potassium channels. *Neuron*, **39**, 467–481.
22. Lee, H. C., Wang, J. M. & Swartz, K. J. (2003). Interaction between extracellular Hanatoxin and the resting conformation of the voltage-sensor paddle in Kv channels. *Neuron*, **40**, 527–536.
23. Cohen, B. E., Grabe, M. & Jan, L. Y. (2003). Answers and questions from the KvAP structures. *Neuron*, **39**, 395–400.
24. Gandhi, C. S., Clark, E., Loots, E., Pralle, A. & Isacoff, E. Y. (2003). The orientation and molecular movement of a K(+) channel voltage-sensing domain. *Neuron*, **40**, 515–525.
25. Broomand, A., Mannikko, R., Larsson, H. P. & Elinder, F. (2003). Molecular movement of the voltage sensor in a k channel. *J. Gen. Physiol.* **122**, 741–748.
26. Sigworth, F. J. (2001). Potassium channel mechanics. *Neuron*, **32**, 555–556.
27. Gobel, U., Sander, C., Schneider, R. & Valencia, A. (1994). Correlated mutations and residue contacts in proteins. *Proteins: Struct. Funct. Genet.* **18**, 309–317.
28. Shindyalov, I. N., Kolchanov, N. A. & Sander, C. (1994). Can three-dimensional contacts in protein structures be predicted by analysis of correlated mutations? *Protein Eng.* **7**, 349–358.
29. Kass, I. & Horovitz, A. (2002). Mapping pathways of allosteric communication in GroEL by analysis of correlated mutations. *Proteins: Struct. Funct. Genet.* **48**, 611–617.
30. Oliveira, L., Paiva, A. C. & Vriend, G. (2002). Correlated mutation analyses on very large sequence families. *Chembiochem*, **3**, 1010–1017.
31. Valencia, A. & Pazos, F. (2002). Computational methods for the prediction of protein interactions. *Curr. Opin. Struct. Biol.* **12**, 368–373.
32. Lockless, S. W. & Ranganathan, R. (1999). Evolutionarily conserved pathways of energetic connectivity in protein families. *Science*, **286**, 295–299.
33. Suel, G. M., Lockless, S. W., Wall, M. A. & Ranganathan, R. (2003). Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nature Struct. Biol.* **10**, 59–69.
34. Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556.

† http://ashtoret.tau.ac.il/~sarel/CorrMut.html

35. Miyata, T., Miyazawa, S. & Yasunaga, T. (1979). Two types of amino acid substitutions in protein evolution. *J. Mol. Evol.* **12**, 219–236.
36. Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. (1979). A model of evolutionary change in proteins. In *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., ed.), pp. 353–358, National Biomedical Research Foundation, Washington, DC.
37. Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423, 623–656.
38. Bradley, E. & Tibshirani, R. (1993). *An Introduction to the Bootstrap*, Chapman & Hall, New York.
39. Lebart, L., Morineau, A. & Warwick, K. M. (1984). *Multivariate Descriptive Statistical Analysis*, Wiley, New York.
40. Hong, K. H. & Miller, C. (2000). The lipid–protein interface of a Shaker K(+) channel. *J. Gen. Physiol.* **115**, 51–58.
41. Li-Smerin, Y., Hackos, D. H. & Swartz, K. J. (2000). A localized interaction surface for voltage-sensing domains on the pore domain of a K^+ channel. *Neuron*, **25**, 411–423.
42. Holmgren, M., Shin, K. S. & Yellen, G. (1998). The activation gate of a voltage-gated K^+ channel can be trapped in the open state by an intersubunit metal bridge. *Neuron*, **21**, 617–621.
43. del Camino, D. & Yellen, G. (2001). Tight steric closure at the intracellular activation gate of a voltage-gated K(+) channel. *Neuron*, **32**, 649–656.
44. Labro, A. J., Raes, A. L., Bellens, I., Ottschytsch, N. & Snyders, D. J. (2003). Gating of shaker-type channels requires the flexibility of S6 caused by prolines. *J. Biol. Chem.* **278**, 50724–50731.
45. Ranganathan, R., Lewis, J. H. & MacKinnon, R. (1996). Spatial localization of the K^+ channel selectivity filter by mutant cycle-based structure analysis. *Neuron*, **16**, 131–139.
46. Li-Smerin, Y., Hackos, D. H. & Swartz, K. J. (2000). alpha-helical structural elements within the voltage-sensing domains of a K(+) channel. *J. Gen. Physiol.* **115**, 33–50.
47. Blaustein, R. O., Cole, P. A., Williams, C. & Miller, C. (2000). Tethered blockers as molecular "tape measures" for a voltage-gated K^+ channel. *Nature Struct. Biol.* **7**, 309–311.
48. Santacruz-Toloza, L., Huang, Y., John, S. A. & Papazian, D. M. (1994). Glycosylation of shaker potassium channel protein in insect cell culture and in Xenopus oocytes. *Biochemistry*, **33**, 5607–5613.
49. Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.
50. Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. (2002). TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics*, **18**, 502–504.
51. Lu, Z., Klem, A. M. & Ramu, Y. (2002). Coupling between voltage sensors and activation gate in voltage-gated K^+ channels. *J. Gen. Physiol.* **120**, 663–676.
52. Monks, S. A., Needleman, D. J. & Miller, C. (1999). Helical structure and packing orientation of the S2 segment in the Shaker K^+ channel. *J. Gen. Physiol.* **113**, 415–423.
53. Carter, P. J., Winter, G., Wilkinson, A. J. & Fersht, A. R. (1984). The use of double mutants to detect structural changes in the active site of the tyrosyl-tRNA synthetase (*Bacillus stearothermophilus*). *Cell*, **38**, 835–840.
54. Perozo, E. (2000). Structure and packing orientation of transmembrane segments in voltage-dependent channels. Lessons from perturbation analysis. *J. Gen. Physiol.* **115**, 29–32.
55. Lu, Z., Klem, A. M. & Ramu, Y. (2001). Ion conduction pore is conserved among potassium channels. *Nature*, **413**, 809–813.
56. Smith-Maxwell, C. J., Ledwell, J. L. & Aldrich, R. W. (1998). Uncharged S4 residues and cooperativity in voltage-dependent potassium channel activation. *J. Gen. Physiol.* **111**, 421–439.
57. Dutzler, R., Campbell, E. B. & MacKinnon, R. (2003). Gating the selectivity filter in ClC chloride channels. *Science*, **300**, 108–112.
58. Bairoch, A. & Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucl. Acids Res.* **28**, 45–48.
59. Eddy, S. R. (1996). Hidden Markov models. *Curr. Opin. Struct. Biol.* **6**, 361–365.
60. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680.
61. Pupko, T., Bell, R. E., Mayrose, I., Glaser, F. & Ben-Tal, N. (2002). Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics*, **18**, S71–S77.
62. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
63. Merritt, E. A. & Bacon, D. J. (1997). Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524.

Edited by G. von Heijne

(Received 21 January 2004; received in revised form 26 April 2004; accepted 30 April 2004)