

Evolutionary coupling in the $K_v1.2-\beta_2$ complex

Shreedhar Natarajan,^{1†} Jay Mashl² and Eric Jakobsson^{1-3,*}

¹Biophysics and Computational Biology; ²National Center for Supercomputing Applications; ³Department of Molecular and Integrative Physiology; University of Illinois at Urbana-Champaign; Urbana-Champaign, IL USA

[†]Current address: Department of Biology; University of Pennsylvania; Philadelphia, PA USA

Key words: channels, hypoxia, $K_v1.2$, β_2 , co-evolution, RACK1, NADPH, oxidoreductases

$K_v1.2$ is a potassium channel protein whose electrophysiological properties, including an oxygen sensing response, are modulated by auxiliary beta subunits. The beta subunits are homologous to oxidoreductases, supporting a hypothesis that the coupling of the two subunits helps connect the redox state of the cell to the electrical activity of the membrane. However the exact mechanism of the coupling has not been discovered to-date.

We apply evolutionary correlation analysis to infer previously unknown components of the interaction network regulating the response to hypoxia of the $K_v1.2/\beta_2$ complex. Briefly, evolutionary correlation analysis involves finding correlated amino acid substitutions in functionally equivalent proteins (for both subunits) across a range of species. The method thus depends on a reliable method of inferring functionally equivalent (orthologous) proteins in different species, which method we describe in a paper recently published in PLoS ONE. One key finding is the characterization of a network of motif interactions between the α and the β subunits. By significance testing, we show that the likelihood of the correlations shown in the $K_v1.2-\beta$ two interaction motifs arising by chance is less than 0.0003, which shows that the correlations are statistically highly significant. We therefore believe that the correlations are likely to be biologically relevant. Other major findings are correlations between specific motifs in the $K_v1.2-\beta_2$ complex and motifs in other proteins such as RACK1 and Eif3s6ip, which in turn are connected to the hypoxia response. Our paper combines this correlation with the literature evidence on potassium channel inactivation and hypoxia response to identify specific motifs to serve as experimental targets for studies focused on this response.

This work aims to add to our general understanding of two major issues in ion channel science: (1) how multi-protein complexes including ion channels function in coordinated fashion and (2) how ion channels mediate the conversation between the intracellular and extracellular environments. We also aim to apply to ion channel science the principle that domain-domain interactions in proteins can be inferred from correlation of amino-acid substitutions in sets of functionally equivalent proteins.

Introduction

Voltage gated potassium channels (K_v) are a large and diverse family¹ of ion channels that allow selective permeation of potassium ions at specific transmembrane voltages. $K_v1.2$ is a Shaker-like channel that works as a delayed rectifier, to bring the membrane potential back to its resting state after an action potential. The recently solved crystal structure of the $K_v1.2-\beta_2$ complex² reveals a homo-tetrameric structure. Each $K_v1.2 \alpha$ subunit has 6 transmembrane helices (S1-S6) with the S1-S4 forming the voltage sensing apparatus, and the S5-S6 helices along with the pore helix, forming part of the pore region responsible for permeation of K^+ ions. Many ion channels, and in particular voltage-gated potassium channels, have auxiliary subunits attached to them that can modulate their electrophysiological parameters.³ The β subunits are one type of cytoplasmic auxiliary subunit that can modulate the inactivation kinetics of K_v channels.⁴ At least 3 types of β subunits have been identified, β_1 ,

β_2 , β_3 , each of which may have multiple isoforms.⁴ These beta subunits also tetramerize like the $K_v1.2 \alpha$ subunits and attach to the corresponding α subunits via the T1 domain. The β subunits are homologous to aldo-keto reductases and have the typical TIM-barrel fold that is characteristic of this family. The crystal structure of the $K_v1.2-\beta_2$ complex also has a NADP⁺ co-factor bound to the β_2 subunit. It was found that the binding of NADP⁺ could influence potassium channel trafficking.⁵ Differential binding of the reduced and oxidized forms also had effects on the degree of inactivation of the $K_v \alpha$ subunits.^{6,7} Since $K_v\beta_2$ does not have the long N-terminal ball and chain region that inactivates K_v channels,⁸ the exact functional advantage of $K_v1.2-\beta_2$ complexation is still to be resolved. Both $K_v1.2$ and the β subunits have been shown to be important for oxygen sensing.^{2,9,10} In addition, hypoxia can cause inactivation of $K_v1.2$ channels.^{9,11} Given the large number of combinatorial possibilities of $K_v-\beta$ complexation, experimental exploration will be greatly aided by computational techniques that can pinpoint interaction sites and

*Correspondence to: Eric Jakobsson; Email: jake@ncsa.uiuc.edu

Submitted: 05/12/10; Revised: 06/20/10; Accepted: 06/29/10

Previously published online: www.landesbioscience.com/journals/channels/article/12813

DOI: 10.4161/chan.4.5.12813

functionally relevant regions for further analysis. In this paper, we explore the evolutionary coupling between the two subunits with a view to understanding their overall functional coupling.

Evolutionary correlation or co-evolution between proteins has been explored both at the sequence level and at the whole genome level. Phylogenetic profiling¹² assumes that the presence/absence of proteins in genomes will be correlated if they have co-evolved to maintain an interaction. Phylogenetic congruency¹³ is an extension of the phylogenetic profiling technique, but at the level of sequences, rather than the presence/absence alone. The underlying premise is again that of correlated evolution, whereby mutations that occur in one protein of an interacting pair would cause correlated mutations in its interacting partner, with a view to preserving the interaction. The correlations are detected by comparing the sequence evolution of both proteins across several species. Since sequences are represented by symbols instead of numbers, the comparison is done using distance matrices that accompany the construction of phylogenetic trees.

One of the early applications of this method was for the ligand-receptor systems.¹⁴ The technique was benchmarked with domain interactions in PGK (phosphoglycerate kinase). Pazos et al.¹⁵ extended the technique to genome-wide identification of protein interaction networks. Kim et al. extended the technique to studying the list of all structural domain interactions in the PDB using PSIMAP.¹⁶ In most of the analyses carried out using the phylogenetic congruency technique, a significant level of background correlation was found to exist even between proteins that were not functionally related. This background correlation was attributed to speciation. This means that differences in sequence between human and chimp for two different proteins might simply arise due to species wide pressures for humans and chimps. Several attempts have been made to subtract the effect of the “speciation-vector”.^{17,18} There was a decrease in false positives, but in some of the techniques, the proportion of false negatives increased. A recent effort¹⁹ attempted to identify protein interactions using the mirror-tree approach on a genome-wide scale. Rather than subtract speciation vectors, they used the idea of higher-order correlations and partial correlations to extract highly specific interactions. The overall accuracy was very high, making coevolution a strong candidate for identifying genome wide protein interaction networks.

There are several techniques that predict domain-domain interactions from a given protein network. Since our analysis is limited to two proteins, we shall not cover those methods here; the interested reader is referred to a recent review on this topic.¹³ We also limit our discussion to sequence based techniques; a good review of structure-based techniques can be found in Aloy et al.²⁰ Early studies²¹ used the InterPro²² databases to learn which of the domain/motif pair combinations were more frequent in interacting proteins than expected by chance. The studies were aimed at providing an interaction screen to reduce the experimental search space of all interacting pairs. Several other techniques²³ have built on these ideas with sophisticated algorithms; most depend on PFAM or InterPro domains and test their accuracy with experimentally predicted interactions.

Kim et al.²⁴ took a different approach to inferring domain interactions. They applied the idea of phylogenetic profiles to

protein domains instead of whole proteins. Their technique could pick up functional links previously missed by the original phylogenetic profiling technique. They extended this approach²⁵ to studying correlations at the residue level, and identified potential interacting residues that were contiguous in sequence space and could correspond to domain-domain interactions. Their system successfully predicted domain interactions in chemotaxis proteins in *E. coli* and the proteins involved in regulation of the high-affinity potassium transport ATPase. Jothi et al.²⁶ extended the technique (RCDP—Relative coevolution of domain pairs) and predicted known interactions in iPFAM²⁷ with an accuracy of approx. 65% as compared to 55% accuracy of randomized predictions. In a subsequent paper,²⁸ they attempted to remove background correlations using entropy thresholds in addition to the previously mentioned speciation subtraction techniques. They hypothesized that in regions of low entropy, coevolution resulting from preservation of function was less likely to be saturated with species divergence. The entropy reduction step (ERS) was an input to the speciation subtraction and was useful in decreasing false positives.

In addition to the above domain-based methods, several techniques have been developed to identify coevolving residues between and within proteins.²⁹ Notably, Mintseris et al.³⁰ studied obligate and transient interactions and concluded that the substitution rate at the interfaces of obligate interaction partners were slower and thus showed good correlations, while the residues at the interfaces of transient interactions had much more plasticity and evolved at too rapid a rate to leave evidence of correlated mutations. Studies involving intra-protein residue correlation focused primarily on predicting structural contacts within the protein.^{31,32} Recently, Yeang et al.³³ introduced a continuous time Markov process model of sequence coevolution and showed that coevolving residues were not only involved in functionally linked domains; they were also spatially coupled. Almost all of the above-mentioned methods have had different approaches to computing the datasets used in the correlation analysis. Each method uses a set of heuristics to allow for uncertainties in their methodology as well as in the validation datasets. A critical review by Halperin et al.³⁴ found that the method of evolutionary correlation was useful in predicting intramolecular contacts, but with present techniques would not generally be useful for prediction of intermolecular contacts, except for specific cases. One limiting factor cited was the inability to identify a substantial number of truly orthologous protein sequence pairs in multiple species for each potential interacting pair. In our paper, we have attempted to overcome that limitation by application of our MoLFuncs method for identification of functional counterparts across multiple organisms. The mirror-tree method that we adopt in our analysis, requires orthologs of the proteins being investigated across multiple species. In a previous paper,³⁵ we described a technique to extract the Most Likely Functional Counterparts (MoLFuncs) of a protein from a set of species. The MoLFunc is analogous to the concept of ortholog, but works at the protein level in contrast to the gene level and aims to address the issue of picking the right isoform of the protein being investigated. This is especially relevant for biophysical studies where tissue-specific

splice variants might have both subtle changes in their sequences and consequently their function as well. The MoLFunCs algorithm resulted in a sufficiently large and diverse sample size for the coevolution analysis.

Our determination of MoLFunCs is based on the idea of “authority”—the term refers to the degree of confidence in the annotation of a protein’s function. The $K_v1.2$ and the β_2 from rat were chosen as the authorities due to the availability of their crystal structure.² We are interested in the following primary goals:

(1) Develop a protocol to accurately identify the motif pairs that mediate the physical connection between two interacting proteins.

(2) Identify functionally relevant long-range correlations between motif pairs in the two proteins.

(3) Identify emergent properties of the correlation network and analyze them in the context of known functionality of the protein complex.

The rest of the paper describes the methodology of extracting motifs, computing correlations between motifs, estimating the significance of the correlations and a detailed analysis of the correlations that we hope will shed some light on the overall functionality of the $K_v1.2$ - β_2 complex.

Results

Highly correlated motif pairs. Table 1 shows the 30 motif pairs between $K_v1.2$ and β_2 that exhibit the most highly correlated evolution of the 491*359 total potential motif pairs. The bold entries show 10 motif pairs that are proximal in the 3D structure while the italicized entries show 20 that are not proximal. The strong correlation of motif pairs in structural proximity serves to validate the method, since it is clear that evolutionary change in these areas must be correlated. We note that the $K_v1.2$ motifs in the structurally relevant pairs are in the T1 region of the channel.² The strong correlation of some motif pairs that are distant from each other suggests that both regions are subject to some common adaptive pressure.

Recall/precision analysis. Figure 1 shows the precision (see Methods for definition) at the standard recall levels for the method applied with respect to the structurally relevant (i.e., most proximal) set of motif pairs (as defined in Methods). The significance tests for this figure correspond to *Case 1a* (see Methods for definition). It is clear that precision falls rapidly as recall increases for all the measures used; i.e., a higher fraction of the motif pairs recalled are not structurally relevant. The rapid fall of precision indicates the accumulation of “structural false positives”; i.e., apparently highly correlated motif pairs that are not in regions of strong spatial interaction. It should be noted that designation as a “false positive” does not necessarily mean that the high correlation is either accidental or artifactual. Since the correlations are so strong as to be unlikely to arise accidentally, the more likely interpretation is that the two regions are participating in a coordinated response to the same evolutionary pressure. Note that it is not to be expected that all of the structurally relevant motif pairs will score equally highly, since only a subset of them represent actual structural arrangements of the

Table 1. Table listing the highly correlated motifs between β_2 and $K_v1.2$ sorted in decreasing order of correlation values

Beta2 motif start	Beta2 motif end	$K_v1.2$ motif start	$K_v1.2$ motif end	Correlation score	p-value
229	237	63	71	0.9907	0.0002
229	237	62	70	0.9902	0.0002
<i>158</i>	<i>166</i>	<i>377</i>	<i>385</i>	<i>0.9901</i>	<i>0.0002</i>
<i>158</i>	<i>166</i>	<i>378</i>	<i>386</i>	<i>0.9897</i>	<i>0.0002</i>
<i>158</i>	<i>166</i>	<i>379</i>	<i>387</i>	<i>0.9897</i>	<i>0.0002</i>
<i>158</i>	<i>166</i>	<i>380</i>	<i>388</i>	<i>0.9897</i>	<i>0.0002</i>
233	241	62	70	0.989	0.0002
235	243	64	72	0.9882	0.0002
<i>157</i>	<i>165</i>	<i>376</i>	<i>384</i>	<i>0.9865</i>	<i>0.0002</i>
234	242	62	70	0.9865	0.0002
233	241	63	71	0.986	0.0002
<i>157</i>	<i>165</i>	<i>377</i>	<i>385</i>	<i>0.9857</i>	<i>0.0002</i>
<i>159</i>	<i>167</i>	<i>382</i>	<i>390</i>	<i>0.9853</i>	<i>0.0002</i>
<i>159</i>	<i>167</i>	<i>383</i>	<i>391</i>	<i>0.9853</i>	<i>0.0002</i>
<i>226</i>	<i>234</i>	<i>382</i>	<i>390</i>	<i>0.9845</i>	<i>0.0002</i>
<i>226</i>	<i>234</i>	<i>383</i>	<i>391</i>	<i>0.9845</i>	<i>0.0002</i>
<i>159</i>	<i>167</i>	<i>378</i>	<i>386</i>	<i>0.9838</i>	<i>0.0002</i>
<i>159</i>	<i>167</i>	<i>379</i>	<i>387</i>	<i>0.9838</i>	<i>0.0002</i>
<i>159</i>	<i>167</i>	<i>380</i>	<i>388</i>	<i>0.9838</i>	<i>0.0002</i>
<i>158</i>	<i>166</i>	<i>376</i>	<i>384</i>	<i>0.9835</i>	<i>0.0002</i>
<i>226</i>	<i>234</i>	<i>375</i>	<i>383</i>	<i>0.9833</i>	<i>0.0002</i>
234	242	63	71	0.9832	0.0002
233	241	64	72	0.9832	0.0002
<i>50</i>	<i>58</i>	<i>384</i>	<i>392</i>	<i>0.9827</i>	<i>0.0003</i>
<i>331</i>	<i>339</i>	<i>378</i>	<i>386</i>	<i>0.9825</i>	<i>0.0003</i>
<i>331</i>	<i>339</i>	<i>379</i>	<i>387</i>	<i>0.9825</i>	<i>0.0003</i>
<i>331</i>	<i>339</i>	<i>380</i>	<i>388</i>	<i>0.9825</i>	<i>0.0003</i>
230	238	62	70	0.9825	0.0003
231	239	62	70	0.9825	0.0003
<i>331</i>	<i>339</i>	<i>377</i>	<i>385</i>	<i>0.9822</i>	<i>0.0003</i>

Bold indicates true positives and italics indicate “structural false positives”.

interacting surfaces. Therefore, getting even one of the relevant pairs as the highest correlated pair, can be considered a successful validation of the method. By this logic, we need to consider primarily the precision at 0% recall, which is really an extrapolated precision. The overall precision at 0% recall is 100% and six of the top eleven hits are structurally relevant (Table 1). Although structural false positives accumulate quickly as one goes lower down the list of descending correlations, the relatively high hits are not distributed randomly across the sequences, but rather are clustered around a few regions in both proteins. We shall investigate these high-scoring pairs in the next section. Figure 2 shows these clustered pairs of interacting regions on the 3D structure.

Figure 3 describes our results for *Case 1b* (see Methods for definition), in which we start with a motif on one protein and

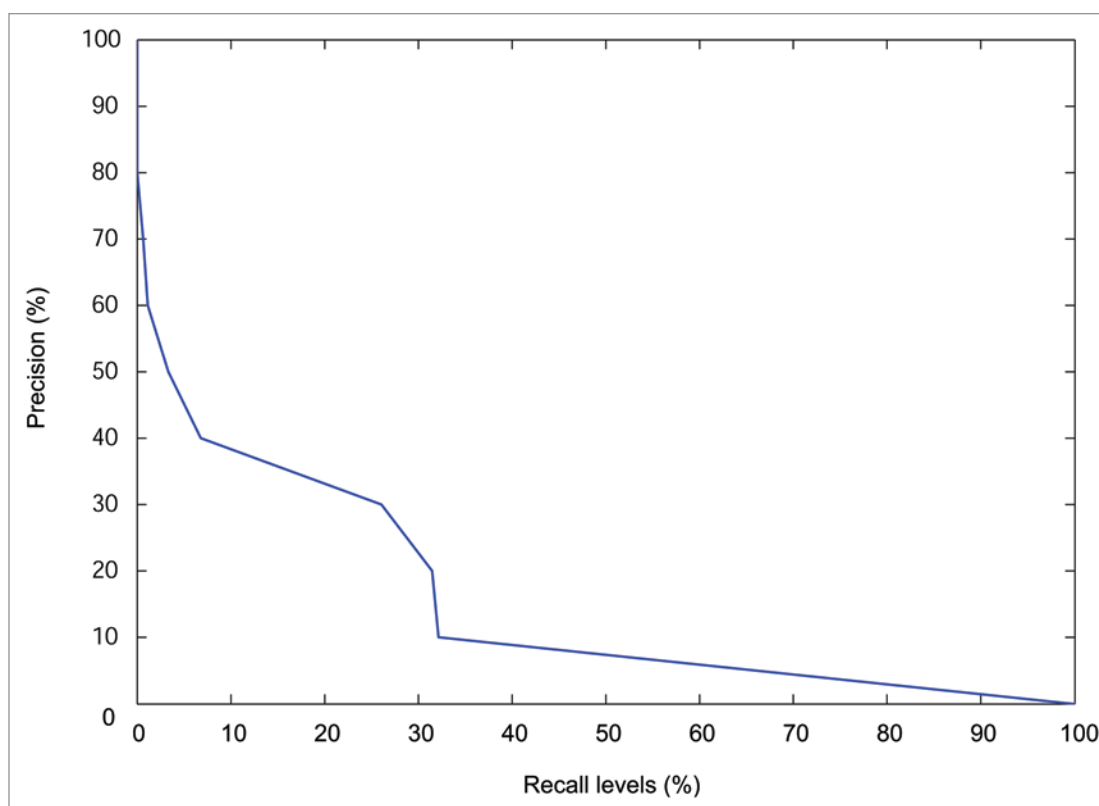


Figure 1. Precision Recall curve. Query is two proteins that may interact (pair shown is $K_v1.2$ and β_2) and the results are a sorted list of highly correlated motif pairs. As discussed in the text, each motif is a gapless 9-mer.

predict the interacting motifs on the other protein. All the motifs used as queries are extracted from Table 2. In Figure 3A, we see an image with the horizontal axis representing the starting residue position of a $K_v1.2$ motif used as query, the vertical axis represents the percent recall and the grayscale intensity represents the percent precision; the values corresponding to the grayscale level is given on the right. The $K_v1.2$ motifs represented in this figure are in the region where the surfaces of the two proteins physically come together. We can see that for all the $K_v1.2$ motifs shown, we are able to retrieve the interacting motifs from the β_2 subunit with 100 percent precision for several levels of recall. Similarly, Figure 3B represents the performance of our method with respect to β_2 motifs as query. As with the Figure 3A motifs, these motifs are in the region where the surfaces of the two proteins physically come together. Again, (except for the first two motifs), we get a high level of precision for several recall levels. Overall, we can conclude that coevolution analysis is acceptably good at predicting interacting sites on proteins. In the next few sections, we shall take a closer look at the “false positives” and “possible false negatives” and assess their significance in the context of biophysical knowledge of the $K_v1.2$ - β_2 complex.

C-terminus docking and the major “false positive” in Table 1. A recent paper suggested, based on electron microscopy, that the C-terminus of $K_v1.2$,³⁶ which was not resolved fully in the 3D structure,² could interact with a region on the β_2 subunit.³⁶ Figure 4A shows the tetramer with all subunits as visualized from the X-ray structure. The proposed region of β_2 that interacts with

the C-terminus is also shown. It is seen that the X-ray structure does not contain the putative interacting partner in $K_v1.2$. In addition to the lack of structure, there are gaps in the alignments near the C-terminus of $K_v1.2$, which prevent us from seeking direct correlations that might correspond to the interaction seen in 36.³⁶ However, projecting from our correlation analysis (Table 1) onto Figure 4A, we see that a very high-scoring β_2 region (157–167) is also in the proposed β_2 interaction region reported by Sokolova et al.³⁶ Our correlation suggests that the 157–167 region of the β_2 strongly interacts with the 375–388 region of the α subunit (Fig. 2B), which is a loop connecting the selectivity filter to the extracellular end of the inner helix S6. We suggest that their interaction is related to their physical connection to the opposite ends of the inner helix S6, which is the part of the permeation pathway that moves during opening and closing of the channel. This raises the issue of whether the gating of the channel in the β_2 - $K_v1.2$ complex could involve motions involving both the C-terminus and the β_2 subunit.

An examination of the 3D structure (Fig. 4A) suggests that if we postulated that both the C-terminus and the N-terminus of $K_v1.2$ were associated with the same monomer of β_2 , the path of the C-terminus of $K_v1.2$ to the proposed docking site in the associated β_2 subunit would be sterically hindered by the presence of the T1 region. The extent of hindrance is even more marked if we consider the movement of the C-terminus between the open structure and the putative closed structure (Fig. 4B) from the Rosetta models.³⁷ On the other hand, we note that the end of S6

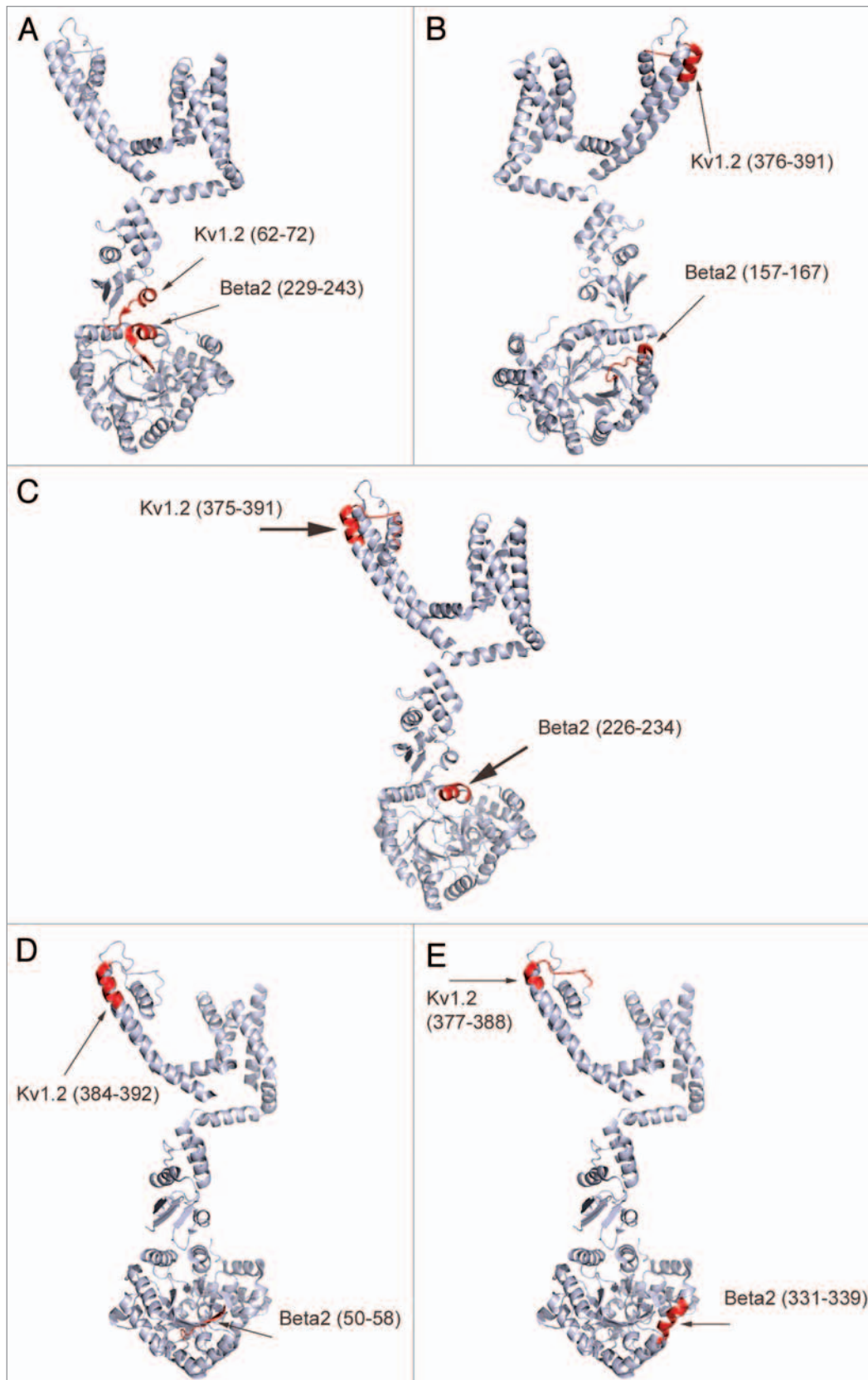
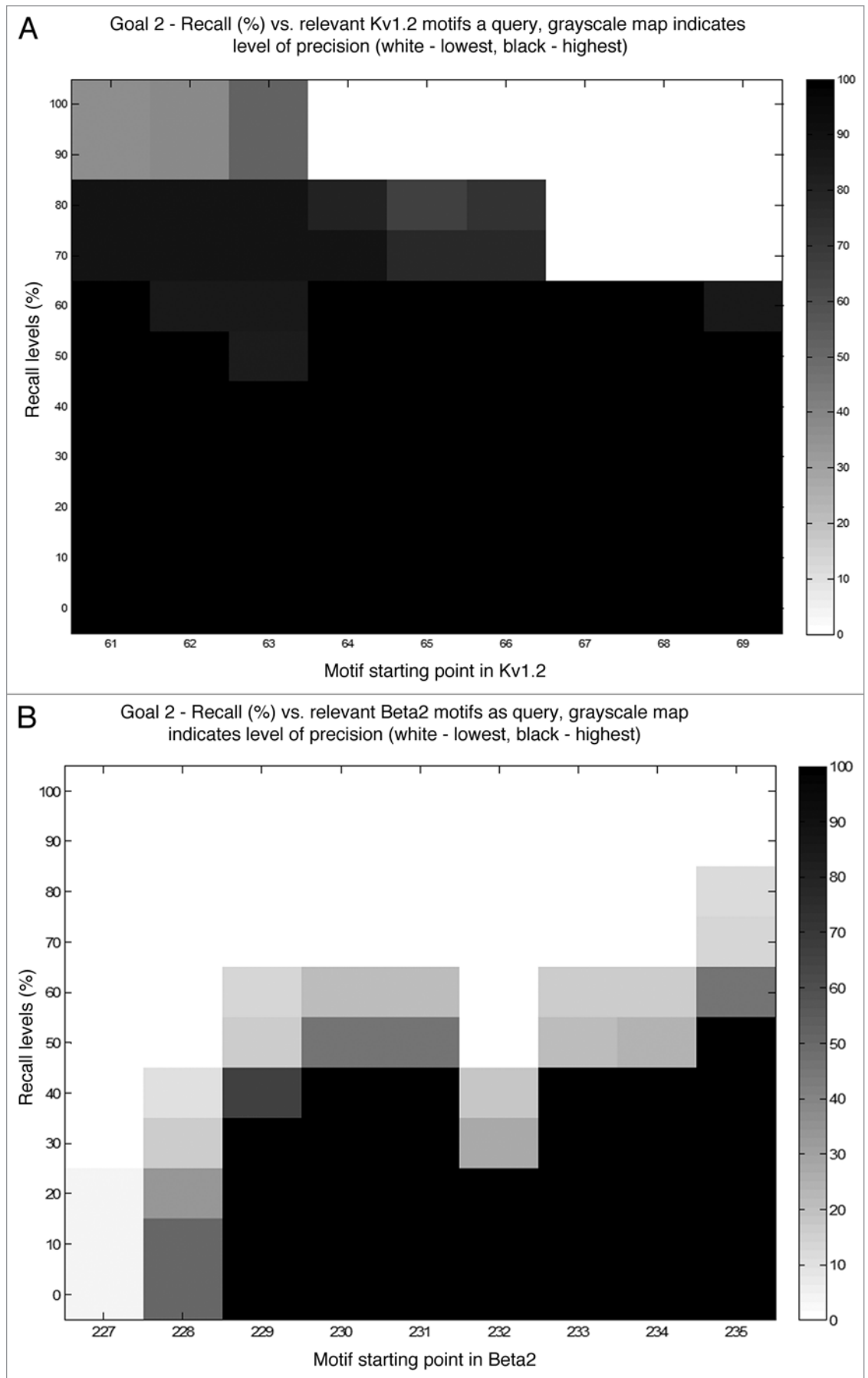


Figure 2. The top 30 most correlated pairs can be grouped by merging overlapping motifs in each protein into five sets of highly correlated regions between $K_v1.2$ and β_2 . These regions are shown on the structure. (A) Region corresponding to the interaction interface in $K_v1.2$ (62–72) and β_2 (229–243). (B) β_2 region (157–167) and $K_v1.2$ (376–391). (C) β_2 region (226–234) and $K_v1.2$ (375–391). (D) β_2 region (50–58) and $K_v1.2$ (384–392). (E) β_2 region (331–339) and $K_v1.2$ (377–388).

Figure 3. Image of Precision levels for various recall levels and motifs used as queries. The x-axis represents the starting point of motif used as query (to the correlation engine that identifies the interacting partners), the y-axis represents the 11 standard recall levels (in percentage) and the grayscale level corresponding to each cell represents the precision at a specific recall level for the specific motif. The grayscale level code, representing degree of precision, is shown on the right of the image. (A) $K_v1.2$ motifs as queries. (B) β_2 motifs as queries.



in the structure seems to point toward the adjacent β_2 monomer. In **Figure 4A**, β_2 monomer B1, is attached to $K_v1.2$ monomer K1, whose C-terminus is annotated. The β_2 monomer B_2 , is attached to $K_v1.2$ monomer K2, whose C-terminus and points away from the docking site on B_2 . Based on our argument, the N-terminus of K1 interacts with B1 while the C-terminus of K1 interacts with B_2 .

Figure 4C shows a schematic depicting the putative intermonomer docking. This figure suggests that during gating the “attachment points” between the $K_v1.2$ and β_2 remain fixed and that the β_2 rotates. Since experimentally the $K_v1.2$ can gate without the presence of the β_2 ,^{38,39} and since the voltage sensor is contained in the $K_v1.2$, we postulate that in ordinary voltage-dependent gating, the movement of the β_2 follows the gating, rather than driving it. However, the coupling to the β_2 provides a mechanism for other factors to modulate the gating. The open and closed structures of the $K_v1.2$ channel from the Rosetta model (**Fig. 4B**) suggest that the C-terminus of $K_v1.2$ moves inward toward the pore during channel closing while the N-terminus also has a slight movement, possibly in a direction counter to that of the C-terminus region. The circular arrows in **Figure 4C** suggest a rotation of the entire β_2 unit during gating, with the C-terminus of the $K_v1.2$ moving towards the center of the complex on closing, and towards the periphery of the complex on opening. We note that the 157–167 regions in different monomers of β_2 are likely to come close to each other upon rotation to the closed state (**Fig. 6**). Specifically, a glutamate in this region, from each monomer, is likely to face the pore upon rotation. We also note that movement of the β_2 implies movement of the T1 region of the $K_v1.2$ during gating, which has been postulated previously.⁴⁰ Assuming a rigid connection between T1 and the β_2 , a rotation to the T1 domain would force buried polar residues toward the channel axis, including a completely conserved glutamate. This tetrameric glutamate site is reminiscent of the postulated selectivity filter for calcium channels.⁴¹ In this case we suggest that these glutamates provide an attractive interacting site for calcium ions, explaining the role of calcium ions in facilitating slow inactivation in the $K_v1.2$ - β_2 complex.⁴² We note that the above picture depends on the hypothesis that the interface between the $K_v1.2$ and β_2 is unchanged during gating, a view supported by the correlation between the $K_v1.2$ region 62–72 and the β_2 region 229–243. There are a number of structural experiments that might be devised to test the above hypothesis. One would be structure determination in high calcium, to test the hypothesis that the calcium inactivated structure involves a cage of glutamates as suggested above. Also, structure determination has been achieved for another potassium channel mutated to stabilize the voltage sensor at different configurations.⁴³ Perhaps this approach could be extended to understand the interaction across the $K_v1.2$ - β_2 complex during gating.

While the rotation as a unit suggested in **Figure 4C** is conceptually the simplest motion to postulate for β_2 , it is possible that β_2 changes conformation in some more complex manner. We note that the interaction between $K_v1.2$ and β_2 is “polarized”,⁴⁴ in that the N-terminal region of the $K_v1.2$ interacts with the C-terminal core of β_2 and the C-terminal region of $K_v1.2$ interacts with the N-terminal core of β_2 . Therefore, a relative motion of the N and

Table 2. The motifs possibly involved in interaction on the $K_v1.2$ and β_2 subunits

Starting and ending residue positions ($K_v1.2$ subunit)	Starting and ending residue positions (Beta subunit)
61–69	227–235
62–70	228–236
63–71	229–237
64–72	230–238
65–73	231–239
66–74	232–240
67–75	233–241
68–76	234–242
69–77	235–243

These form part of the *Structurally Relevant Set*. Starting and Ending positions of motifs are shown.

C-termini of $K_v1.2$ during gating could imply a relative motion between the N-terminal region and the C-terminal region of β_2 , which could occur via a hinge region that might lie in-between. This hinge will be discussed in more detail later in this paper.

The 197–209 region of β_2 —a possible false negative. Motifs in the 197–209 region of β_2 (**Fig. 5A**) are close to the $K_v1.2$ interaction interface, and indeed are structurally close to two different $K_v1.2$ monomers, but show only weak correlation to the motifs physically close to them in $K_v1.2$. Since the motifs are close to each other, we hypothesize that the strongest interactions that govern the composition of this region are between different monomers of β_2 . **Figure 5B** shows in detail the spatial relationships between these motifs in different β_2 monomers. An almost fully conserved phenylalanine at position 205 serves as a marker. The only other residue that appears at 205 is a tyrosine, so the aromatic property at 205 is fully conserved. It can be seen that the motifs in question are part of helices that point almost directly at each other from adjacent β_2 monomers. The direction of the helices is such as to create a strong electronegative region that might, in vivo, contain an ion. It might also be that the conserved phenylalanine can provide a stabilizing ring stacking environment when the channel is closed. (In **Fig. 5B**, the phenylalanines are not stacked, but that represents the open structure only). Pi-ring stacking could lend considerable stabilization to the closed state.^{45,46} Another possibility is the likelihood of a cation-pi interaction with positive ions that might be found in that region, especially since the helix dipole of motif 197–205 is pointed toward the channel axis; this makes it highly attractive to positive charges (possibly K^+ ions). Interestingly, in our autocorrelation analysis of the β_2 subunit (Suppl. S3), the motifs starting from 202 through 212 are highly correlated not with other motifs within β_2 , but only with each other. We cannot determine if these high correlations are because of contiguity in sequence space or because they come together in 3D space from different monomers. On the other hand, these motifs do show fairly strong correlation with motifs in the pore region (Complete correlation results, as shown in Suppl. material Dataset S1). We note that the same rotation of both the T1 and the β_2 can produce negative charge concentration near the channel axis. Overall, there is a high likelihood of a positive ion

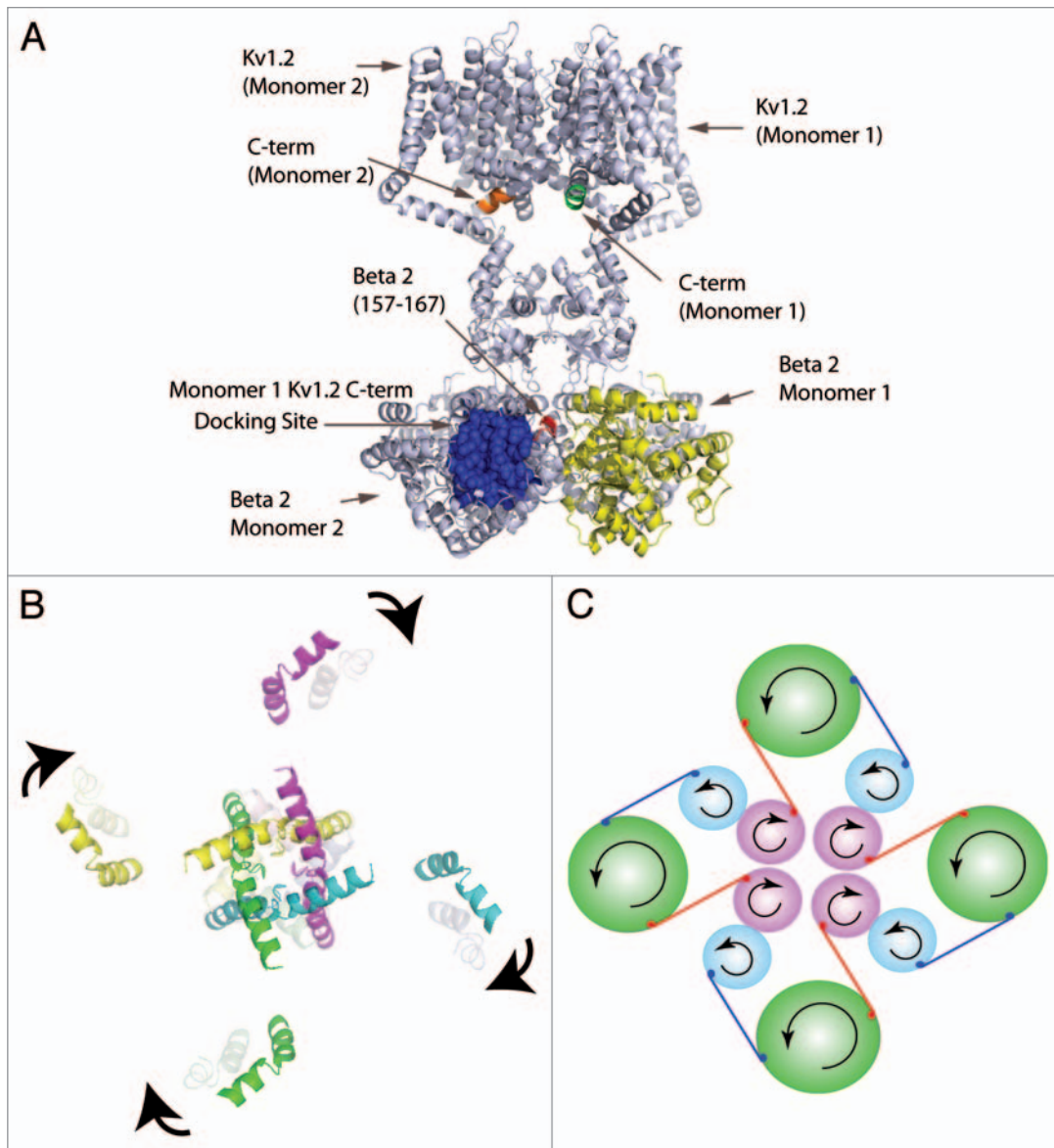


Figure 4. Interaction of $K_v1.2$ C-terminus with β_2 . (A) Figure showing the tetrameric protein with both $K_v1.2$ and β_2 . The relevant regions shown are: the possible docking site for the $K_v1.2$ C-terminus (based on a visual comparison with Fig. 7 and reviewed in ref. 36), the 157-motif that is close to this docking site and is one of the β_2 regions in Table 1, the end of the C-terminus for the $K_v1.2$ monomer (numbered 2), the end of the C-terminus of the adjacent $K_v1.2$ monomer (numbered 1). $K_v1.2$ monomer 1 is attached to the β_2 monomer 1. Our hypothesis is that the $K_v1.2$ C-term from monomer 1 docks with the adjacent β_2 monomer 2 at the indicated docking site. (B) Open and closed structure of the $K_v1.2$ subunit (tetrameric) from the Rosetta models. The view is from the intra-cellular side. The central region represents the S6 helices from the 4 monomers and the distant helices represent the S1 helices from the 4 monomers. The shadowed helices represent the closed state while the solid helices represent the open state. The direction of arrows indicates the overall movement of the S6 and S1 helices in going from the open to the closed state. The figure was constructed by superimposing the structure in both states, with the selectivity filter serving as a reference. (C) Schematic showing the proposed interaction points between $K_v1.2$ and β_2 and the rotation scheme associated with $K_v1.2$ movement during gating. Shown is the view from the extra-cellular region. The large circle represents the β_2 subunit. The circle with anticlockwise arrows represents the S1 helix (attached to the N-terminus) and the circle with clockwise arrows represents the S6 helix (attached to the C-terminus). The straight lines represent the connections to the appropriate β_2 monomer. The figure represents the open state, the direction of the arrows show the direction of rotation for each region upon channel closing.

entering the central pore (formed by the β_2 monomers) from the cytoplasmic side (in the open state) and stabilizing the rotated conformation. This could contribute to a form of inactivation, which might involve the $K_v1.2$ pore region as well. Slow inactivation and the slow recovery from inactivation⁴⁷ could involve the coupling of conformational changes between $K_v1.2$ and β_2 and

the presence of a stabilizing cation. These results also support the idea of a closed-inactivated state as proposed in recent work.⁴⁸ The authors of this work suggest that $K_v1.2$ shows minimal or no inactivation in the absence of the beta subunits. They further outline a series of residues⁴⁸ that could be important in the slow inactivation of $K_v1.2$. In our alignment, all of these residues are

completely conserved and hence invisible to evolutionary correlation analysis; however most of these residues are near to, or in the, pore region. Our hypothesis is that the coupling with the β_2 subunit can induce slow inactivation in the overall complex by the mechanisms outlined above. We hypothesize that the inter-monomer coupling of $K_v1.2$ and β_2 , by allowing for cooperativity, can accelerate the long time scale that might be required for such a large-scale conformational change. We note that cooperativity has been implicated in potassium channel gating.⁴⁹

The 226–234 region of β_2 . The 226–234 motif in β_2 has major overlap with the β_2 motifs that are part of the structurally relevant motif pairs (Table 2). Note that these structurally relevant motif pairs define the structural interface between the N-terminus of the $K_v1.2$ and the β_2 subunit. As discussed previously, we saw that the β_2 region that is likely to form the interface with the C-terminus of the $K_v1.2$ has a strong correlation with the 375–388 pore region from $K_v1.2$. From Table 1 we see that the 226–234 region of β_2 is also very strongly correlated with the pore region motifs starting at 376, 382 and 383 in the $K_v1.2$ subunit (Fig. 2C). In the context of elucidation of the control network, the full import of Figure 2 is that there are two types of functional domains in the β_2 protein. One is critical for the complexation with $K_v1.2$ and that is 229–243, which has coevolved with $K_v1.2$ 62–72 to preserve the favorable contact, as shown in Figure 2A. The other type of functional domain has to do with the function of β_2 as an oxidoreductase, which is functionally (although not directly by structural contact) correlated with the extracellular end of $K_v1.2$, as shown in Figure 2B–E. Further analysis, summarized in Figure 9, will indicate that this second interaction is effected by phosphorylation mediated by Protein Kinase-C and Tyrosine Kinase. The partial overlap of two types of functional domain, as β_2 229–243 (Fig. 2A and complexation function) and β_2 226–234 (Fig. 2C, oxidoreductase function), is unusual but not unknown. For example it is well established that the activation and degradation domains overlap in some transcription factors.⁵⁰

Possible hinge motions in β_2 . From the 3D structure, we can see that parts of the β_2 interaction interface and the 197–205 region of β_2 lie on the same plane. It is possible that rather than an entire rotation of the β_2 subunit, these regions move relative to each other, e.g., a tilt of the 197–205 region toward the β_2 interaction interface. Such a tilt would bring the phenylalanine rings almost parallel to the channel axis and line them up like hoops, in a circular fashion across monomers. However, any relative motion between these regions requires a hinge to be present. Since β_2 belongs to a family of aldo-keto reductases, we compared the sequence to AKR1B1—an aldo-keto reductase where putative hinge regions were identified.⁵¹ Our sequence alignment (Dataset S2) showed that the hinge regions of AKR1B1 that are implicated in the binding of the cofactor correspond to the 258–278 region of β_2 . This region is far away from the $K_v1.2$ interaction interface. We further examined the β_2 structure for hinges using the online

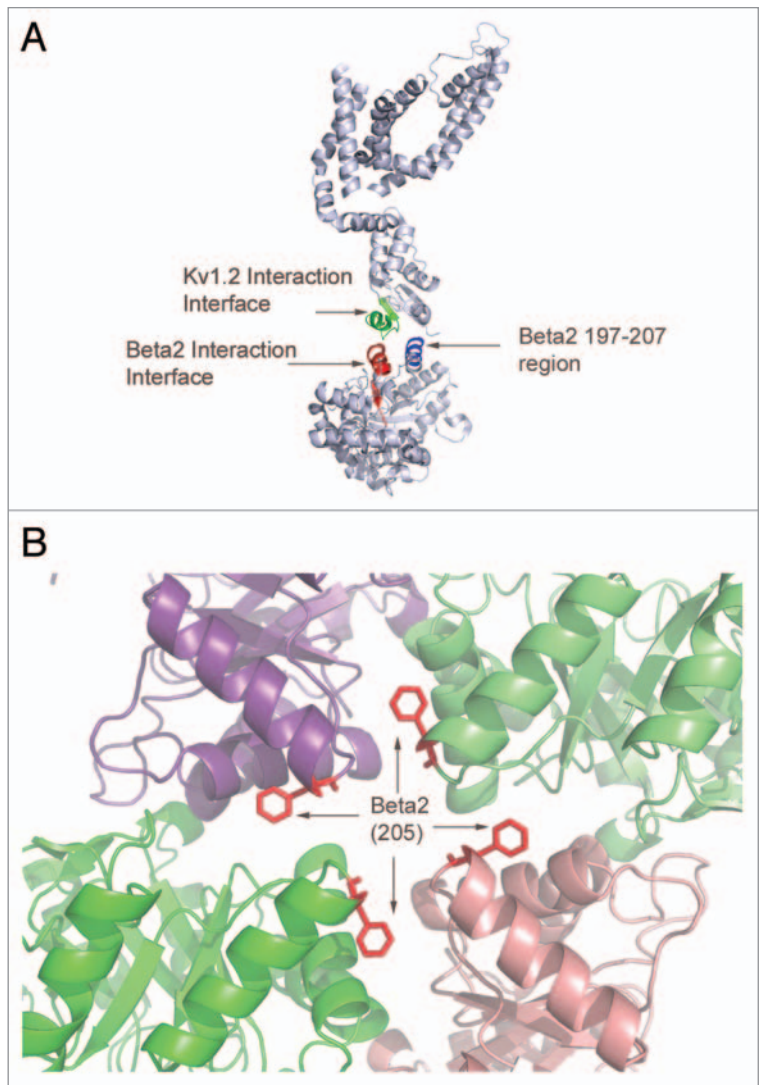


Figure 5. (A) Monomer structure showing the “possible false negative”, the 197–207 region in β_2 . For reference, the interaction regions of $K_v1.2$ and β_2 are also shown. (B) Phenylalanines at position 205 in the beta monomers. An inward anticlockwise rotation will bring the rings close to each other. View is from the extra-cellular end.

hinge prediction program from the Database of Macromolecular Motions.⁵² We found that the most likely hinge region surrounds a completely conserved proline at 210 that lies between the 197–205 motif and the β_2 interaction interface and below the plane of those regions (Fig. 7). This could explain a possible relative motion of the two regions. The attractiveness of the hypothesis suggested by such a relative motion is that it provides a possible connection between voltage-dependent gating by $K_v1.2$ and the binding and release of messenger molecules, such as NADPH, by β_2 (as the conformational change of β_2 resulting from gating would no longer provide an effective binding pocket for NADPH/NADP⁺).⁵³

Other “structural false positives”. The final set of “structural false positives” (correlations without obvious structural significance) is the correlations of the β_2 motif 331–339 (Fig. 2E), with the $K_v1.2$ pore region and the β_2 motif 50–58 (Fig. 2D)

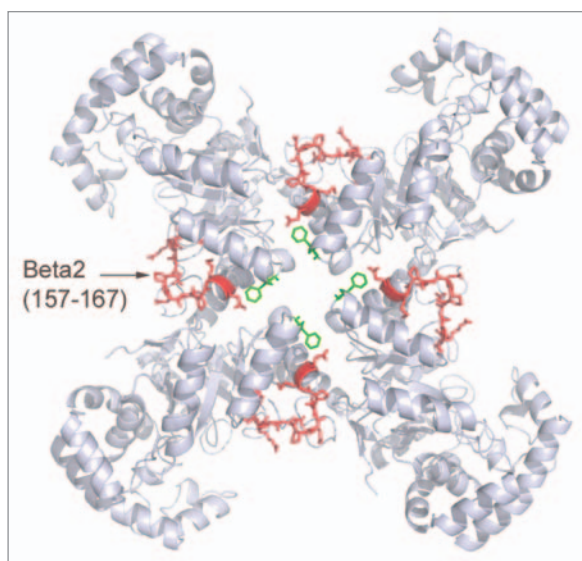


Figure 6. The 157–167 region from each monomer. An anticlockwise rotation will bring these regions from each monomer close to each other. View is from the extra-cellular end.

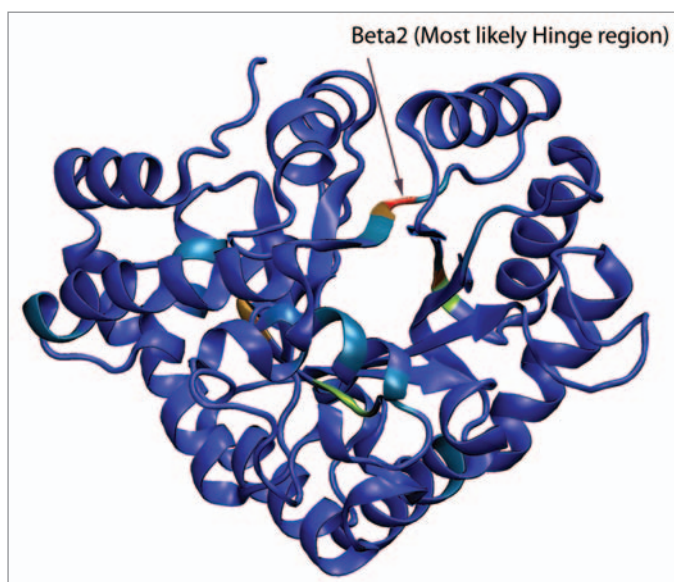


Figure 7. Hinge prediction by HingeMaster for the β_2 subunit. The location of the most probable hinge is shown.

with the $K_v1.2$ pore region. The motif starting at 50 includes residue 56 that is proximal to the $NADP^+$ molecule as shown in the β_2 structure.⁵³ Residue 333 in β_2 was shown to be critical for $NADPH$ binding.⁵⁴ Therefore we suggest that this set of correlations may be relevant in the context of the proposed coupling between the redox state of the cell and the electrical activity of the membrane.⁵³

Emergent properties of the correlation network. In addition to computing correlations between the $K_v1.2$ and β_2 motifs, we also carried out an autocorrelation analysis of the $K_v1.2$ motifs and the β_2 motif sets. The correlation matrices can be interpreted

as a graph, where nodes represent β_2 and $K_v1.2$ motifs with interconnecting edges, the edge weights corresponding to the degree of correlation.

Without imposing cutoffs on edge weights, our network will be very dense and it will become hard to analyze its properties. For our analysis, we imposed a cutoff of 0.9 for correlations between proteins and 0.95 for intra-protein correlation. Since full-length protein alignments were shown to have a correlation score of 0.8 to interact,¹⁴ we chose a threshold value of 0.9 for the motif level correlation, as motifs are expected to be a more appropriate interaction site and therefore have a higher degree of evolutionary coupling. (The full-length sequence correlation of $K_v1.2$ and β_2 was ~ 0.84). We also imposed a p-value cutoff of 0.05. This makes the network more manageable. In continuing our analogy with the Internet, we decided to streamline the network by adding directionality to the edges. Note that correlation is a symmetric property; however our p-value analysis when centered on a motif, is directional. That is to say that motif A might interact with motif B significantly, but the converse may not be true. This would be the case if motif B were promiscuous or non-specific. Sorting the correlations for each motif and taking only the highest correlations will also achieve directionality. Thus, for each node, we identified the motif that it was most correlated to (within significance level), both on the same protein and on the interacting partner. This represents the outgoing edge for the node. It is possible that there are several outgoing edges for each node with same degree of correlation; we retain all of these in our analysis. Similarly, each node can have several incoming edges. Borrowing from Internet terminology,⁵⁵ we call nodes with a high degree of incoming links as *authorities*. Intuitively, the biological relevance of authority is that a mutation in that region would necessitate changes in a large number of motifs that are functionally coupled to it. Datasets S3-S4 contains autocorrelation results for $K_v1.2$ and β_2 respectively. Datasets S5-S8 contains *authority* analysis for $K_v1.2$ w.r.t β_2 , $K_v1.2$ w.r.t $K_v1.2$, β_2 w.r.t $K_v1.2$ and β_2 w.r.t β_2 , respectively.

Figure 8A and C show the authority profiles in terms of incoming edges for $K_v1.2$ and β_2 respectively, for both the intra-protein autocorrelation (continuous line) and the inter-protein cross-correlation (dotted line). The authority profile for a protein is the graph of the number of other motifs that find a given motif (along the X-axis) as the highest correlated. For inter-protein correlation (dotted line in **Fig. 8A**), the most authoritative region in $K_v1.2$ is the motif starting at 376, which is the linker between the selectivity filter and the extra-cellular end of S6. A closer examination reveals that if we collapse all the overlapping regions in the neighborhood, then the region whose starting points range from 375–384 are highly authoritative (**Fig. 8B**). In the β_2 subunit's inter-protein authority profile (**Fig. 8C** and dotted line), the highest authorities are motifs starting at 50, 141 and 235 (**Fig. 8B**). If we collapse overlapping regions, these regions remain separately, highly authoritative. The region around 235 is the β_2 interaction interface. As discussed earlier, the motif starting at 50 includes residue 56 that is proximal to the $NADP^+$ molecule as shown in the β_2 structure.⁵³ The motif starting at 141 is pointing toward the cytoplasm. Although there is no clear

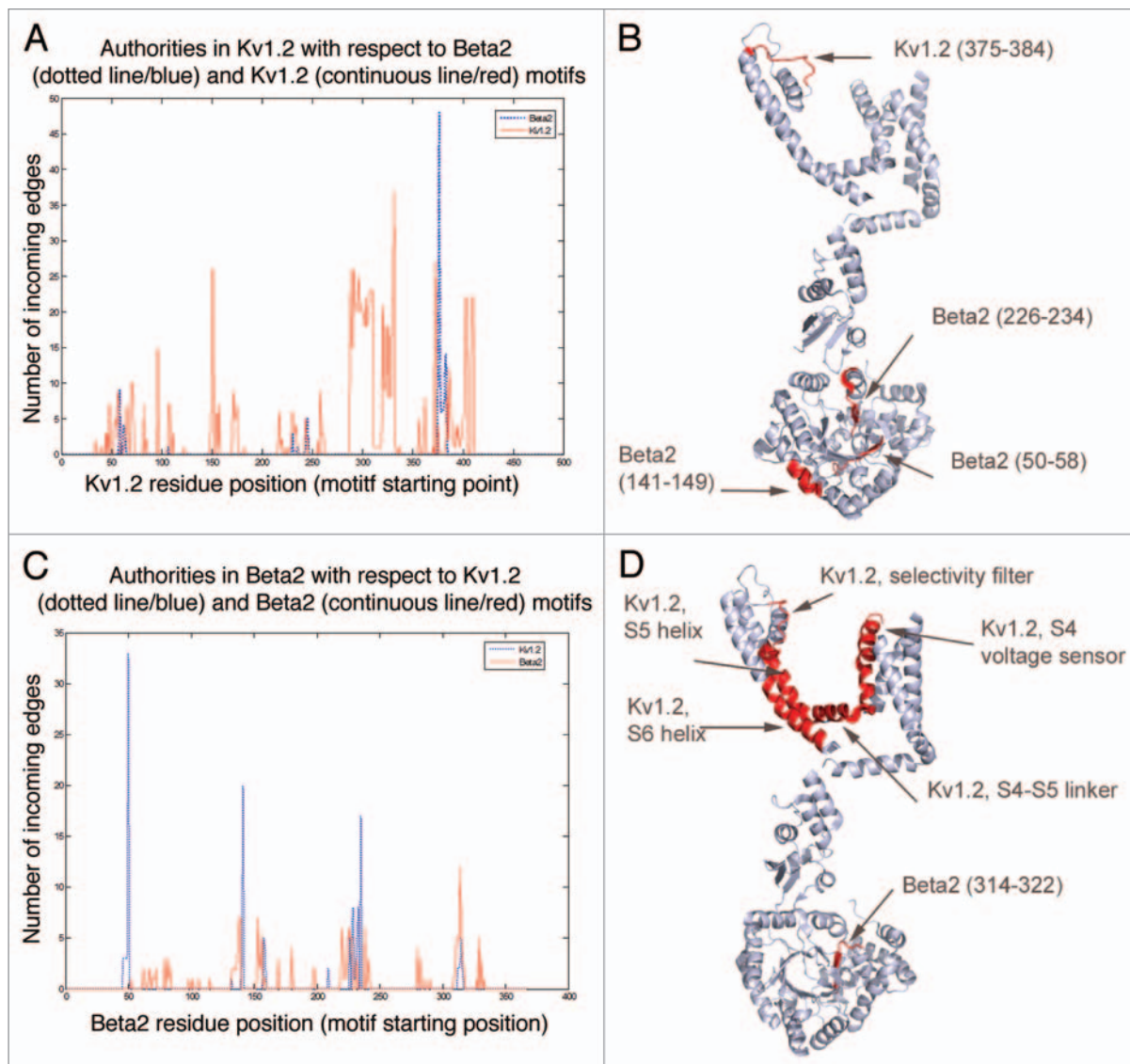


Figure 8. (A) Authority profile—number of incoming edges for $K_v1.2$ from the inter-protein correlation (blue/dotted line) and the autocorrelation analysis (red/continuous line). (B) Authorities from the inter-protein correlation projected on to the 3D structure for both $K_v1.2$ and β_2 . (C) Authority profile—number of incoming edges for β_2 from the intra-protein correlation (blue/dotted line) and the autocorrelation analysis (red/continuous line). (D) Authorities from the intra-protein correlation projected on to the 3D structure for both $K_v1.2$ and β_2 .

evidence in the literature on its function, we observe that there are some glutamates that stick out from this region and could possibly be charged. These negative charges could facilitate entry of cations into the channel from the cytoplasm. The 141–149 motif is examined later in this paper.

The authorities from the autocorrelation analysis for β_2 and $K_v1.2$ are quite different from the authorities from the inter-protein analysis. This implies that the evolutionary coupling within proteins and across proteins are different, presumably because intra-protein correlation includes a component focused on preserving a stable structure in addition to functional significance and relaying allosteric activity throughout the protein. In the β_2 protein, the authoritative region is the 314-motif (Fig. 8C, red/continuous line and D). This does not seem to have an immediate function, nor has it been implicated in cofactor binding.

We looked at the incoming edges to motifs in this region and its immediate neighborhood. Most of the links are from the β_2 interface region. Interestingly, most of the outgoing links from the 314-region is also pointing to the β_2 interface region. There must be a degree of structural coupling as these regions are somewhat close by in the 3D structure; more experiments need to be carried out to verify this connection. The authorities in the $K_v1.2$ autocorrelation (Fig. 8A, red/continuous line and D) reveal an interesting story. Regions 150 and 151 that have about 26 incoming edges are in the cytoplasmic region just before the start of S1. This region might play a role in the movement of K^+ ions through side-portals,² but precision in thinking about this is limited by lack of structural data from this region. Regions from 288–318 and 320–339 also show a high degree of incoming links. These include the voltage sensor S4 region, the S4-S5 linker and the

lower half of the S5 helix. Region 372 is the selectivity filter region. The region from 402–418 is the lower part of S6 that forms the gate. All these regions are extremely important for the overall functionality of the channel and their identification as authorities suggests that evolutionary pressure at these regions necessitate the coevolution of the rest of the structure. The turret region, which lies between the end of S5 and the pore helix, does not show very high correlations either within the $K_v1.2$ subunit or with the β_2 subunit motifs. This is not surprising as the turret regions are also under evolutionary pressure⁵⁶ to coevolve with toxins that are known to bind these regions and disrupt channel function. In this case, however the coevolution is to avoid binding to the toxins, rather than to preserve the interaction.

We notice that regions that are proximal across monomers in both $K_v1.2$ and β_2 are not highly correlated with each other. The only exceptions are the pore regions and the 204-regions that are close to themselves from other monomers. Some of the motifs that are very close to each other at the inter-monomer interface of the β_2 subunits do not show the strongest correlations either. For example, the motif centered on residue 132 is very close to the motif centered on residue 46, yet there is weak correlation between them. These regions however show a high degree of conservation. We hypothesize that evolutionary correlation within and between proteins will tend to be more prominent in regions whose primary function is to mediate coordinated function, while strong conservation will tend to be more prominent in regions whose primary function is to confer structural stability.⁵⁷ These two functions are not completely orthogonal; it may be that conservation to achieve a stable foundation will be followed by correlated substitutions to allow for functional variability. In this regard, note that the interface regions of the $K_v1.2$ and the β_2 , which are highly correlated, are important not just for structural stability, but can also function as allosteric communicators. In the family of potassium channels, a large number of channels are expected to have the Shaker-like fold and yet they have variations in their overall functionalities. This hypothesis has been explored under the theme of stability promoting evolvability, as shown in cytochrome P450 proteins.⁵⁸

We have also done analysis on correlation of the $K_v1.2$ and β_2 with other proteins implicated in interactions with Shaker potassium channel protein complex. The results of this analysis are summarized in **Figure 9**. Some of the components of **Figure 9** should be regarded as speculative; i.e., the weight of the evidence plus our analysis supports the relationships shown but in some cases the evidence is suggestive rather than conclusive and invites further experimental and computational confirmation. The steps in the analysis are described in the Methods Section and repeated in outline as follows: We picked particular proteins to identify MoLFunCs, based on their having a definitive annotation. Then, we picked from that group proteins to subject to correlation analysis based on having the same species distribution of MoLFunCs as $K_v1.2$ and β_2 . We selected random motifs from this set of proteins and computed correlations between all the $K_v1.2$ and β_2 motifs and this set of randomly selected motifs. These correlations provided a baseline for assessing the significance of the inferred β_2 - $K_v1.2$ interactions.

In the course of analyzing those results we observed that the 141-motif in β_2 (score ~ 0.96) and an N terminus region in $K_v1.2$ (score ~ 0.93), are both highly correlated to a G-protein beta subunit homolog. The homolog—GN_{B2}L1, also called RACK1, is known to be an adaptor protein for Protein Kinase-C (PKC).⁵⁹ Our computed correlation between RACK1 and $K_v1.2$ and β_2 appears as a diamond with the text “SC” in **Figure 9**. Both PKC⁶⁰ and RACK1,⁶¹ have been implicated in the cellular response to hypoxia, which makes these correlations relevant in the context of the fact that $K_v1.2/K_v\beta$ complexes have been found to have oxygen-sensing mechanisms.^{9,10} Yet another protein that we found β_2 141-motif to be highly correlated (score ~ 0.96) to is a translation initiation factor interacting protein (Eif3s6ip). In addition, we found that the region on $K_v1.2$ that correlates with RACK1 also correlates with Eif3s6ip (score ~ 0.97). Eif3 can interact with Eif2⁶², which has been implicated in hypoxia response;^{63–65} one of the early events in this response is the phosphorylation and inhibition of this factor. These correlations between the Eif3s6ip on the one hand and β_2 and $K_v1.2$ are shown as diamonds with the text “SC” in **Figure 9**.

Both $K_v1.2$ and β_2 interact with kinases and the $K_v\beta$ interaction itself is influenced by phosphorylation.⁶⁶ Specifically, β_2 subunits are known to be phosphorylated by PKCs.^{67,68} (Shown as a diamond with text “P” in **Fig. 9**). There is evidence that G-protein-coupled receptors in general and the muscarinic acetylcholine receptor in particular, can interact with tyrosine kinases that are involved in modulation of $K_v1.2$ activity.⁶⁹ Specifically, the C-terminus of $K_v1.2$ has been implicated in this response.⁶⁹ It has been shown that tyrosine kinase and PKC have a combined inhibitory effect on $K_v1.3$ currents.⁷⁰ This chain of interactions is shown as diamonds with text “A” for activation, “P” for phosphorylation and “CS” for current suppression, in **Figure 9**. Although this combined effect has not been specifically reported for $K_v1.2$ (which is also a delayed rectifier like $K_v1.3$), the inhibition of $K_v1.2$ current by tyrosine kinases was studied in the context of downstream effects of the M1 muscarinic acetylcholine receptor.⁶⁹ It is known that M1 muscarinic acetylcholine receptors can activate both PKCs and tyrosine kinases through second messengers.^{69,71,72} The observed inhibitory effect of $K_v1.2$ by tyrosine kinases⁶⁹ in this context, could be due to the presence of activated PKCs as well, and will be in agreement with the combined inhibitory effect observed for $K_v1.3$. Interestingly, RACK1 can also function as an adaptor for tyrosine kinases.⁷³ From the totality of these observations, we infer it as likely that $K_v1.2$ and $K_v1.3$ have common interaction modes with tyrosine kinase and PKC and that the data on both can be federated with respect to this particular interaction. We indicate that on **Figure 9** by connecting $K_v1.2$ and $K_v1.3$ with a rectangular shape.

It has also been shown that potassium channel phosphorylation could influence the rate of C-type inactivation.⁷⁴ We found that one of the highest correlations of the RACK1 protein with the K^+ channel is in the $K_v1.2$ interaction interface, which we indicate with a diamond with text “SC” in **Figure 9**. We note that RACK1 would not have direct access to the $K_v1.2$ interaction interface due to complexation of $K_v1.2$ with β_2 . In the absence of accessibility to these sites, phosphorylation

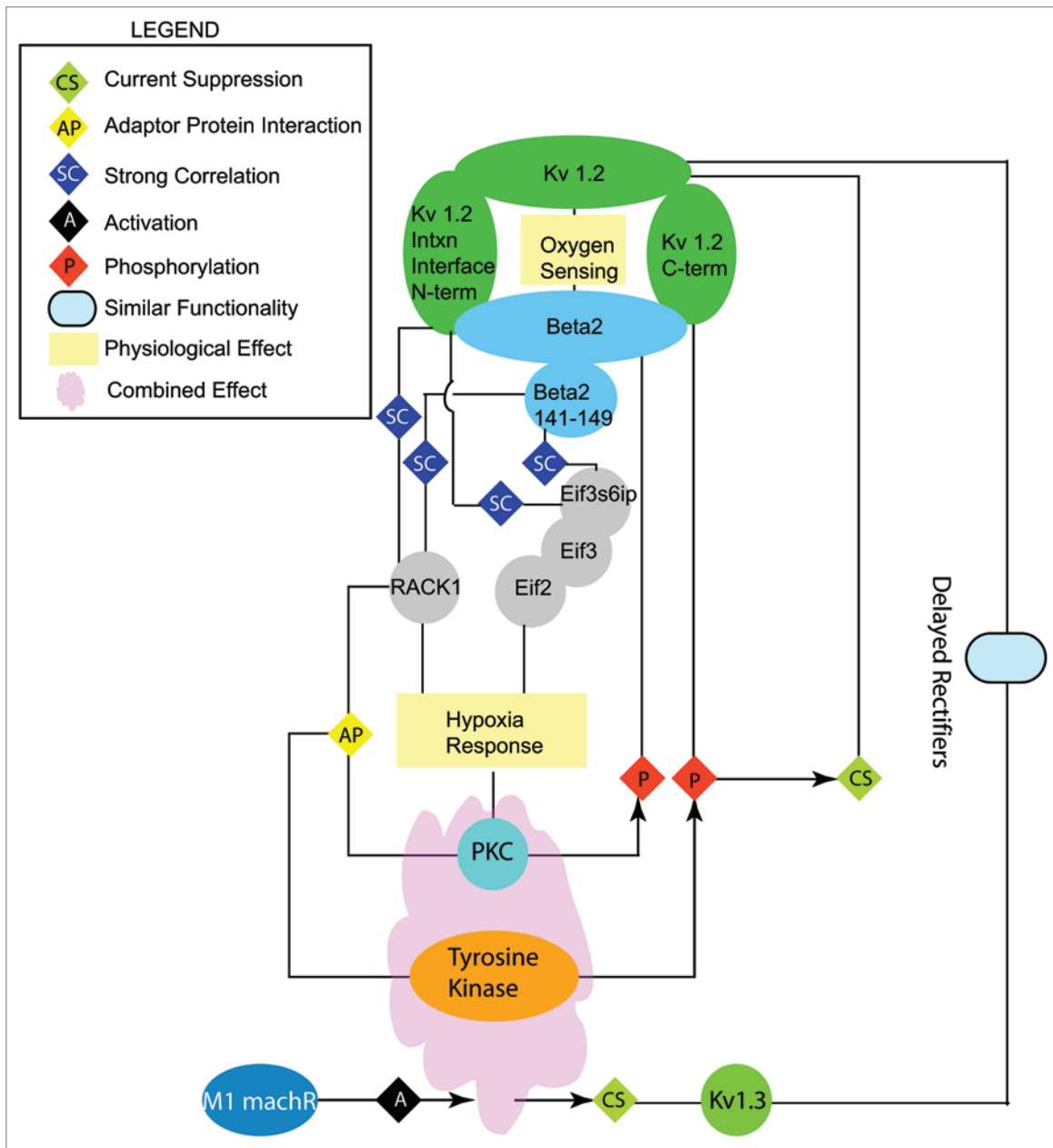


Figure 9. Summary of correlations, literature-based evidence for interactions/functional linkage and structural evidence, for the β_2 141–149 region and its overall implication. The legend shows the meaning of the symbols. The molecular entities are shown in ellipses or circles and physically interacting motifs/proteins are shown as overlapping. The direction of the arrows in some of the connections indicates the direction of cause and effect.

of β_2 residues by protein kinases could provide modulation of inactivation kinetics (indicated by a diamond with text “P” on Fig. 9). It is also possible that β_2 provides the docking site for the RACK1 protein and promotes phosphorylation of the $K_v1.2$ C-term residues. This possibility is indicated by a diamond with text “P” from the tyrosine kinase to the $K_v1.2$ C-terminus in Figure 9. Although it is not immediately evident from the existing (incomplete) structure, due to lack of resolution in that region, the $K_v1.2$ C-terminus could be somewhat proximal to the β_2 141–149 motif, with sufficient room for both adaptor protein and kinase. We postulate that the region around and

including motif 141–149 could be a binding site for molecules that couple signaling events with electrical activity and that the overall coupling of the $K_v1.2$ - β_2 complex is significant for hypoxic response. Indeed, it has been suggested that the oxygen sensing apparatus of the potassium channels might be mediated through redox sensors in the cytoplasm.⁷⁵ The fact that the β_2 subunit is homologous to oxidoreductases, has a NADP⁺ bound in the crystal structure and has catalytic activity as demonstrated recently,⁷⁶ points to such a mechanism. Recent work^{7,77} has identified NADPH (bound tightly to the β_2) as being the primary species (as opposed to NADP⁺) responsible for inactivation of K_v channels. The paper⁷ referred to $K_v1.5$ - $\beta_{1,3}$ interactions and

suggested that the observed inactivation might be mediated by the N-terminus of $\beta_{1,3}$, which is not present in β_2 . However the finding is still consistent with the proposed rotation of the β_2 subunit in our earlier discussion. Recall that the rotation can cause high electron density at the center of the pore, which argues for the possibility of a positive ion there. A positive ion acts as an electron sink. It is likely that the NADP⁺, which is electron deficient, might not provide the extra stability for the interaction at the center of the pore, whereas the electron rich NADPH can provide extra stability. This might be a cause for the inactivation to be weaker or absent when NADP⁺ is bound as opposed to NADPH. The mechanistic view presented here is very relevant to the role of the $K_v1.2$ - β_2 as oxygen sensors. In the hypoxic state, NADPH (reduced state) is likely to be the dominant species. It is possible that the RACK1 and the kinases interacting with the $K_v1.2$ and β_2 subunits might facilitate structural rearrangements (using the hinge) allowing for NADP⁺/NADPH exchange, thus ensuring that the bound species is the reduced form (NADPH) to inactivate the channel. The positive ion that we think is most likely involved in inhibition of $K_v1.2$, is Ca²⁺. Post et al.⁴² showed that release of calcium from intracellular stores can contribute directly to inhibition of $K_v1.2$ current under hypoxic conditions. This fits in with our hypothesis that Ca²⁺ might act as the molecular wrench that helps inactivate the current by interacting with the electronegative pore region created by the rotation of β_2 , which creates a structural motif similar to the EEEE motif postulated in the selectivity filter of calcium channels.⁴¹ Recent work on the interaction of NADPH with β_2 suggested that the mode of conversion of NADPH to NADP⁺ might be catalysis.⁷⁷ This might be the mode of recovery from inactivation as well when the cell responds to hypoxia.

Finally, we consider the impact on cell homeostasis of a reduction in resting potassium permeability according to the pump-leak model for cell osmoregulation.⁷⁸ According to that model, inhibition of the K channel will conserve energy in the short term, by reducing the passive ionic fluxes across the membrane and therefore also reducing the rate of active transport necessary to balance the passive leak. Thus potassium channel inhibition seems a useful component of the cellular response to temporary hypoxia. However the model also shows that a reduction in resting potassium permeability will cause deleterious cell swelling in the long term, with the rate of cell swelling being dependent on the permeability to chloride ions. Consistent with the pump-leak model, the chloride channel blocker DIDS rescues cultured neuronal cells from hypoxia-induced damage.⁷⁹

Discussion

We conclude that the evolutionary correlation analysis is very useful in uncovering functional coupling between and within proteins. Such coupling is not limited to structurally proximal motifs, but is also prominent among distant motifs participating in a distinct physiological response. We identified several regions that have not been experimentally probed; these could be critical not only for the $K_v1.2$ - β_2 interaction, but also couple signaling

events and cellular response to stress conditions such as hypoxia, to the electrical activity of the cell. We also proposed a series of conformational changes in the complex that might accompany channel opening/closing and identified regions that would be affected by these changes. We believe that more structural studies and experiments are needed to fully elucidate the overall coupling of $K_v1.2$ and β_2 ; the analysis presented here serves as a useful starting point for focused exploration of specific regions in this system and other K_v - β complexes.

A possible avenue for future exploration would be to look at the rates of evolution at the interface vis-à-vis the rest of the protein. Hub proteins would provide a good test bed for such analysis as they can bind to several other proteins at different sites and it would be interesting to identify the degree to which coevolution at an interface propagates through the structure, especially in light of multiple competitive evolutionary pressures. In principle, the analysis of interacting motifs could be applied to simulation studies focused on binding/docking of proteins. The algorithm could be biased by potentials derived from the correlation scores, thus decreasing the search space and achieving complexation faster. A similar use of this strategy is in folding proteins or predicting 3-D structure, wherein the bias in potentials comes from the correlations inherent in the structure itself. There is some precedence to this idea; Ranganathan and coworkers⁸⁰ used the correlation scores to design new proteins that had the same fold as the proteins from which the analysis was derived.

We have tried to follow a discovery-driven approach and let the emergent patterns in the data relate the story. The only phylogenetic techniques that we use are implicit in the theory of substitution matrices and the alignment techniques. One of the limitations of our method is that it ignores gapped motifs. For example, in this paper, we did not consider a region in the β_2 N-terminus that is known to be essential for tetramerization,⁴⁴ because of the gaps in the alignment in this region. A thorough calibration of correlations in gapped regions and their relevance to function would provide a more complete picture of evolutionary correlation between proteins. Note that we did not use any substitution matrix in determining the distance; we counted either identities or differences. Although the chemical nature of substitutions would be relevant in proximal motifs, consideration of substitution matrices for distant motifs might be less critical as they do not directly interact. On the other hand, the use of substitution matrices might help distinguish between structural and functional coupling. Moreover, benchmarking across different substitution matrices in the context of correlated evolution might provide a basis to reevaluate some of the matrices.

We fully explored only one protein pair in this work; we envisage that the technique can be extended to whole proteome analysis as well. An evolutionary correlation technique has been applied at the full-protein sequence level rather than the motif level to the *E. coli* genome¹⁹ and the results were very promising. In addition to indicating the promise of the evolutionary correlation technique, our results also validate the earlier procedure³⁵ of collecting the Most Likely Functional Counterparts (MoLFuncs) of the proteins being investigated.

Table 3. MoLFunCs for β_2 and $K_v1.2$ used as input to the correlation analysis

Species	$K_v1.2$ MoLFunCs	Beta2 MoLFunCs
<i>Rattus_norvegicus</i>	NP_037102.1	NP_059000.1
<i>Mus_musculus</i>	NP_032443.2	NP_034728.2
<i>Homo_sapiens</i>	NP_004965.1	NP_003627.1
<i>Canis_familiaris</i>	ENSACFP00000029266	XP_858333.1
<i>Equus_caballus</i>	XP_001496518.1	XP_001497164.1
<i>Macaca_mulatta</i>	XP_001101557.1	XP_001090173.1
<i>Bos_taurus</i>	XP_588328.1	NP_001014405.1
<i>Monodelphis_domestica</i>	XP_001372701.1	ENSMODP00000019659
<i>Ornithorhynchus_anatinus</i>	XP_001508339.1	ENSOANP00000013127
<i>Gallus_gallus</i>	NP_989794.1	XP_001232971.1
<i>Xenopus_laevis</i>	NP_001079222.1	NP_001083842.1
<i>Xenopus_tropicalis</i>	ENSXETP00000038520	ENSXETP00000006332
<i>Gasterosteus_aculeatus</i>	ENSGACP00000006912	ENSGACP00000002211
<i>Oryzias_latipes</i>	ENSORLP00000008061	ENSORLP00000005422
<i>Danio_rerio</i>	XP_691312.1	XP_001334335.1
<i>Ciona_intestinalis</i>	ENSCINP00000007870	ENSCINP00000005438
<i>Ciona_savignyi</i>	ENSCSAVP00000011900	ENSCSAVP00000017319
<i>Drosophila_melanogaster</i>	CG12348-PE	NP_511104.3
<i>Apis_mellifera</i>	XP_391895.2	XP_624840.1
<i>Tribolium_castaneum</i>	XP_968511.1	XP_970125.1
<i>Aedes_aegypti</i>	AAEL00242-PA	AAEL006650-PA

Methods

Defining the sample sets. The MoLFunCs were determined as described in a previous paper.³⁵ Since the correlation analysis of motifs is based on the alignments of the MoLFunCs, we removed species that either caused gaps in functionally relevant regions or caused gaps in regions that would otherwise be well aligned. The final list of species and MoLFunCs for β_2 and $K_v1.2$ is shown in Table 3. We aligned both sets using the MUSCLE⁸¹ program. This alignment forms the input for the motif extraction and the correlation analysis steps. Datasets S9 and S10 contain the $K_v1.2$ and β_2 alignments, respectively.

Extracting relevant motifs. We are concerned with finding correlations between motifs rather than the entire protein sequences. One technique of finding motifs is through standard motif databases such as PFAM or Interpro and structural databases such as SCOP and CDD.²² However, it is highly likely that the motifs found using these resources would be associated with the primary function of the protein such as “K⁺ selectivity filter” or “Voltage sensor domain”. The interaction information might not be present unless the domain is a generic binding domain like the PDZ-domain or has been specifically studied experimentally. We choose a brute-force approach to identify all possible relevant motifs in the protein sequence. In the simplest case, a sliding window of appropriate size is iteratively sampled from each protein sequence giving a total of $(L - \text{Winlen} + 1)$ motifs, where L is the length of the sequence and Winlen is the window size. Our choice of window size was based on two factors. First, for every residue involved in an interaction, we wish to maximize the likelihood

of finding at least one more interface residue in the neighboring sequence space in either direction. Ofran et al.⁸² reported that 70% of proteins had an interface residue within four residues of another interface residue. Applying this principle in both directions from an interface residue, we get a window size of nine. Second, we argue that evolutionary pressure that acts on a residue would involve at least a marginal rearrangement of the local structure. Therefore, an ideal window size would be the maximum length along the protein polymer chain that would be affected by a mutation at a certain position. The same window size can be inferred by reference to knowledge-based potentials developed for protein structure prediction.⁸³ These potentials describe probability density functions (*pdfs*) for structural fragments that have been successfully employed to assess and refine protein structures. The *pdfs* are described for five residues, $(n \pm 4)$; thus the net structural impact of a particular residue will be on neighboring 4 residues in each direction, giving a total of nine residues. Remarkably, both arguments agree on a window size of nine residues. Therefore, our length for the sliding window strategy is fixed to nine.

Since our arguments are based on protein sequences rather than alignments, we need to extract the windows from the sequence, which in this case is the authority protein sequence for both the $K_v1.2$ and the β_2 subunit from rat. For every window, the corresponding positions in the alignment are identified and the sub-alignment is extracted. Note here that the alignment contains gaps; therefore each sub-alignment might have a length greater than nine. Since we chose to ignore gapped motifs in the present study, this did not pose a problem for our analysis. Following this procedure, we got 491 motif alignments

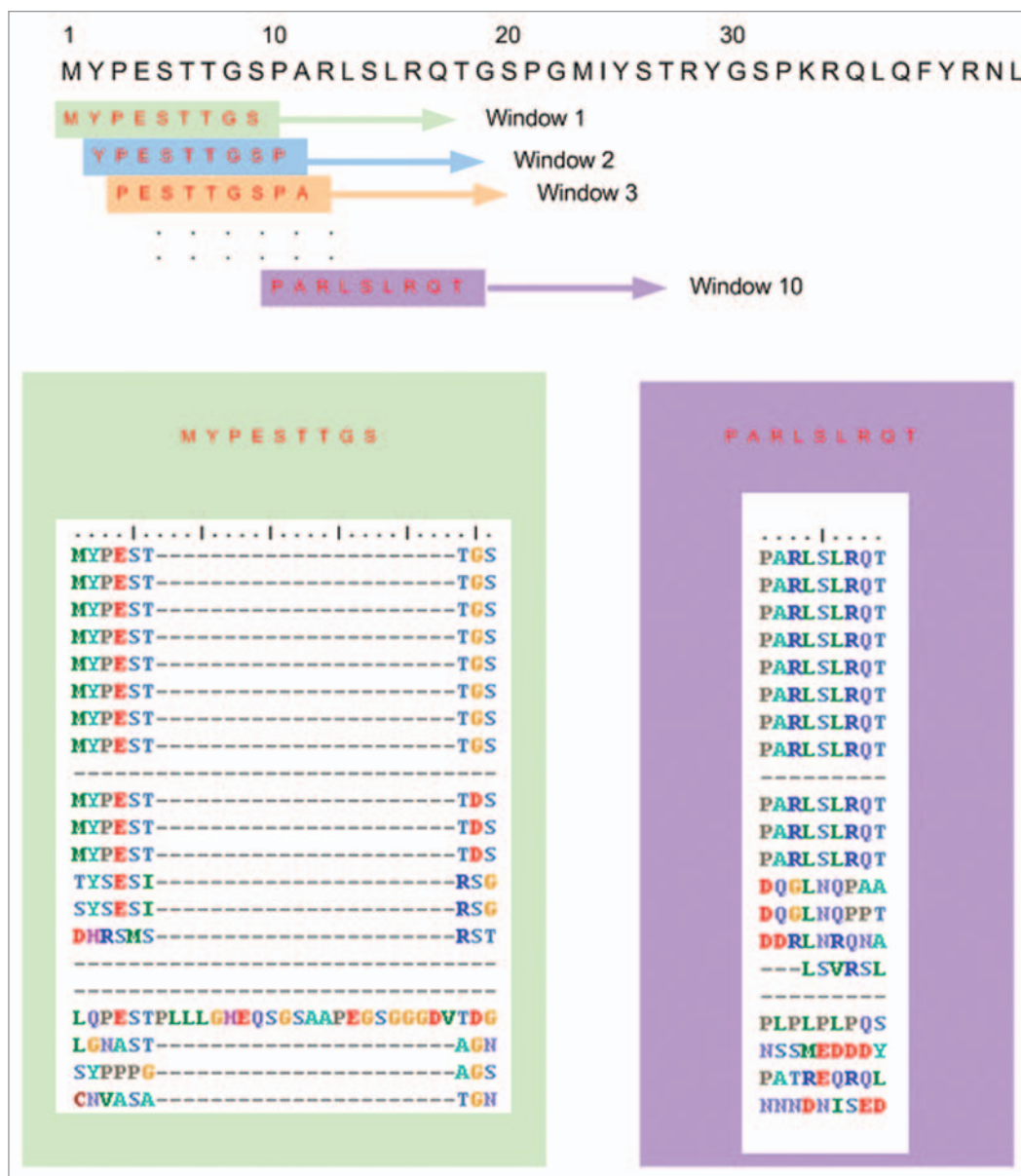


Figure 10. The Sliding Window strategy. Shown is an extract from the β_2 sequence and a few sample sliding windows. The sub-alignments corresponding to window 1 and window 10 are also shown. The numbers above the topmost sequence indicate residue position.

for K_v 1.2 and 359 for β_2 . The sliding window protocol is illustrated in Figure 10.

Distance/similarity assessment. The next step is defining a distance/similarity measure between sequences. For each sub-alignment extracted from the main alignments of K_v 1.2 and β_2 , we computed a distance matrix using CLUSTALW⁸⁴ using the default settings, that essentially computes distance as a fraction of mismatches. The actual identity of the amino acids in both sequences is not important for this distance calculation. The strategy is similar to Jothi et al.²⁶ wherein we obtain a $N \times N$ upper triangular matrix of distances/similarities between sequences; each row or column corresponding to a MoLFunC from a species (total species = N). The matrix is converted to a vector with $N \times (N - 1)/2$ entries that serves as the input for the correlation analysis. Since we had

21 species common to K_v 1.2 and β_2 , we obtained 491 vectors for K_v 1.2 and 359 for β_2 , each of length 210; i.e., $21 \times (21 - 1)/2$.

Correlation analysis. We then computed correlations between all the distance vectors (motifs) of K_v 1.2 with all the distance vectors (motifs) of β_2 , using Pearson's correlation coefficient, which gives the linear correlation between a pair of distance vectors.

$$R_{x,y} = \frac{n \sum_{i=1}^n x_i y_i \cdot \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\left(\sqrt{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \right) \left(\sqrt{n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2} \right)}$$

Almost all of the mirror-tree based techniques implemented previously used this measure to infer interactions; we adopted the same measure for the sake of consistency. The correlation analysis resulted in a 491 x 359 matrix of correlations. In the following section, we describe our method for significance testing of correlations. The correlation was computed using the “*corrcoef*” function in MATLAB.

Significance testing. The standard procedure for significance testing has been bootstrapping with random shuffling of both sequence vectors or the original sequences in the alignment.¹⁴ The shuffling is repeated 10^3 – 10^6 times and a z score or p-value is computed from the distribution, for the correlation matrix. Another technique of assessing significance for multiple motif pairs in two proteins has been to repeatedly sample random pairs and compute the correlations.²³ The p value of each pairing can be computed from the distribution. Yet another method that has been employed is the construction of artificial sequences and evolving them along a phylogenetic tree based on an evolutionary model.³² Certain global properties of each sequence such as amino acid composition are maintained in this process.

The random shuffling of sequences is especially useful in assessing if the original pairing was the right pairing. So if we have several ligand-receptor pairs as in Ramani et al.⁸⁵ the hypothesis is that each ligand is paired up with the appropriate receptor. In our case we have established the pairing by our previously cited methodology of determining MoLFunCs and postulating that the respective MoLFunCs of the rat $K_v1.2$ and rat β_2 will interact with each other in all other species. So the random shuffling technique is not directly applicable to our problem. Further, we believe that the random sampling of possible domain pairs might contribute to heavy re-sampling of the data set, as previously suggested in CAPS,³² which includes a correction for multiple testing. Nevertheless, it is advantageous to use a large sample of motif pairs beyond the pairs whose correlations are being investigated. Model-generated sequences are inherently dependent on the model. In our work, we have tried not to assume any specific evolutionary model and our analysis is dependent only on correlations in the raw data. The only dependence is on substitution matrices that have an implicit dependence on evolutionary models. Wherever possible, we prefer to evaluate our significance with real sequences, as they most certainly reflect the evolutionary patterns across species.

In order to estimate background distribution of possible correlations, we extracted MoLFunCs for as many proteins as possible from the authority genome, across all species under consideration. To do this rigorously by the methods employed for $K_v1.2$ and β_2 would be an enormous calculation; so we simplified the process. For the other proteins we only carried out the first stage of BLAST searches starting from the authority and accepted the resulting list of proteins as putative MoLFunCs. Thus there were no iterations performed starting from the authority seed and no refinement process was employed. Although this might not be the most accurate representation of MoLFunCs, we hope that the errors will only be marginal, as we will be sampling a large population of motifs from these proteins. We note that the idea of authority cannot be fully applied to all proteins, as several are

not annotated properly and we do not have structures for many of them. This does not compromise the identification of a set of sequences from the 21 species as a group of MoLFunCs, but we could not be sure what the function is. However, for the sake of consistency and to prepare for a possible scenario wherein we might actually take a closer look at this background data set for interesting correlation patterns, we defined an approximate notion of authority for these sequences. From the rat genome, we extracted all proteins that did not have the keywords *putative*, *possible* or *hypothetical* in their definition line. The rest of the sequences can be assumed to have some degree of functional definition and thus can be assumed to be authoritative. The total number of authoritative sequences obtained was 9192. Of these, only 566 had MoLFunCs in all of the 21 species obtained for $K_v1.2$ and β_2 . We refer to this set as the “*Background Set*”. The MoLFunC sequences were extracted and aligned using MUSCLE software.

There are two kinds of tests that we carry out—(1A) Assess the overall specificity of the $K_v1.2$ - β_2 interaction by assessing the significance of correlations in the context of all possible motif pair interactions in the proteome. (1B) Assess the specificity of a particular motif-pair in the context of the interactions that each of the motifs in the pair has with other motifs in the proteome. The second case is more specific in that it attempts to separate the less promiscuous motif from a given motif pair. This is especially relevant when we analyze the emergent properties of the correlation network.

For the *Case 1a*, the null hypothesis states that the correlation between motif pairs can be obtained by chance alone. If this is true, any two randomly selected motifs from any two randomly selected proteins are likely to be as highly correlated as the motif pair under investigation. In order to test this hypothesis, we randomly selected a protein and then randomly sampled a gapless 9-residue window from the *Background Set* of 566 proteins and the $K_v1.2$ and β_2 alignments. This was repeated 20,000 times. The first motif was paired with the second, the third with the fourth and so on. This gave us a motif-pairs sample of 10,000. We refer to this set as the “*Random Pairs Set*”. We then computed correlations between the 10,000 pairs using the same formula as in the $K_v1.2$ - β_2 motif correlations. A significance matrix of 491*359 entries is computed; each row and column index of this matrix can be mapped to the indices of the correlation matrix. The significance of a motif pair (corresponding to a cell in the correlation matrix) is simply the probability of obtaining a correlation equal to or greater than the correlation between the motifs. This is computed as the number of motif pairs in the *Random Pairs Set* with correlation equal to or higher than that of the motif pair under consideration, divided by the total size of the sample (10,000). After filtering the correlation matrix by ensuring that the corresponding entry in the significance matrix is lesser than or equal to a p-value threshold, there are 10,612 motif pairs in $K_v1.2$ and β_2 that appear to be significantly correlated. The thirty highest scoring matrix pairs are shown in **Table 1**. Further results and analysis will be provided in the results and discussion sections.

For *Case 1b*, we are investigating the correlations of “a given motif” with all other motifs from the interacting protein. The

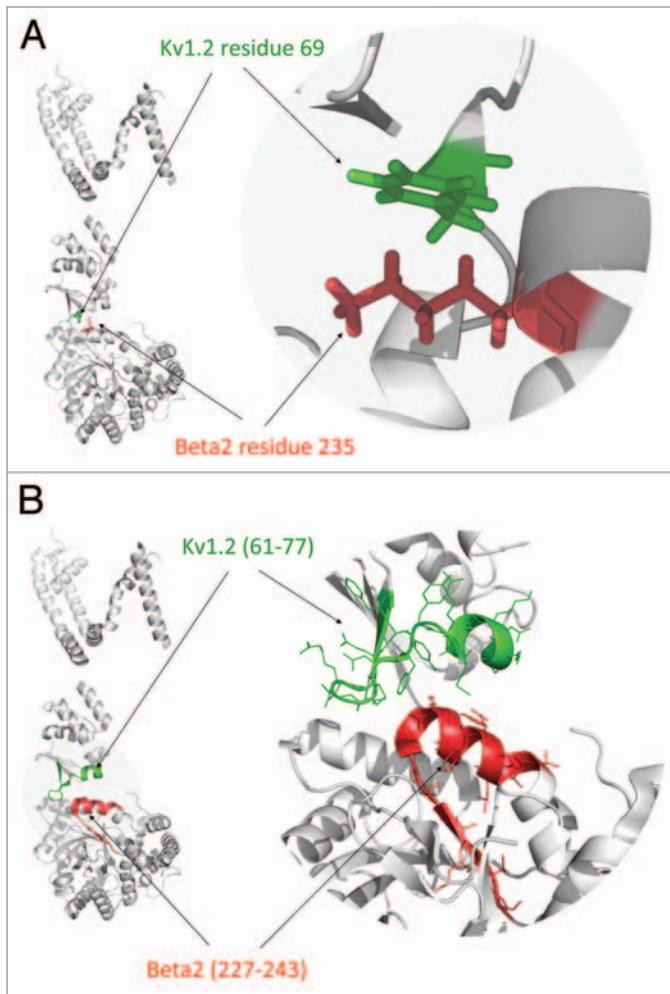


Figure 11. (A) The closest residue pair between the two subunits, residue 235 on the β_2 subunit and residue 69 on the $K_v1.2$ subunit. (B) The motifs possibly involved in interaction (as judged by structural proximity), region 227–243 on the β_2 subunit and region 61–77 on the $K_v1.2$ subunit.

null hypothesis again states that these correlations could have been obtained by chance. This can be rephrased as: What is the likelihood that the “given motif” would interact with a randomly selected motif from the genome? We carried out a similar procedure as before and extract random motifs from random proteins; and repeat the process 10,000 times. We will refer to this set as the “*Random Motif Set*”. A correlation vector was computed between the “given motif” and all motifs in the *Random Motif Set*. We call this the “*Motif Background vector*”. We then extracted the vector of correlation values of “the given motif” with all motifs from the interacting protein from the correlation matrix (in this case it is the $K_v1.2$ - β_2 motif correlation matrix). We call this the “*Motif Interaction vector*”. The significance of each cell in this vector was computed as the number of cells in the *Motif Background Vector* with correlation equal to or higher than the value of the cell in the *Motif Interaction vector*, divided by the total sample size (10,000). We repeated this procedure for all motifs in both sequences giving us a 491×359 significance matrix as before.

The correlation analysis, the significance testing and plots were carried out in MATLAB. The *Background Set* calculation and the distance computations and the random motif list extractions were all carried out using Perl Scripts. The visual analysis of structural details was carried out using Pymol software.⁸⁶

Recall/precision analysis. Our methods have been inspired from the field of Information Retrieval. The standard measures for measuring success of a retrieval engine are *Precision* and *Recall*. *Recall* measures the proportion of relevant documents (to a query) that was retrieved by the engine in response to the query. *Precision* measures the proportion of retrieved documents that is relevant. High precision implies very few false positives; high recall implies very few false negatives. The goal is to achieve high precision and high recall, but that is seldom the case. A useful graphical aid to analyze the performance of the retrieval engine/algorithm is to plot precision at varying recall levels. The standard recall levels that are used range from 0% to 100%, with an increment of 10%. Thus, there are 11 standard recall levels. However, in some cases, our results may not have recall values corresponding to the standard recall levels, but rather may have values in between them. In such cases, the precision at the standard level is interpolated from the next highest recall level beyond the standard. Thus, if there are recall levels of 33, 66 and 100%, the precision at recall levels 0, 10, 20 and 30% is equal to the precision at recall level of 33%. Maintaining a high precision at all levels of recall is an indicator of the power of the algorithm in retrieving the most relevant documents among the top few in the output list.

We analyzed the results for both goals of this paper using these measures and graphs. For *Case 1a*, the query corresponds to the protein pair; the output is a putative list of interacting motif pairs. Since there are several combinations of motif pairs between the two proteins, we used a cut-off for the significance of motif-pair correlations. The resulting motif pairs were ranked in the order of decreasing correlation values, analogous to the documents relevant to a query being ranked in the decreasing order of relevance. For *Case 1b*, the query corresponds to a motif from either protein and the results are the most likely interacting motifs from the other protein. Again, we only chose pairs within a significance cut-off and ranked the rest in order of decreasing correlation values. We used several motifs as queries and calculated the precision for all the queries at each recall level. The p value cutoff chosen was 0.05.

The most obvious criterion for relevance is structural proximity; residues and motifs immediately adjacent to each other must either be conserved or undergo correlated evolution to maintain the interaction. We defined distances between motifs in either protein as the minimum distance between the residues of the motif pair in the full octameric (four α -subunits plus four beta-subunits) structure of the protein complex. The distances at the residue level were computed as the average distance of all inter-atomic distances between the residues. Once we had the closest residue pair between the two proteins, we included all the motifs that contain the respective residue from each protein. This should correspond to 9 motifs (due to the choice of window size) from each protein, giving a total of 81

possible interacting pairs as the list of relevant interactions to be predicted. The residue pair that was found to be the closest was residue 69 from K_v1.2 and residue 235 from the β₂ subunit, at a distance of 6.01 Angstroms. **Figure 11A** shows this residue pair in the 3D structure. The list of motifs containing these two residues from either protein is shown in **Table 2** and **Figure 11B** highlights these motifs in the 3D structure. Discovering any one of them among the very top hits would be a promising

indicator of the effectiveness of our method. These motifs form the “*Structurally Relevant Set*”.

Acknowledgements

We would like to thank Dr. Ionel Rata and the entire Jakobsson lab for stimulating discussions on this work. We thank Dr. Cibele Falkenberg for doing new calculations using the pump-leak cell model to infer consequences of K⁺ channel inhibition.

References

- Deutsch C. Potassium channel ontogeny. *Ann Rev Physiol* 2002; 64:19-46.
- Long SB, Campbell EB, MacKinnon R. Crystal structure of a mammalian voltage-dependent Shaker family K⁺ channel. *Science* 2005; 309:897-903.
- Li Y, Um SY, McDonald TV. Voltage-gated potassium channels: Regulation by accessory subunits. *Neuroscientist* 2006; 12:199-210.
- Torres YP, Morera FJ, Carvacho I, Latorre R. A marriage of convenience: beta-subunits and voltage-dependent K⁺ channels. *J Biol Chem* 2007; 282:24485-9.
- Campomanes CR, Carroll KI, Manganas LN, Hershberger ME, Gong B, Antonucci DE, et al. K_vbeta Subunit Oxidoreductase Activity and K_v1 Potassium Channel Trafficking. *J Biol Chem* 2002; 277:8298-305.
- Bhatnagar A, Kumar R, Tipparaju SM, Liu S-Q. Differential pyridine nucleotide coenzyme binding to the β-subunit of the voltage-sensitive K⁺ channel: a mechanism for redox regulation of excitability? *Chemico-Biological Interactions* 2003; 143:613-20.
- Tipparaju SM, Saxena N, Liu S-Q, Kumar R, Bhatnagar A. Differential regulation of voltage-gated K⁺ channels by oxidized and reduced pyridine nucleotide coenzymes. *Am J Physiol Cell Physiol* 2005; 288:366-76.
- Hoshi T, Zagotta W, Aldrich R. Biophysical and molecular mechanisms of Shaker potassium channel inactivation. *Science* 1990; 250:533-8.
- Conforti L, Bodi I, Nisbet JW, Millhorn DE. O₂-sensitive K⁺ channels: role of the K_v1.2 α-subunit in mediating the hypoxic response. *J Physiol* 2000; 524:783-93.
- Coppock EA, Tamkun MM. Differential expression of KV channel α—and beta—subunits in the bovine pulmonary arterial circulation. *Am J Physiol Lung Cell Mol Physiol* 2001; 281:1350-60.
- Weir EK, Olschewski A. Role of ion channels in acute and chronic responses of the pulmonary vasculature to hypoxia. *Cardiovasc Res* 2006; 71:630-41.
- Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. *Proc Natl Acad Sci USA* 1999; 96:4285-8.
- Jothi R, Przytycka TM. Computational approaches to predict protein-protein and domain-domain interactions. In: Zelikovsky MA, Ed. *Bioinformatics Algorithms: Techniques and Applications*. New York: Wiley Press 2008.
- Goh CS, Bogan AA, Joachimiak M, Walthers D, Cohen FE. Co-evolution of proteins with their interaction partners. *J Mol Biol* 2000; 299:283-93.
- Pazos F, Valencia A. Similarity of phylogenetic trees as indicator of protein-protein interaction. *Protein Engineering* 2001; 14:609-14.
- Kim WK, Bolser DM, Park JH. Large-scale co-evolution analysis of protein structural interlogues using the global protein structural interactome map (PSIMAP). *Bioinformatics* 2004; 20:1138-50.
- Pazos F, Ranea JAG, Juan D, Sternberg MJE. Assessing protein co-evolution in the context of the tree of life assists in the prediction of the interactome. *J Mol Biol* 2005; 352:1002-15.
- Sato T, Yamanishi Y, Kanehisa M, Toh H. The inference of protein-protein interactions by co-evolutionary analysis is improved by excluding the information about the phylogenetic relationships. *Bioinformatics* 2005; 21:3482-9.
- Juan D, Pazos F, Valencia A. High-confidence prediction of global interactomes based on genome-wide coevolutionary networks. *Proc Natl Acad Sci USA* 2008; 105:934-9.
- Aloy P, Russell RB. Structural systems biology: modeling protein interactions 2006; 7:188-97.
- Kim WK, Park J, Suh JK. Large scale statistical prediction of protein-protein interaction by potentially interacting domain pair. *Genome Inform* 2002; 13:42-50.
- Galperin MY. The Molecular Biology Database Collection: 2008 update. *Nucl Acids Res* 2008; 36:2-4.
- Nye TMW, Berzuini C, Gilks WR, Babu MM, Teichmann SA. Statistical analysis of domains in interacting protein pairs. *Bioinformatics* 2005; 21:993-1001.
- Kim Y, Subramaniam S. Locally defined protein phylogenetic profiles reveal previously missed protein interactions and functional relationships. *Proteins* 2006; 62:1115-24.
- Kim Y, Koyuturk M, Topkara U, Grama A, Subramaniam S. Inferring functional information from domain co-evolution. *Bioinformatics* 2006; 22:40-9.
- Jothi R, Cherukuri PF, Tasneem A, Przytycka TM. Co-evolutionary analysis of domains in interacting proteins reveals insights into domain-domain interactions mediating protein-protein interactions. *J Mol Biol* 2006; 362:861-75.
- Finn RD, Marshall M, Bateman A. iPfam: visualization of protein-protein interactions in PDB at domain and amino acid resolutions. *Bioinformatics* 2005; 21:410-2.
- Kann MG, Jothi R, Cherukuri PF, Przytycka TM. Predicting protein domain interactions from coevolution of conserved regions. *Proteins* 2007; 67:811-20.
- Anthony AF, Richard WA. Influence of conservation on calculations of amino acid covariance in multiple sequence alignments. *Proteins* 2004; 56:211-21.
- Mintseris J, Weng Z. Structure, function and evolution of transient and obligate protein-protein interactions. *Proc Natl Acad Sci USA* 2005; 102:10930-5.
- Horner DS, Pirovano W, Pesole G. Correlated substitution analysis and the prediction of amino acid structural contacts. *Brief Bioinform* 2008; 9:46-56.
- Fares MA, Travers SAA. A novel method for detecting intramolecular coevolution: Adding a further dimension to selective constraints analyses. *Genetics* 2006; 173:9-23.
- Yeang C-H, Haussler D. Detecting coevolution in and among protein domains. *PLoS Comput Biol* 2007; 3:211.
- Halperin I, Wolfson H, Nussinov R. Correlated mutations: Advances and limitations. A study on fusion proteins and on the Cohesin-Dockerin families. *Proteins* 2006; 63:832-45.
- Natarajan S, Jakobsson E. Functional equivalency inferred from authoritative sources in networks of homologous proteins. *PLoS ONE* 2009; 4.
- Sokolova O, Accardi A, Gutierrez D, Lau A, Rigney M, Grigorieff N. Conformational changes in the C terminus of Shaker K⁺ channel bound to the rat K_vβ₂-subunit. *Proc Natl Acad Sci USA* 2003; 100:12607-12.
- Yarov-Yarovoy V, Baker D, Catterall WA. Voltage sensor conformations in the open and closed states in rosetta structural models of K⁺ channels 2006; 103:7292-7.
- Trimmer JS. Regulation of ion channel expression by cytoplasmic subunits. *Curr Opin Neurobiol* 1998; 8:370-4.
- Accili EA, Kiehn J, Yang Q, Wang Z, Brown AM, Wible BA. Separable K_vβ subunit domains alter expression and gating of potassium channels. *J Biol Chem* 1997; 272:25824-31.
- Minor DL, Lin Y-F, Mobley BC, Avelar A, Jan YN, Jan LY, et al. The polar T1 interface is linked to conformational changes that open the voltage-gated potassium channel. *Cell* 2000; 102:657-70.
- Sather WA, McCleskey EW. Permeation and selectivity in calcium channels. *Ann Rev Physiol* 2003; 65:133-59.
- Post JM, Gelband CH, Hume JR. [Ca²⁺] Inhibition of K⁺ channels in canine pulmonary artery: Novel mechanism for hypoxia-induced membrane depolarization. *Circ Res* 1995; 77:131-9.
- Tao X, Lee A, Limapichat W, Dougherty DA, MacKinnon R. A gating charge transfer center in voltage sensors. *Science* 2002; 298:67-73.
- Xu J, Yu W, Wright JM, Raab RW, Li M. Distinct functional stoichiometry of potassium channel beta subunits. *Proc Natl Acad Sci USA* 1998; 95:1846-51.
- Kannan N, Vishveshwara S. Aromatic clusters: a determinant of thermal stability of thermophilic proteins. *Protein Eng* 2000; 13:753-61.
- McGaughey GB, Gagne M, Rappe AK. pi-Stacking interactions. Alive and well in proteins. *J Biol Chem* 1998; 273:15458-63.
- Shen W, Hernandez-Lopez S, Tkatch T, Held JE, Surmeier DJ. K_v1.2-containing K⁺ channels regulate subthreshold excitability of striatal medium spiny neurons. *J Neurophysiol* 2004; 91:1337-49.
- Cordero-Morales JF, Jogini V, Lewis A, Vasquez V, Cortes DM, Roux B, et al. Molecular driving forces determining potassium channel slow inactivation. *Nat Struct Mol Biol* 2007; 14:1062-9.
- Haliloglu T, Ben-Tal N. Cooperative transition between open and closed conformations in potassium channels. *PLoS Comput Biol* 2008; 4:1000164.
- Salghetti SE, Muratani M, Wijnen H, Futcher B, Tansey WP. Functional overlap of sequences that activate transcription and signal ubiquitin-mediated proteolysis. *Proc Natl Acad Sci USA* 2000; 97:3118-23.
- Bohren KM, Brownlee JM, Milne AC, Gabbay KH, Harrison DHT. The structure of Apo R268A human aldose reductase: Hinges and latches that control the kinetic mechanism. *Biochimica et Biophysica Acta (BBA)—Proteins & Proteomics* 2005; 1748:201-12.
- Krebs WG, Gerstein M. Survey and summary: The morph server: a standardized system for analyzing and visualizing macromolecular motions in a database framework. *Nucl Acids Res* 2000; 28:1665-75.
- Gulbis JM, Mann S, MacKinnon R. Structure of a voltage-dependent K⁺ channel β subunit. *Cell* 1999; 97:943-52.

54. Liu SQ, Jin H, Zacarias A, Srivastava S, Bhatnagar A. Binding of pyridine nucleotide coenzymes to the β -subunit of the voltage-sensitive K⁺ channel. *J Biol Chem* 2001; 276:11812-20.
55. Kleinberg JM. Authoritative sources in a hyperlinked environment. *J ACM* 1999; 46:604-32.
56. Xu CQ, Zhu SY, Chi CW, Tytgat J. Turret and pore block of K⁺ channels: what is the difference? *Trends Pharmacol Sci* 2003; 24:446-8.
57. Keskin OMB, Nussinov R. Hot regions in protein—protein interactions: the organization and contribution of structurally conserved hot spot residues. *J Mol Biol* 2005; 345:1281-94.
58. Bloom JD, Labthavikul ST, Otey CR, Arnold FH. Protein stability promotes evolvability 2006; 103:5869-74.
59. Schechtman D, Mochly-Rosen D. Adaptor proteins in protein kinase C-mediated signal transduction. *Oncogene* 2001; 20:6339-47.
60. Short M, Fox S, Stenmark KR, Das M. Hypoxia-induced alterations in protein kinase C[ζ] signaling result in augmented fibroblast proliferation. *Chest* 2005; 128:582.
61. Liu YV, Hubbi ME, Pan F, McDonald KR, Mansharamani M, Cole RN, et al. Calcineurin promotes hypoxia-inducible factor 1 α expression by dephosphorylating RACK1 and blocking RACK1 dimerization. *J Biol Chem* 2007; 282:37064-73.
62. Jivotovskaya AV, Valasek L, Hinnebusch AG, Nielsen KH. Eukaryotic translation initiation factor 3 (eIF3) and eIF2 can promote mRNA binding to 40S subunits independently of eIF4G in Yeast. *Mol Cell Biol* 2006; 26:1355-72.
63. Koritzinsky M, Magagnin MG, van den Beucken T, Seigneuric R, Savelkoul K, Dostie J, et al. Gene expression during acute and prolonged hypoxia is regulated by distinct mechanisms of translational control. *EMBO J* 2006; 25:1114-25.
64. Liu L, Cash TP, Jones RG, Keith B, Thompson CB, Simon MC. Hypoxia-induced energy stress regulates mRNA translation and cell growth 2006; 21:521.
65. Hinnebusch AG. eIF3: a versatile scaffold for translation initiation complexes. *Trends Biochem Sci* 2006; 31:553-62.
66. Jeffrey RM, Yong-Geun Kwak, Tamkun MM. Modulation of K_v channel α/β subunit interactions. *Trends Cardiovasc Med* 1999; 9:253-8.
67. Wang X, Zhang J, Berkowski SM, Knowleg H, Chandramouly AB, Downens M, Prystowsky MB. Protein kinase C-mediated phosphorylation of K_v β ₂ in adult rat brain. *Neurochemical Research* 2004; 29:1879-86.
68. Liu SQ, Yin J, Bhatnagar A. Protein kinase C-dependent phosphorylation of the [beta]-subunit of the voltage-sensitive potassium channels (K_v β ₂). *Chemico-Biological Interactions* 2003; 143:597-604.
69. Hattan D, Nesti E, Cachero TG, Morielli AD. Tyrosine phosphorylation of K_v1.2 modulates its interaction with the actin-binding protein cortactin. *J Biol Chem* 2002; 277:38596-606.
70. Strauss O, Rosenthal R, Dey D, Beninde J, Wollmann G, Thieme H, et al. Effects of protein kinase C on delayed rectifier K⁺ channel regulation by tyrosine kinase in rat retinal pigment epithelial cells. *Invest Ophthalmol Vis Sci* 2002; 43:1645-54.
71. Felsch JS, Cachero TG, Peralta EG. Activation of protein tyrosine kinase PYK2 by the m1 muscarinic acetylcholine receptor. *Proc Natl Acad Sci USA* 1998; 95:5051-6.
72. Nesti E, Everill B, Morielli AD. Endocytosis as a mechanism for tyrosine kinase-dependent suppression of a voltage-gated potassium channel. *Mol Biol Cell* 2004; 15:4073-88.
73. Betty Y, Chang, Rachel A, Harte, Cartwright CA. RACK1: a novel substrate for the Src protein-tyrosine kinase. *Oncogene* 2002; 21:7619-29.
74. Kupper JBM, Marom S, Levitan IB. Intracellular and extracellular amino acids that influence C-type inactivation and its modulation in a voltage-dependent potassium channel. *Pflugers Arch* 1995; 430:1-11.
75. Tang XD, Santarelli LCH, Heinemann S, Hoshi T. Metabolic regulation of potassium channels. *Ann Rev Physiol* 2004; 66:131-59.
76. Tipparaju SM, Barski OA, Srivastava S, Bhatnagar A. Catalytic mechanism and substrate specificity of the β -subunit of the voltage-gated potassium channel. *Biochemistry* 2008; 47:8840-54.
77. Oleg A, Barski, Srinivas M, Tipparaju, Bhatnagar A. Kinetics of nucleotide binding to the β -subunit (AKR6A2) of the voltage-gated potassium (K_v) channel. *Chemico-Biological Interactions* 2009; 178.
78. Jakobsson E. Interactions of cell volume, membrane potential and membrane transport parameters. *Am J Physiol* 1980; 238:196-206.
79. Xue J, Zhou D, Yao H, Haddad GG. Role of transporters and ion channels in neuronal injury under hypoxia. *Am J Physiol Regul Integr Comp Physiol* 2008; 294:451-7.
80. Socolich M, Lockless SW, Russ WP, Lee H, Gardner KH, Ranganathan R. Evolutionary information for specifying a protein fold 2005; 437:512-8.
81. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl Acids Res* 2004; 32:1792-7.
82. Ofran Y, Rost B. Predicted protein-protein interaction sites from local sequence information. *FEBS Letters* 2003; 544:236-9.
83. Atipat Rojnuckarin SS. Knowledge-based interaction potentials for proteins. *Proteins: Structure, Function and Genetics* 1999; 36:54-67.
84. Thompson JD DGH, 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; 22:4673-80.
85. Ramani AK, Marcotte EM. Exploiting the co-evolution of interacting proteins to discover interaction specificity. *J Mol Biol* 2003; 327:273-84.
86. DeLano WL. The PyMOL Molecular Graphics System. Palo Alto: DeLano Scientific 2002.