Domain Insertion Permissibility is a Measure of Engineerable Allostery in Ion Channels

Willow Coyote-Maestas[1], Yungui He[2], Chad L. Myers[3] & Daniel Schmidt[2*]

[1]Dept. of Biochemistry, Molecular Biology & Biophysics, [2]Dept. of Genetics, Cell Biology & Development, [3]Dept. of Computer Science and Engineering, University of Minnesota, Minneapolis, MN

*To whom correspondence should be addressed: schmida@umn.edu

Abstract

Allostery is a fundamental principle of protein regulation that remains poorly understood and hard to engineer, in particular in ion channels. Here we use human Inward Rectifier K^+ Channel Kir2.1 to establish domain insertion 'permissibility' as a new experimental paradigm to identify engineerable allosteric sites. We find that permissibility is best explained by dynamic protein properties, such as conformational flexibility. Many allosterically regulated sites in Kir2.1 or sites equivalent to those regulated in homologs such as G-protein-gated inward rectifier K^+ channels (GIRK) have *differential* permissibility; that is permissibility depends on the structural properties of the inserted domain. Our data and the well-established link between protein dynamics and allostery led us to propose that differential permissibility is a metric of both existing and latent allostery in Kir2.1. In support of this notion, inserting light-switchable domains into either existing or latent allosteric sites, but not elsewhere, renders Kir2.1 activity sensitive to light.

The propagation of information between distal sites in proteins is called allosteric regulation. By providing a means for long-range communication, allostery enables the evolution of natural multi-domain proteins (e.g., (*1*)) and the engineering of synthetic proteins (e.g., (*2*)) through end-to-end joining or the insertion of functionally and structurally discrete protein domains. Despite its ubiquity in natural protein systems and utility in synthetic proteins, the emergence of allosteric regulation is poorly understood.

For allosteric regulation to evolve, the allosteric modulator and the modulated protein must *simultaneously possess* compatible allosteric signaling pathways, that could impart a selectable advantage (*3*, *4*). The most parsimonious way this can happen is if the required signal transduction networks *already exist* within the modulated protein (**Fig. 1A**). The term latent allostery has been coined for such hidden –latent– phenotypes that are not under selection, but that could be exploited by the same amino acid sequence if selection pressures change (*5*). Co-opting latent phenotypes allows for adaptation toward new functions that are complementary with an existing dominant function. As one example, a scaffold protein (Ste5) allosterically regulates Erk-like kinases that diverged *before* the evolution of Ste5 itself, implying that the allosteric potential to be regulated by Ste5 *was already present* at that point (*4*). Another example is an ancestral steroid receptor that responds to aldosterone, even though this hormone became available *later* in mammalian evolution (*6*). While latent allostery is a useful conceptual framework with implications for protein evolution and the engineering of synthetic proteins, our current understanding of latent allostery is limited. Methods that have guided the study and engineering of allostery in soluble proteins, such as random circular permutation (*7*), NMR (*8*), and computational approaches (*9*, *10*), rely heavily on structural information or statistical methods. Thus they remain relatively inaccessible to membrane proteins, such as ion channels, which provide essential properties to neuromuscular tissues and are drug targets for many neurologic and cardiac disorders.

Here we establish an experimental method for *de novo* mapping allosteric sites without relying on structural information or statistical measures. For this proof-of-concept, we used the inward rectifier K^+ channel Kir2.1, because (1) extensive functional studies are available (*11*), (2) known somatic mutations are linked to disruption of allosteric signaling and result in disease (e.g., Long QT syndrome (*12*)), (3) high-resolution crystal structures are available across phylogeny (*13*-*15*), and (4) there is great allosteric regulatory diversity between homologs (e.g., G-proteins and ATP *(11)*).

Because allosteric regulation requires both a functional protein and a means to distinguish different signals, we reasoned that for a site in Kir2.1 to have latent allosteric potential, it must have context-dependent plasticity. To test this idea, we mimicked protein evolution and inserted protein domains (PDZ, Cib81, $GSAG_{2x}$, $GSAG_{3x}$) with different properties into every amino acid position of Kir2.1 to measure site-specific domain insertion 'permissibility' (**Fig. 1A**). We define permissibility as the ability of Kir2.1 to accept a domain insertion that does not disrupt folding, assembly, and trafficking to the cell surface. We used the 10kDa syntrophin PDZ domain (PDB 2PDZ) because it is well structured and has been used to study how large domains with known function disrupt activity (*16*, *17*). Cryptochrome-interacting basic-helix-loop-helix (Cib81) is a

similarly sized domain that forms a two-component switchable system with its binding partner, CRY2, after blue-light illumination (*18*). We included flexible $GSAG_{x2}$ and $GSAG_{x3}$ linkers to establish a permissibility baseline. To generate insertion libraries, we used a saturation domain insertion profiling method, domain insertion profiling with sequencing (DIPSeq) (*17*). DIPSeq uses a MuA transposase to insert an antibiotic cassette into random positions of a gene (i.e., all six reading frames) (**Supp. Fig. 1**). Upon antibiotic selection of variants with insertions, we replaced this cassette with a domain of interest (Cib81, PDZ, $GSAG_{x2}$, and $GSAG_{x3}$) using restriction sites at transposon ends that double as flexible linkers. To determine which amino acid positions are permissive and non-permissive to domain insertions, we leveraged the fact that for a channel to be expressed on the cell surface it must fold, oligomerize, and surface traffic (*19*). Therefore, only permissive insertion variants can be fluorescently labeled via a FLAG epitope tag we inserted into an extracellular loop of Kir2.1 (S116) that does not interfere with channel function (*20*). In this way, two cell populations can be collected by fluorescence-activated cell sorting (FACS): those that express channel variants but don't surface express ($EGFP^{high}$/anti-$FLAG^{low}$) and those that do surface express ($EGFP^{high}$/ anti-$FLAG^{high}$). From both populations, we isolated and sequenced plasmid DNA, and aligned reads with the DIPSeq alignment pipeline. We calculate permissibility as site-specific enrichment between 'not surface-expressed' to 'surface-expressed' insertion variants. Apart from some regions near the N-terminus, most notably M1, coverage in the remaining regions is near complete for all three insertion datasets (**Fig. 1C**).

We then mapped permissibility onto the crystal structure of chicken Kir2.2 (PDB 3SPC) (*21*). As expected, domain insertion positions that should not allow surface expression do not (e.g., transmembrane and inter-subunit interfaces, **Fig. 1B, Detail A**). Also unsurprising, the unstructured C-terminus (which *in vivo* interacts with scaffolding proteins not present in HEK293 cells) was highly permissive to any insertion. Predictably, overall flexible peptide insertions are more permissive than larger, more structured domains. Counter-intuitively, many non-conserved, surface-exposed loops (e.g., αG-βH or βN-αH) were also not permissive (**Fig. 1B, Detail B**). From this, we conclude that the rules that govern permissibility, at least in Kir2.1, differ from cytosolic proteins. Perhaps this is due to more stringent selection pressures unique to membrane proteins such as the need for proper folding, assembly, surface trafficking, and membrane insertion.

Our data also reveals that permissibility is sensitive to the structural context of the inserted domain. Despite Cib81 and PDZ being of similar size, many sites are *differentially* permissive, suggesting that there is context dependence for permissibility beyond simple sterics (**Fig. 1B, Detail B**). Surprisingly, all of Kir2.1's known allosteric sites and many regulated sites are differentially permissive. This included the PIP_2 binding site at the interface between the pore and cytosolic domain, and the βG loop, a flexible region involved in channel opening (**Fig. 1B**). Particularly interesting is that the remaining highly yet differentially permissive sites –outside unstructured termini– are known to be allosterically regulated in homologs of Kir2.1. This includes the βB-βC loop (which binds ATP in Kir.6x (*22*)), and the βD-βE loop (which binds Gβγ in GIRK (*23*)) (**Fig. 1B**). Correlation matrices of permissibility profiles sorted by insertion sites reveal distinct clusters that coincide with structural features involved in allosteric regulation of Kir2.1 (**Supp. Fig. 2**).

We more quantitatively compared permissibility profiles of biological replicates of Kir2.1 for inserted domains (structured vs. flexible) by clustering correlation matrices (**Supp. Fig. 3**). Biological replicates were in very good agreement indicating relatively little noise was introduced through the transposition, heterologous expression, and cell sorting steps. We found that domains cluster by structure (Cib81 and PDZ cluster discretely from each other, but flexible insertions do not) and size (Cib81 and PDZ cluster closer than flexible peptides).

Given the unexpected patterns of domain insertion permissibility, we explored what protein features most contribute to permissibility. We calculated and compared structure-, conservation, and dynamic-based properties for Kir2.1 with different domain permissibility profiles (**Fig. 2A**). We found that domain insertion permissibility is correlated with dynamic features and does not correlate well with static and conservation-related protein properties. This result is entirely in agreement with the proposed mechanism through which dynamic protein properties enable latent allosteric properties to be exploited in response to novel regulatory factors (3).

To determine the degree to which computed properties can explain permissibility, and which properties best approximate permissibility, we constructed decision tree classification models. Consistent with the result that permissibility best correlates with dynamics properties, dynamic features had the greatest predictive power across the three types of domains (**Fig. 2B, Supp. Fig. 4-6**). While some profiles' predictive models perform better than others ($GSAG_{x2}$ was best and PDZ was worse) and none fully explain permissibility, we *can* build models for all domains, whose performance is far better than random as assessed by receiver operating characteristic (ROC) curves and other performance criteria. Our ability to generate predictive models for permissibility demonstrates that features used in predicting measured permissibility are meaningful and that the 'indefinable qualities' of permissibility play minor roles. Furthermore, the necessity for a non-linear approach (decision trees) demonstrates that permissibility, and by extension latent allostery, is an emergent phenotype from non-linear interactions between multiple protein properties, as opposed to a linear combination of, for example, conservation and surface exposure.

Since permissibility only reports on whether a Kir2.1 insertion variant can fold and traffic to the cell surface, we focused on a representative sample of insertion positions to assess whether they remain functional (i.e., able to conduct K^+) upon domain insertion. We subjected this subset to a flow cytometry-based optical activity assay that measures population-level resting membrane potential (RMP) in HEK293FT cells using an oxonol voltage-sensor, DIBAC4(3) (**Fig. 3A**) (*24*). Since Kir2.1 drives the RMP towards the reversal potential of K^+, cells expressing functional Kir2.1 are more hyperpolarized compared to 'empty' cells under our culture conditions (**Fig. 3B**). As expected, insertions into flexible and highly permissive regions (e.g., M24, A371) of Kir2.1 have little impact on function. All insertions into regions critical for gating, the G-loop or the conduction pore exit, break channel function. Whereas, insertions into the exposed extracellular loop (S116) have different impacts on function. Here, both Cib81 (presumably unstructured) and a flexible linker are well tolerated, while PDZ impairs function significantly. The emerging theme of differential impact on function continues in other regions of Kir2.1, including the PIP_2 binding site. Here, P187 is permissive to all insertions but partial function remains with large domain insertions

(PDZ & Cib81) while a flexible linker completely breaks channel function. Conversely, permissibility and impact on function tracked quite well for insertions into the D-E loop. Here, both PDZ and flexible insertions allowed gating, while Cib81 broke channel function. Overall as with permissibility, the functional assay shows that domain insertions in allosteric sites impact function in a context dependent manner that can't be explained with simple sterics.

What permissibility and functional assays tell us is that many allosterically regulated sites are sensitive to the structure of the inserted domain (differentially permissive) and retain at least partial function. The fact that sites involved in allosteric signaling in Kir2.1 homologs (GIRK and Kir6.2) have similar features, suggests that these sites might have latent allosteric potential (**Supp. Fig. 7**). To test this idea, we tested insertion Cib81 variants in these sites for light-dependent modulation, as well as several others that we predicted are non-allosteric (e.g, C-termini), a sterically inaccessible extracellular loop, and negative controls (wildtype and poor gating mutant, V302M). Initially, there was no optimization of flanking linkers. We reasoned that if Cib81 is sterically accessible and a position has allosteric potential, then a light-mediated association of the channel with co-expressed Cry2 (size 70 kDa) would modulate channel gating even if binding interfaces are not optimized. We adopted the flow cytometry RMP assay to measure Kir2.1 activity with and without blue light illumination. As expected, Wildtype channel and a gating mutant has no light-dependent modulation. Furthermore, when Cib81 is inserted into predicted non-allosteric sites terminal regions (e.g., E392) recruitment of Cry2 to the channel does not affect Kir2.1 activity. Similarly, when Cib81 is inserted into an inaccessible extracellular site (S116) there was no light-dependent effect on Kir2.1 activity. Remarkably, even though channel function was severely impaired, when Cib81 was inserted in the PIP_2 binding site, illumination markedly decreased the remaining Kir2.1 activity (**Fig. 4A, B**). We validated this with patch clamp electrophysiology, which shows that the open probability of Kir2.1(P187CIB), which is low to begin with, is further *decreased* with blue light illumination (**Fig. 4C, E**). We also observed Kir2.1 light-modulation when Cib81 was inserted into the pore-facing side of the αH Helix (A362), but not the outward-facing side (L361) (**Fig. 4B, E**). Furthermore, patch clamp validation of this insertion showed that open probability is higher than wild-type channel in the absence of illumination (**Fig. 4C**) and further *increased* with illumination (**Fig. 4D**). Most interestingly, we noticed weak light modulation when Cib81 was inserted into predicted latent allosteric sites which are part of the D-E and L-M loops (S238, L332, R335) (**Fig. 4B**). The weak impact of Cry2 recruitment in these regions could be due to none-optimized binding interfaces in contrast to $G\beta\gamma$ to GIRK's D-E and L-M loops (*23*). When we patched cell expressing one of these insertion mutants (S238), we observed higher Kir2.1 activity again even in the absence of illumination (**Fig. 4C**), suggesting that insertions into the DE loop are activating. We explored linker optimization to obtain a cleaner photoswitching phenotype. Flanking the inserted Cib81 by five amino acids, but not three or nine, improved light modulation of Kir2.1 significantly (**Fig. 4B**).

In summary, from the combination of permissibility and function assays, we coarsely classify Kir2.1 regions as (**Fig. 4E**): (1) 'Unimportant sites' which have *high* permissibility to any insertion and where impact on function is minimal, irrespective of insertion type e.g., C-terminal extension. (2) Scaffold sites, which have *low* permissibility to any insertion, and where any type of insertion severely impacts function e.g., transmembrane helices. (3) Allosteric sites, which have *differential*

permissibility that depends on both, the structural context of the insertion position as well the properties of the inserted domain. In these sites, we find that impact on function is not correlated with permissibility. We note that these sites are often involved in either allostery in Kir2.1 (e.g., PIP_2 binding site) or exploited in homologs, e.g., D-E loop (binding to $G\beta\gamma$ in GIRK).

That we can use computed dynamic features to predict permissibility, and the well-established connection between allostery and latent phenotypes emerging from dynamic protein features, allows us to link existing ideas about latent allostery to an experimentally accessible metric – differential permissibility. In the case of Kir2.1, differential permissibility pointed to the D-E and L-M loops as possessing latent allosteric potential. Insertion into the loops, even in the absence of light-modulation had the same phenotype (activating) as G-proteins binding to the equivalent regions of GIRK. This is strong support for the notion that latent allostery can explain the specific path an early inward rectifying channel took to evolve G-protein regulation in GIRK.

Interpreted broadly, mapping how permissibility–and by extension allostery–changes through phylogeny might be useful in explaining how the metazoan ion channel set diversified regulation. Perhaps latent allostery is an emergent property of recombination and sequence drift that defines evolutionary branch points by giving rise to emergent phenotypes that are subsequently exploited. Given this tantalizing possibility, further experiments are needed to establish how well differential permissibility is conserved along the evolutionary tree of ion channels.

To more practical interest, we demonstrate that inserting switchable domains only into existing allosteric sites or predicted latent allosteric sites, renders the target channel light-switchable. We do not yet know whether this is generalizable to multiple ion channels. However, what this proof-of-principle experiment demonstrates is that mapping latent allosteric sites via experimentally measured permissibility could greatly simplify engineering ion channels with useful function, such as sensitivity to biorthogonal stimuli.

References and Notes

1. S. T. Glantz *et al.*, Functional and topological diversity of LOV domain photoreceptors. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E1442–E1451 (2016).
2. L. Morsut *et al.*, Engineering Customized Cell Sensing and Response Behaviors Using Synthetic Notch Receptors. *Cell*. **164**, 780–791 (2016).
3. T. Sikosek, H. S. Chan, Biophysics of protein evolution and evolutionary protein biophysics. *Journal of The Royal Society Interface*. **11**, 20140419–20140419 (2014).
4. S. M. Coyle, J. Flores, W. A. Lim, Exploitation of latent allostery enables the evolution of new modes of MAP kinase regulation. *Cell*. **154**, 875–887 (2013).
5. T. Sikosek, H. S. Chan, E. Bornberg-Bauer, Escape from Adaptive Conflict follows from weak functional trade-offs and mutational robustness. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 14888–14893 (2012).
6. J. T. Bridgham, S. M. Carroll, J. W. Thornton, Evolution of hormone-receptor complexity by molecular exploitation. *Science*. **312**, 97–101 (2006).
7. G. Guntas, M. Kanwar, M. Ostermeier, Circular Permutation in the Ω-Loop of TEM-1 β-Lactamase Results in Improved Activity and Altered Substrate Specificity. *PLoS ONE*. **7**, e35998 (2012).

8. A. Mittermaier, L. E. Kay, New Tools Provide New Insights in NMR Studies of Protein Dynamics. *Science*. **312**, 224–228 (2006).
9. J. Lee *et al.*, Surface sites for engineering allosteric control in proteins. *Science*. **322**, 438–442 (2008).
10. O. Dagliyan *et al.*, Engineering extrinsic disorder to control protein activity in living cells. *Science*. **354**, 1441–1444 (2016).
11. H. Hibino *et al.*, Inwardly rectifying potassium channels: their structure, function, and physiological roles. *Physiol. Rev.* **90**, 291–366 (2010).
12. T. Hattori *et al.*, A novel gain-of-function KCNJ2 mutation associated with short-QT syndrome impairs inward rectification of Kir2.1 currents. *Cardiovascular Research*. **93**, 666–673 (2012).
13. M. Nishida, M. Cadene, B. T. Chait, R. MacKinnon, Crystal structure of a Kir3.1-prokaryotic Kir channel chimera. *EMBO J.* **26**, 4005–4015 (2007).
14. X. Tao, J. L. Avalos, J. Chen, R. MacKinnon, Crystal Structure of the Eukaryotic Strong Inward-Rectifier K+ Channel Kir2.2 at 3.1 Å Resolution. *Science*. **326**, 1668–1674 (2009).
15. M. R. Whorton, R. MacKinnon, Crystal Structure of the Mammalian GIRK2 K+ Channel and Gating Regulation by G Proteins, PIP2, and Sodium. *Cell*. **147**, 199–208 (2011).
16. B. Z. Harris, W. A. Lim, Mechanism and role of PDZ domains in signaling complex assembly. *J. Cell. Sci.* **114**, 3219–3231 (2001).
17. B. L. Oakes, D. C. Nadler, A. Flamholz, C. Fellmann, Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. *Nature*. **34**, 646–651 (2016).
18. A. Taslimi *et al.*, An optimized optogenetic clustering tool for probing protein interaction and function. *Nat Commun*. **5**, 4925 (2014).
19. W. N. Green, Ion channel assembly: creating structures that function. *J. Gen. Physiol.* **113**, 163–170 (1999).
20. C. Stockklausner, J. Ludwig, J. P. Ruppersberg, N. Klocker, A sequence motif responsible for ER export and surface expression of Kir2.0 inward rectifier K+ channels. *FEBS Lett.* **493**, 129–133 (2001).
21. S. B. Hansen, X. Tao, R. MacKinnon, Structural basis of PIP2 activation of the classical inward rectifier K+ channel Kir2.2. *Nature*. **477**, 495–498 (2011).
22. G. M. Martin *et al.*, Cryo-EM structure of the ATP-sensitive potassium channel illuminates mechanisms of assembly and gating. *Elife*. **6**, 213 (2017).
23. M. R. Whorton, R. MacKinnon, X-ray structure of the mammalian GIRK2-βγ G-protein complex. *Nature*. **498**, 190–197 (2017).

ACKNOWLEDGEMENTS

We thank Matthew R. Whorton, Mikael Elias, Sivaraj Sivaramakrishnan, and the entire Schmidt lab for helpful feedback and discussion, Therese Martin with flow cytometry technical advice and support, David Savage and Avi Flamholz for technical advice and development of the DIPseq alignment pipeline, Alina Zdechlik, Tejas Gupte, and Daniel Sorenson for helpful feedback on the manuscript, and Steffan Okorafor for assistance with single mutant construction. W.C.M. is funded by a National Science Foundation Graduate Research Fellowship.

SUPPLEMENTARY MATERIALS

Material and Methods
Fig. S1-S7
References (24-34)

FIGURE LEGENDS

Figure 1: Domain Insertion Permissibility. (A) Cartoon model for exploited and latent allostery. In addition to existing pathways (solid line) that communicate the binding of an allosteric modulator to a distal effector site or domain, there exist untapped –latent– pathways (dashed lines). Latent allosteric sites can be discovered by domain insertion (blue shapes) because they possess context-dependent plasticity. **(B)** Permissibility data is mapped on the crystal structure of chicken Kir2.2 (PDB 3SPC). Sites sensitive to the structural context of the inserted domain (differential permissibility) are indicated in red. Vignettes provide detail for helix bundle crossing and G loop (**Detail A**), as well as D-E loop and surface-exposed residues in the G-H loop (**Detail B**). **(C)** The primary sequence of human Kir2.1 (GI: 4504835) and secondary-structure elements are shown along the permissibility score for three types of inserted domains (indicated on the left). Key residues with functional relevance are indicated by color-coded spheres below.

Figure 2: Parameter Correlation and Model Performance. (A) Spearman correlations between permissibility for the inserted domain indicated left with the calculate property indicated on the top. Vertical bars separate feature category (Static, Conservation, and Dynamic). **(B)** Model performance determined with receiver operator characteristic (ROC) curves for the indicated domain, along with the top predictive properties used in each recursive 10-fold cross-validated decision tree model. Examples of ROC curves for models with different performance are shown on the left-most panel.

Figure 3: Domain Insertion Impact on Kir2.1 Function. (A) Population-level DiBAC4(3) fluorescence in HEK293FT cells expressing Kir2.1 as a function of external K^+. Increasing K^+ depolarizes the cells, resulting in less membrane-partitioning of the dye thus increasing measured fluorescence. **(B)** Shown are the percent hyperpolarized cells expressing the indicated Kir2.1 variant and inserted domain. Higher numbers indicate function, and lower numbers indicate disruption. Reference measurements are provided for HEK239FT expressing miRFP670 alone and wildtype Kir2.1 coexpressed with miRFP670 (yellow box). Reference levels of WT and no channel are indicated by horizontal green and red lines, respectively. Regions discussed in the text are indicated by black boxes.

Figure 4: Light-modulated Kir2.1 variants. (A) Representative examples showing the difference (indicated by red areas) in population-level DiBAC4(3) fluorescence with and without blue light illumination for HEK293FT cells expressing the indicated Cib81 insertion mutant or WT channel and Cry2. **(B)** Quantitation of dissimilarity (X^2) compared to channel function. Highlighted are insertions into the PIP_2 binding site and αH helix (yellow), D-E loop (blue), gating mutant V302M (green), and wildtype Kir2.1 (red). Significance of light modulation is tested by pairwise comparisons using Dunnett's test for multiple comparisons with wild type as control and post-hoc multiple comparison adjustment. Significance levels: *** $p < 0.001$; ** $p < 0.01$, $p > 0.05$ (not significant) **(C)** Open Probability determined by on-cell patch clamp electrophysiology for wt Kir2.1 or the indicated insertion mutant. Error bars are s.e.m. (n = 3–12). **(D)** Gaussian functions ($Kir2.1_{A362CIB}$, n = 6; $Kir2.1_{P186CIB}$ n = 2) fit to all points current amplitude histograms for representative examples of the indicated insertion mutant. Before, during, and recovery (>2 min

post-illumination) are indicated by black, red, and blue lines, respectively. Representative current traces are shown on the right. Dashed lines indicate zero current level. (**E**) A coarse classification of Kir2.1 into 'unimportant' sites, which have high permissibility to any insertion (yellow), 'scaffold' sites, which due to biophysical constraint are non-permissive (red), and allosteric sites, which display differential permissibility (green). Names indicate the corresponding secondary element.

Figure 1

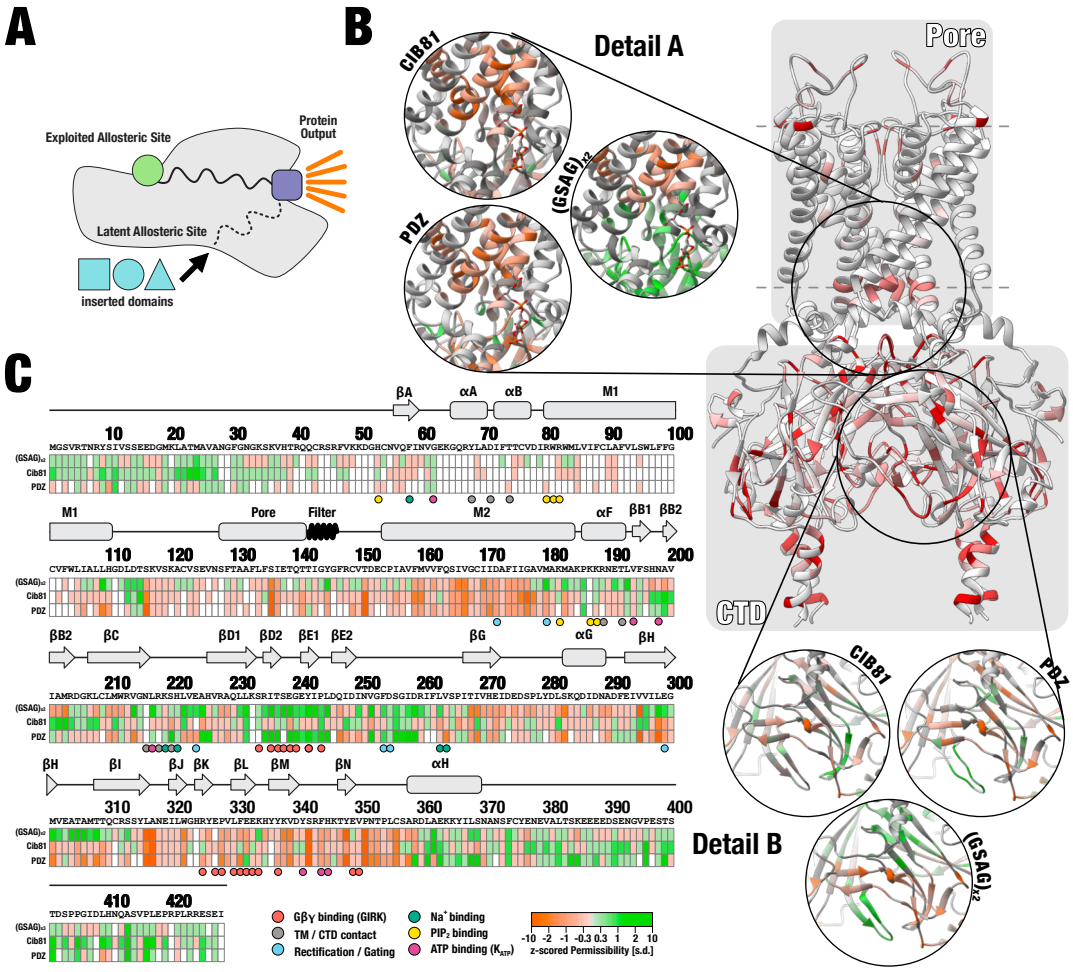

A
Exploited Allosteric Site
Protein Output
Latent Allosteric Site
inserted domains
B
Detail A
CIB81
PDZ
(GSAG)x2
Pore
CTD
Detail B
C
βA αA αB M1
M1 Pore Filter M2 αF βB1 βB2
βB2 βC βD1 βD2 βE1 βE2 βG αG βH
βH βI βJ βK βL βM βN αH
(GSAG)x2
Cib81
PDZ
Gβγ binding (GIRK)
TM / CTD contact
Rectification / Gating
Na+ binding
PIP2 binding
ATP binding (KATP)
-10 -2 -1 -0.3 0.3 1 2 10
z-scored Permissibility [s.d.]

Figure 2

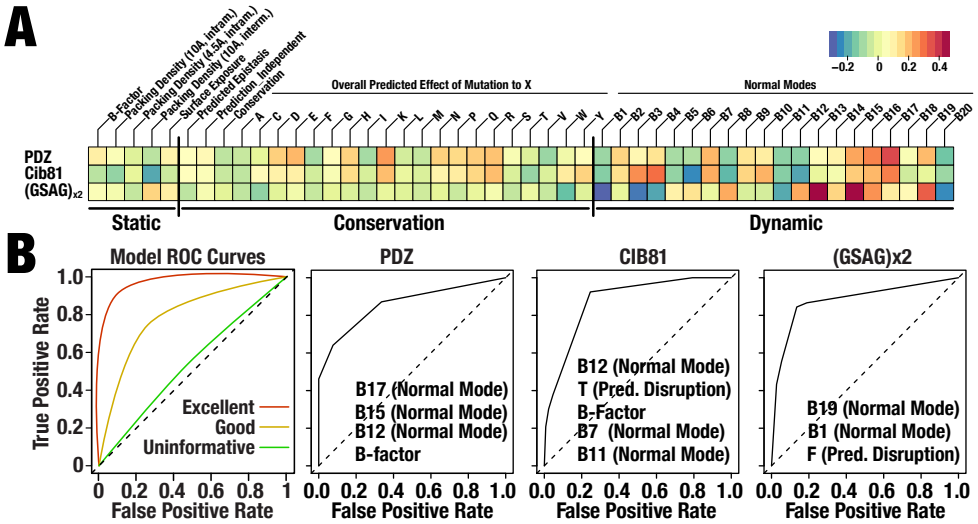

A
B-Factor
Packing Density (10Å, intram.)
Packing Density (4.5Å, intram.)
Packing Density (10Å, interm.)
Surface Exposure
Predicted Epistasis
Prediction, Independent
Conservation
Overall Predicted Effect of Mutation to X
A C D E F G H I K L M N P Q R S T V W Y
Normal Modes
B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 B11 B12 B13 B14 B15 B16 B17 B18 B19 B20
−0.2 0 0.2 0.4
PDZ
Cib81
(GSAG)x2
Static
Conservation
Dynamic
B
Model ROC Curves
PDZ
CIB81
(GSAG)x2
True Positive Rate
False Positive Rate
Excellent
Good
Uninformative
B17 (Normal Mode)
B15 (Normal Mode)
B12 (Normal Mode)
B-factor
B12 (Normal Mode)
T (Pred. Disruption)
B-Factor
B7 (Normal Mode)
B11 (Normal Mode)
B19 (Normal Mode)
B1 (Normal Mode)
F (Pred. Disruption)

Figure 3

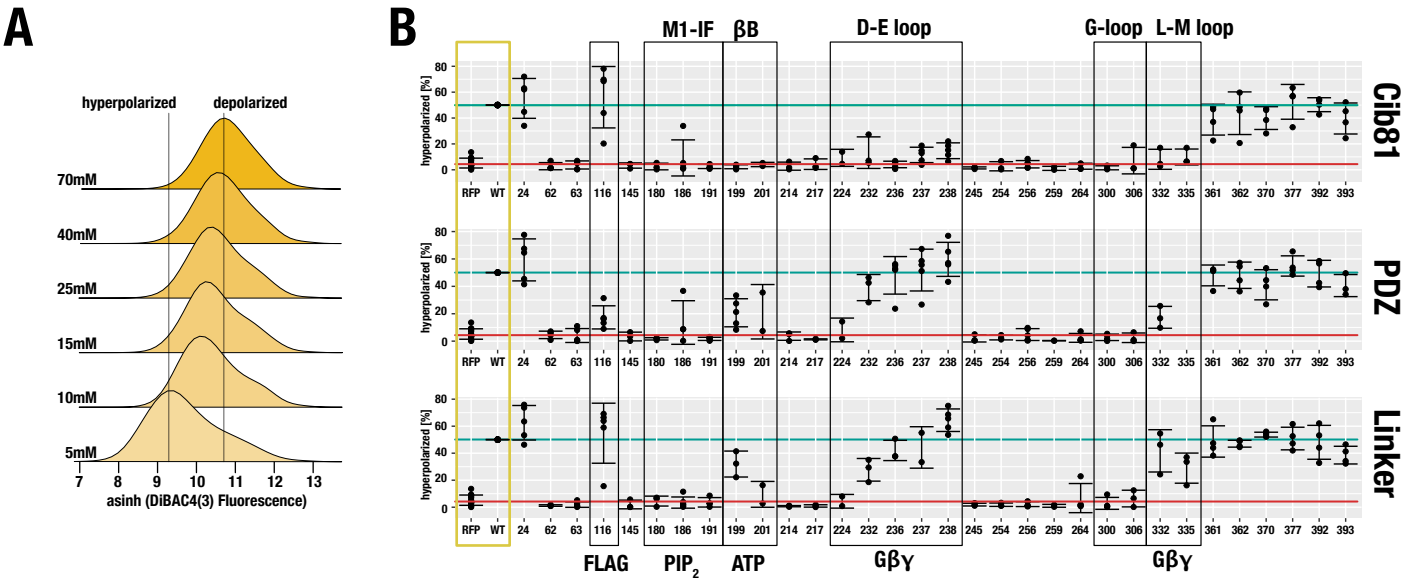
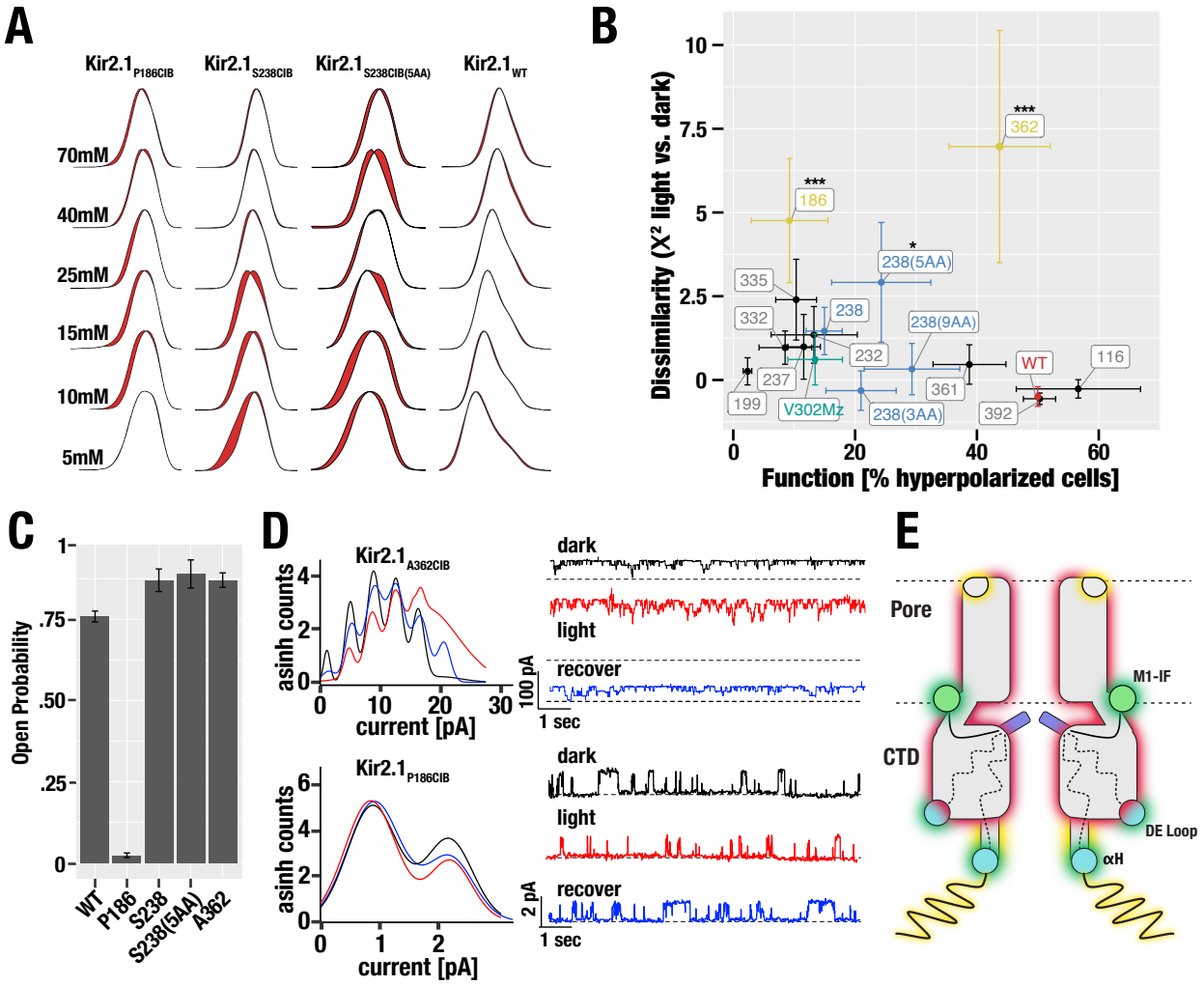

A
hyperpolarized
depolarized
70mM
40mM
25mM
15mM
10mM
5mM
7 8 9 10 11 12 13
asinh (DiBAC4(3) Fluorescence)
B
M1-IF
βB
D-E loop
G-loop
L-M loop
hyperpolarized [%]
0 20 40 60 80
RFP WT 24 62 63 116 145 180 186 191 199 201 214 217 224 232 236 237 238 245 254 256 259 264 300 306 332 335 361 362 370 377 392 393
Cib81
PDZ
Linker
FLAG
PIP2
ATP
GβY
GβY

Figure 4


A
Kir2.1 P186CIB
Kir2.1 S238CIB
Kir2.1 S238CIB(5AA)
Kir2.1 WT
70mM
40mM
25mM
15mM
10mM
5mM
B
Dissimilarity (X² light vs. dark)
Function [% hyperpolarized cells]
362
186
238(5AA)
335
332
238
238(9AA)
232
237
199
V302Mz
238(3AA)
361
WT
116
392
C
Open Probability
WT
P186
S238
S238(5AA)
A362
D
Kir2.1 A362CIB
asinh counts
current [pA]
dark
light
recover
100 pA
1 sec
Kir2.1 P186CIB
2 pA
E
Pore
M1-IF
CTD
DE Loop
αH

Supplementary Materials for

Domain Insertion Permissibility is a Measure of Engineerable Allostery in Ion Channels

Willow Coyote-Maestas[1], Yungui He[2], Chad L. Myers[3] & Daniel Schmidt[2*]

[1]Dept. of Biochemistry, Molecular Biology & Biophysics, [2]Dept. of Genetics, Cell Biology & Development, [3]Dept. of Computer Science and Engineering, University of Minnesota, Minneapolis, MN

*To whom correspondence should be addressed: schmida@umn.edu

This PDF file includes:

Materials and Methods
Figs. S1 to S7
References (24-30)

Materials & Methods

MuA domain insertion library generation

Transposition libraries were generated using 100ng MuA-BsaI engineered transposon and 1:2 molar ratio transposition target DNA in 20ul reactions with 4ul 5x MuA reaction buffer and 1ul 0.22 ug/ul MuA transposon (Thermo Fisher). MuA-BsaI engineered transposon propagation plasmid or pUCKanR-Mu-BsaI was a gift from David Savage (Addgene plasmid # 79769) (*25*). MuA-BsaI engineered transposon was digested with BglII and HindIII Fastdigest enzymes (Thermo Fisher) and gel purified using gel purification kit (Zymo Research).

The transposition target, human Kir2.1 (GI: 4504835) including a porcine teschovirus ribosomal skipping sequence (P2A) (*26*), was codon-optimized for mouse, synthesized (Gen9) and subcloned into pATT-Dest using NEB BamHI and HindIII**.** pATT-Dest was a gift from David Savage (Addgene plasmid # 79770) (*25*). A FLAG tag was inserted after T115 using Q5 site-directed mutagenesis. MuA transposition reactions were incubated at 30degC for 18 hours for transposition, followed by 75degC for 10 minutes for heat inhibition. DNA from reactions was cleaned up (Zymo Research) and eluted in 10ul water. All 10ul were transformed into 30ul electrocompetent 10G ELITE E. coli (Lucigen) in 1.0 mm Biorad cuvettes using a Bio-Rad Gene Pulser II electroporator (settings: 10uF, 600 Ohms, 1.8 kV). Cells were rescued and grown without antibiotics for 1 hour at 37degC. Aliquots were then serially diluted and plated on LB agar plates containing carbenicillin (100 ug/ml) and chloramphenicol (25 ug/ml) to assess library coverage. The remaining transformation mix was grown in 50 ml LB containing carbenicillin (100 ug/ml) and chloramphenicol (25 ug/ml). All transformed libraries yielded greater than 10^5 colonies, which for Kir2.1-P2A (1369bp) is >35x coverage. Plasmid DNA was purified by midi-prep kit (Zymo Research).

Transposition-inserted Kir2.1 variants were subcloned into an expression vector by amplifying channel variant genes, adding on BsmbI sites, using 10 cycles of PCR using Primestar GXL (Takara Clontech) and run on a 1% agarose gel. The larger band was cut out and gel purified (Zymo Research) to isolate channels with inserted transposons. A mammalian expression vector (pcDNA3.1) with EGFP was amplified to add on BsmbI sites complementary to those on Kir2.1-P2A. The Kir2.1-P2A (BsaI-transposon) variants where subcloned into this vector by BsmbI-mediated Golden Gate cloning (*27*). Reactions were cleaned (Zymo Research) and eluted with 10ul water. All 10ul were transformed into 30 ul Lucigen electrocompetent 10G ELITE E. coli and electroporated in 1.0 mm Biorad cuvettes using a Bio-Rad Gene Pulser II electroporator (settings: 10 uF, 600 Ohms, 1.8 kV). Cells were rescued and grown without antibiotics for 1 hour at 37degC then with an aliquot serially diluted plated on LB agar plates containing kanamycin (50 ug/ml) and chloramphenicol (25 ug/ml) to assess library coverage. The remaining transformation mix was grown in LB containing kanamycin (50 ug/ml) and chloramphenicol (25 ug/ml). All transformed libraries yielded greater than 10^5 colonies so for Kir2.1 (1369bp) there is >35x coverage. Plasmid DNA was purified by midi-prep kit (Zymo Research).

Inserted Transposons were replaced with domains in individual reactions using BsaI-mediated Golden Gate cloning (*27*). Domains (PDZ, Cib81, $GSAG_{x2}$, $GSAG_{x3}$) for insertions were ordered

as gblocks (IDT DNA), and PCR amplified to add on BsaI sites complementary to MuA-BsaI transposon sites for Golden Gate cloning. Domain amplicons were gel purified (Zymo Research). The product was further digested with AgeI-HF (NEB) and Plasmid-Safe ATP-dependent DNase (Epicentre) to remove any undigested transposon, then cleaned up (Zymo Research) and eluted with 10ul water. All 10ul were transformed into 30 ul Lucigen electrocompetent 10G ELITE E. coli and electroporated in 1.0 mm Biorad cuvettes using a Bio-Rad Gene Pulser II electroporator (settings: 10 uF, 600 Ohms, 1.8 kV). Cells were rescued and grown without antibiotics for 1 hour at 37degC. An aliquot was serially diluted and plated LB agar plates containing kanamycin (50 ug/ml) to assess library coverage. The remaining transformation mix was grown in LB containing kanamycin (50 ug/ml). All transformed libraries yielded greater than 10^5 colonies so for Kir2.1 (1369bp) there is >35x coverage. Plasmid DNA was purified by midi-prep kit (Zymo Research).

Domain insertion permissibility cell sorting assay

100 ng of each domain insertion library was transfected with 36 ul of turbofect (Thermo Fisher) into 50% confluent HEK293FT(Invitrogen) with 11.9 ug of dummy plasmid (pATT Dest) divided across a single 6 well dish (9.6 cm^2 / well).

Cells from each well were detached using 1 ml accutase (Stemcell Technologies) and twice spun down at 450xg and resuspended in FACS buffer (2% of FBS, 0.1% NaN3, 1xPBS). Cells were incubated with 1:200 anti-flag mouse antibody (Sigma) 1 hour rocking at 4degC, washed twice with FACS buffer, covered with aluminum foil, and then incubated with 1:400 anti-mouse Alexa Fluorophore 568 (Thermo Fisher) for 30 minutes rocking at 4degC. Cells were washed twice, resuspended in 3 ml FACS buffers, and filtered using cell strainer 5 ml tubes (Falcon). Cells were kept on ice and protected from light in the transfer to the flow cytometry core. Before cell sorting, a small aliquot of cells was saved as a control sample for sequencing.

Cells were sorted into EGFP high / Alexa568 low (transfected cells without surface expression) and EGFP high / Alexa Fluorophore 568 high (transfected cells with surface expression) on a BD FACSAria II P69500132 flow cytometer. EGFP fluorescence was excited using a 488 nm laser, recorded with a 525/50 bandpass filter and a 505 long pass filter. Alexa fluorophore 568 fluorescence was excited using a 561 nm laser and recorded with a 610/20 bandpass filter. Cells were gated on Side Scattering and Forward Scattering to separate out whole HEK293FT cells, gated on forward scattering area and width to separate single cells, then gated on coexpressed EGFP to gate out cells that received a plasmid, then gated on cells that were labeled using the anti-flag antibody for surface expressed channels. Gates were determined using single wildtype, EGFP only and unstained library samples. EGFP high / Label low and EGFP high / Label high cells were collected into catch buffer (20% of FBS, 0.1% NaN3, 1xPBS). Between 2,000-100,000 cells were collected for each sample/library pair which is ~4-250x coverage of all potentially productive (i.e., in-frame and forward) domain insertions.

DNA from Control, EGFP high / Label low, and EGFP high / Label high cells for each library were extracted using a Microprep DNA kit (Zymo Research) and triple eluted with water. To remove chromosomal DNA, samples were digested with Plasmid-Safe ATP-dependent DNase (Epicentre). The resulting plasmid DNA was further purified and concentrated using (Zymo

Research). The product was used as a template for 12 cycles of PCR using Primestar GXL (Takara Clontech), run on a 1% agarose gel, and gel purified (Zymo Research) to remove any primer dimers or none amplicon DNA. Purified DNA was quantified using picogreen DNA concentration at the University of Minnesota Genomics Core. Equal amounts of each domain insertion sample were pooled by cell sorting category (control, EGFP high / Label low, EGFP high / Label high were pooled for sequencing library generation and sequencing.

Sequencing

Libraries were generated at University of Minnesota Genomics Core using Nextera XT or Nano Truseq library generation (Illumina) to fragment and add on Illumina sequencing adaptors, and sequenced using either HISEQ or MISEQ sequencing platforms.

Domain insertion permissibility alignment and enrichment

Alignments were done on both forward and reverse reads using a DIPseq pipeline developed by David Savage and coworkers that we slightly modified for compatibility with updated python packages (*17*). Reads with duplicate domain insertion calls in both forward and reverse reads were removed. This pipeline results in plaintext files indicating a domain insertion positions and whether that insertion is in-frame and in the forward direction. Enrichment was calculated by comparing the change in EGFP high /Label low to EGFP high / Label high cells. Only positions with reads in both samples were used in enrichment calculations. All these positions are treated as 'NA' and not considered in downstream analysis and structure mappings, with the exception of calculating correlations between datasets and correlations between sites. In these correlation calculations treat 'NA's as '0's so removing all the data will introduce more noise when comparing between datasets due to limits from sampling.

Fitness function for individual datasets:

$$F(i,j) = \frac{r_j^i[label]}{t_j[label]} - \frac{r_j^i[GFP]}{t_j[GFP]}$$

Where *r* is the number of reads at amino acid position *i*, in the *j*th dataset divided by *t*, the total number of reads in the *j*th given sample. This resulting data from individual sequencing reads are only used to calculate correlations between domains and amino acid positions.

For structure mappings and predictive model training means of fitness for a given domain insertion variant are used. So, the resulting mean fitness function is:

$$G(i,j) = \frac{\sum \frac{r_j^i[label]}{t_j[label]} - \frac{r_j^i[GFP]}{t_j[GFP]}}{n}$$

Where *r* is the number of reads at amino acid position *i*, in the *j*th dataset divided by *t,* the total number of reads in that given sample, summed for all replicates of that domain-channel combination, and divided by *n*, the number of datasets.

Differential fitness was calculated as the mean fitness standard deviation between Cib81, PDZ and $GSAG_{x1}$ datasets.

Mean fitness and differential fitness was z-scored and mapped onto the structure of chicken Kir2.2 (PDB 3SPC) using Chimera (*31*). Mapped dataset for Cib81, PDZ and GSAGx1 linker had adequate coverage: 76.6% Cib81(333/435), 68.9% PDZ (300/435), and 76.7 % (334/435) of potential amino acid positions.

Dataset Correlations

Pearson correlations were used to calculate correlations between domain insertion datasets. Pearson correlations were also used to calculate correlations between amino acid positions across all datasets. Both these correlation matrices were calculated using the dataset that was trimmed to avoid sampling problems such that no more than 0.375 (6/16) datasets are where raw fitness was +10e-4 <x< -10e-4. These correlations were calculated with 63% of possible positions (277/435).

Spearman correlations were used to compare mean domain insertion datasets and calculated protein properties because Spearman correlations are often better at handling non-parametric correlations. Based on lack of structural and conservational data at various amino acid positions many sites had to be trimmed. Data was trimmed from positions where more than half datasets had a raw mean fitness was +10e-4 <x< -10e-4. This resulted in datasets that contained 70% of possible positions (207/293).

Computed protein properties

Static protein properties (B-factor, 10 Angstrom intramonomeric packing density, 4.5 Angstrom intramonomeric packing density, 4.5 Angstrom intermonomeric packing density, and surface exposure) were calculated using the SWIFT web server on chicken Kir2.2 (PDB: 3SPC**)** (*21*, *28*), conservation based properties (overall predicted disruptive effect of a mutation, conservation, and individual predicted disruptive effect of a mutation to specific amino acids (A, C, D, E, F, G, H, I, K, L, M, P, Q, R, S, T, V, W, Y)) were acquired from the EVmutation data server (*29*), and dynamic properties (first 20 normal modes of the B monomer) were calculated on the iGNM 2.0 normal mode webserver (*34*). After trimming, computed protein properties were calculated for all parameters for 67% (293/435) of possible positions.

Recursive decision tree models

We chose decision trees for building predictive models due to their utility in handling and determining non-linear interactions. Prior to training models, data was binarized such that 0 was not permissive and 1 permissive for a given domain insertion. After trimming any data for a given mean raw fitness between +10e-4 <x< -10e-4 datasets models were trained on 62.5% (183/293) Cib81 ,58.0% (170/293) PDZ, and 68.6% (201/293) of possible positions. Models were limited to

a depth of 4 to minimize model overfitting, trained on the computed protein properties to predict Cib81, PDZ, and $GSAG_{x2}$ using the rpart package in R (*30*), and cross-validated 10 times. Model performance was determined using commonly used criteria: receiver operating characteristic (ROC) curves, the complexity parameter (used in minimizing tree size)/tree depth and model residuals, precision vs. recall, accuracy vs. cutoff, precision vs. cutoff, and recall vs. cutoff (Supplemental Fig4). As further validation, models were trained only the four most significant protein properties based on Spearman correlations to demonstrate necessity and utility of using decision trees vs correlation calculations (**Supp. Fig. 5A),** by withholding the most important properties determined with decision tree and by withholding whole classes of protein properties (static, conservation, dynamic, **Supp. Fig. 5B**).

Resting membrane potential functional assay

Single mutants were generated by inserting a BsaI site and a 5 basepair replication identical to those created by transposons that replicated the beginnings and ends of transposon-mediated domain insertions using a Q5 site-directed mutagenesis kit (NEB). Single insertion mutants were created for 32 sites (Amino Acid positions: 24, 62, 63, 116, 145, 180, 186, 191, 199, 201, 214, 217, 224, 232, 236, 237, 238, 245, 254, 256, 259, 264, 300, 306, 332, 335, 361, 362, 370, 377, 392, 393) and then replaced with the domains for which libraries were previously generated. Subsequently, using BsrGI and PstI sites, EGFP was replaced with miRFP670 for all mutants. miRFP670 was amplified from pmiRFP670-N1 which was a gift from Vladislav Verkhusha (Addgene plasmid # 79987) (*31*). The same cloning approach was used to add 3-9 amino acid GSG linkers on either side of Cib81 in the 238 position.

The resting membrane potential assay was conducted on all aforementioned domain insertion mutants in addition to miRFP670 alone (negative control) and wildtype Kir2.1 (positive control). 400 ng of each mutant was transfected with 6 ul Polyethyleneimine (Polysciences) along with 600 ng of dummy plasmid (pATT Dest) across 2 wells of a 24 well dish. For each experiment, wildtype Kir2.1 was transfected as a benchmark and for comparison for mutant function. Cells from each well were detached using 300 ul accutase (Stemcell Technologies), spun down at 450xg three times, and re-suspended in 200 ul Tyrode (125mM NaCl, 2mM KCl, 3mM CaCl2, 1mM MgCl, 10mM HEPES, 30mM glucose, pH 7.3). Bis-[1,3-dibutylbarbituric acid] trimethine oxonol (DiBAC4(3), Thermo Fisher) was added to a final concentration of 950 nM, and cells were filtered in 5 ml cell strainer tubes (Falcon). DiBAC4(3) was diluted every day to exchange buffers from DMSO to Tyrode (*32*). Cells were kept on ice and protected from light in the transfer to the flow cytometry core.

Each sample was run in entirety on a BD Fortessa H0081 flow cytometer. DiBAC4(3) was excited at 488 nm and recorded at 525/50 bandpass, and miRFP670 fluorescence was excited at 640 nm and recorded with a 670/30 nm bandpass filter. Cells were gated on Side Scattering and Forward Scattering to separate out whole HEK293FT cells, gated on forward scattering area and width to separate single cells, then gated on coexpressed miRFP670 to gate out cells that received a plasmid, then a gate was set on the lower 50% of a histogram of wildtype Kir2.1 function, all mutants percentage of cells in this gate are reported. The analysis was performed in FlowJo 10 (FlowJo, LLC).

Flow Cytometry Assay for Light-modulation of Kir2.1 function

The generation of all single mutants used in the optogenetic switching assay was previously described in the resting membrane potential assay methods section. A Cry2-P2A-mKate2 domain was generated using gene fragments (Gen9) and assembled into the expression vector pEGFPN3 (Invitrogen) using BsmbI-mediated Golden Gate cloning (*27*). The Kir2.1(V302M) mutant was generated using Q5 site-directed mutagenesis (NEB).

The light-modulation assay was conducted for Cib81 mutants chosen as representative examples for the various permissibility and functional phenotypes we had observed. In addition, negative controls such as wildtype Kir2.1 and a pore dead mutant V302M (*33*) were included. 4 ug of each mutant, 3ug of dummy plasmid (pATT-Dest), and 100 ug of Cry2-P2A-mKate2 were transiently transfected using 6ul PEI across 16 wells of a 24 well dish at 20% confluency. Cells from each well were detached using 300 ml accutase, washed three times, and resuspended in 4ml Tyrode. DIBAC4(3) was added to a final concentration of 950 nM, and cells were filtered in cell strainer 5 ml tubes (Falcon). Filtered cells were divided into twelve 5 ml tubes (300ul each) and kept on ice and protected from light in the transfer to the flow cytometry core.

Cells expressing each mutant, wt Kir2.1, Kir2.1(V302M) were challenged by the addition of K-gluconate at different concentrations (5 mM, 10 mM, 15 mM, 25 mM, 40 mM, and 70 mM), with and without illumination (455 nm LED (Thorlabs), 30 seconds, 100% duty cycle, 100uW/mm^2). Each sample was run in entirety on a BD Fortessa H0081 flow cytometer. DIBAC4(3) was excited at 488 nm and recorded with 525/50 bandpass and 502 long pass filters, miRFP670 was excited at 640 nm and recorded with a 670/30 bandpass filter, and mKate2 was excited at 561 nm and recorded at 610/20 bandpass and 595 long pass filters.

Each sample was recorded for 5 minutes or until completion. Cells were gated on Side Scattering and Forward Scattering to separate out whole HEK293FT cells, gated on forward scattering area and width to separate single cells, then gated on co-expressed miRFP670 (Kir2.1 mutant) to gate out cells that received a mutant plasmid. For each paired sample (dark and light) a custom gate was created in the non-illuminated sample to include the 15% most hyperpolarized cells (using the flowStats package (*34*)). The number of events falling into this gate were then compared to the corresponding illuminated sample using the Chi-Squared test and reported as Dissimilarity (X^2, light vs. dark). Dissimilarities at different K^+ challenges were normalized to correct for photobleaching and averaged.

Patch Clamp Electrophysiology

HEK293FT cells were transiently transfected with Kir2.1 (wt) or Kir2.1 insertion mutant and Cry2-P2A-mKate using PEI. Cells were screened for mKate2 expression using a 565nm high-power LED (Thorlabs) filtered by a 560±40nm bandpass filter (Semrock) through a 40X lens. K^+ currents were recorded 36-48 hours post-transfection using on-cell patch clamp electrophysiology. Patches with clear channel activity were stimulated with blue (455nm) light delivered by a LED (Thorlabs) at 100 uW/mm^2 for 50 seconds at 100% duty cycle. Analog signals were filtered (2-5 kHz) using the built-in 4-pole Bessel filter of a Sutter Instrument IPA patch clamp amplifier,

digitized and stored. Bath solution contained 125mM NaCl, 2mM KCl, 3mM $CaCl_2$, 1mM $MgCl_2$, 10mM HEPES, 30mM glucose, adjusted to pH 7.3 with NaOH. The pipette solution contained: 125mM K-Gluconate, 8mM NaCl, 0.1mM $CaCl_2$, 0.6mM $MgCl_2$, 1mM EGTA, 10mM HEPES, 4mM Mg-ATP, 0.4mM Na-GTP, adjusted to pH 7.3 with KOH. Osmolarity was adjusted to 295 - 300 mOsm with sucrose. Electrodes were drawn from borosilicate patch glass (Warner Instruments) to a resistance of 2-6 MΩ.

Data analysis was done using custom R scripts. After correction for baseline drift, representative all points current amplitude histograms for sweeps (1) before, (2) during, and (3) after > 2 min after illumination were calculated. Histograms were fit to a sum of Gaussian functions (n = 2–6):

$$f(x) = \sum_{i=1}^{n} C_i \cdot e^{-(x-\mu_i)^2/2\sigma^2}$$

Where *f(x)* is the count of points in a bin with current amplitude *x*, with the mean μ, standard deviation σ, and *C* total number of points belonging to the *i*th component. In a case with a single channel, the open probability is calculated as:

$$P_o = \frac{\alpha}{\alpha+\beta}$$

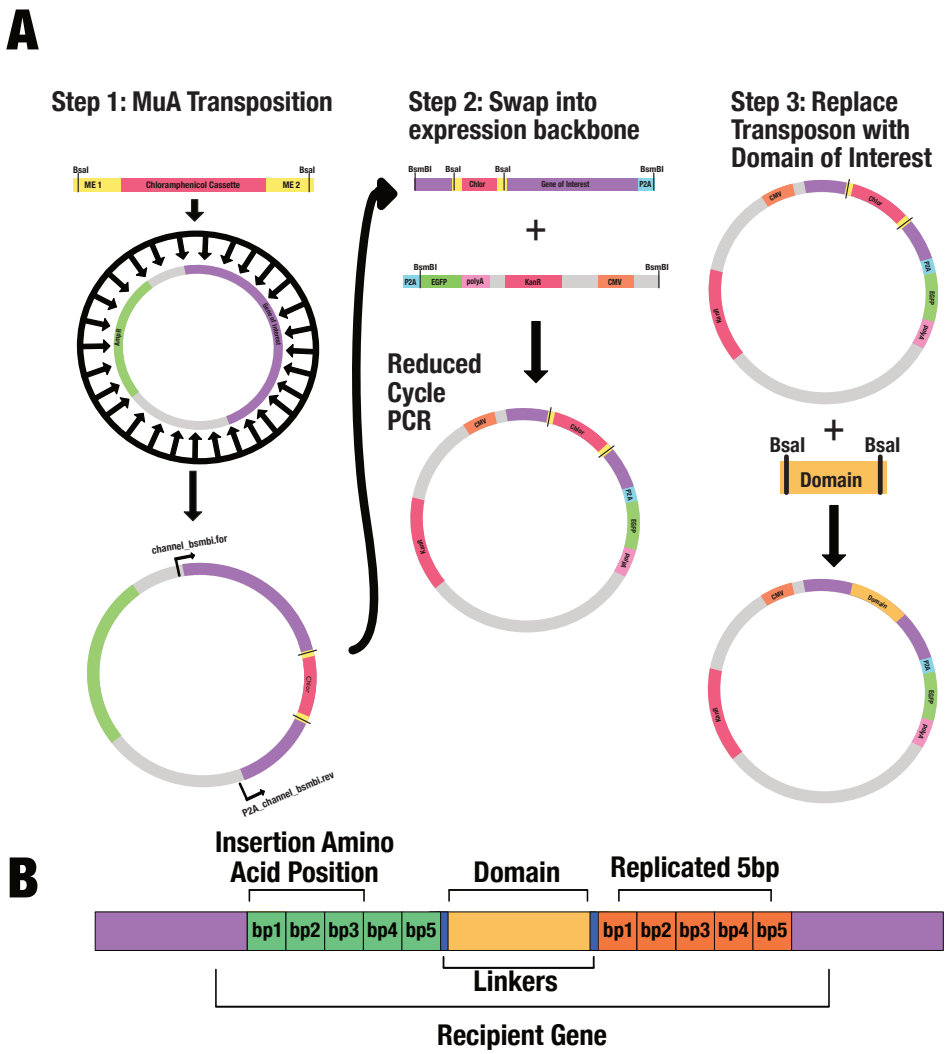
with α and β, the total number of points belonging to the Gaussians that describe the open and closed state, respectively.

References and Notes

24. N. J. Oviedo, C. L. Nicolas, D. S. Adams, M. Levin, Live Imaging of Planarian Membrane Potential Using DiBAC4(3). *CSH Protoc*. **2008**, pdb.prot5055–pdb.prot5055 (2008).
25. D. C. Nadler, S.-A. Morgan, A. Flamholz, K. E. Kortright, D. F. Savage, Rapid construction of metabolite biosensors using domain-insertion profiling. *Nat Commun*. **7**, 12266 (2016).
26. J. H. Kim *et al.*, High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS ONE*. **6**, e18556 (2011).
27. C. Engler, R. Gruetzner, R. Kandzia, S. Marillonnet, Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS ONE*. **4**, e5553 (2009).
28. G. Vriend, WHAT IF: A molecular modeling and drug design program. *Journal of Molecular Graphics*. **8**, 52–56 (1990).
29. T. A. Hopf *et al.*, Mutation effects predicted from sequence co-variation. *Nat. Biotechnol.* **35**, 128–135 (2017).
30. F. Hahne, N. Gopalakrishnan, A.H. Khodabakhshi, C. Wong, K. Lee, flowStats: Statistical methods for the analysis of flow cytometry data. R package version 3.38.0. (2018)
31. D. M. Shcherbakova *et al.*, Bright monomeric near-infrared fluorescent proteins as tags and biosensors for multiscale imaging. *Nat Commun*. **7**, 12405 (2016).
32. D. S. Adams, M. Levin, Measuring resting membrane potential using the fluorescent voltage reporters DiBAC4(3) and CC2-DMPE. *Cold Spring Harbor Protocols*. **2012**, 459–464 (2012).

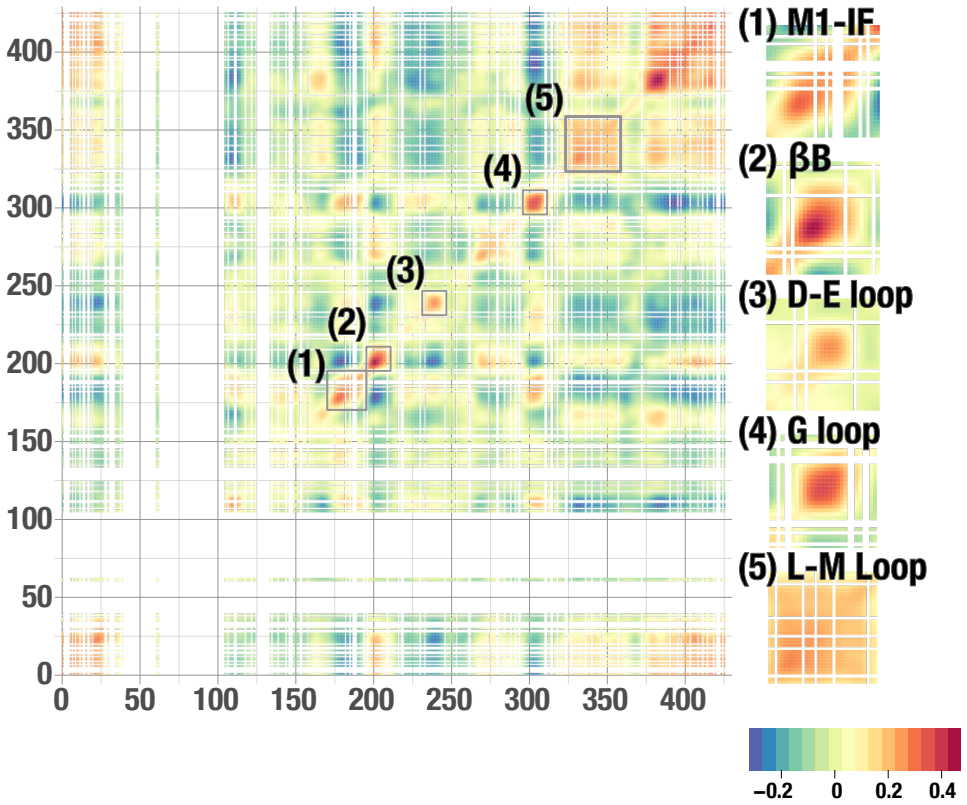
33. D. Ma, X. D. Tang, T. B. Rogers, P. A. Welling, An andersen-Tawil syndrome mutation in Kir2.1 (V302M) alters the G-loop cytoplasmic K+ conduction pathway. *J. Biol. Chem.* **282**, 5781–5789 (2007).
34. F. Hahne, N. Gopalakrishnan, A.H. Khodabakhshi, C. Wong, K. Lee, flowStats: Statistical methods for the analysis of flow cytometry data. R package version 3.38.0. (2018)

Supplemental Figure 1



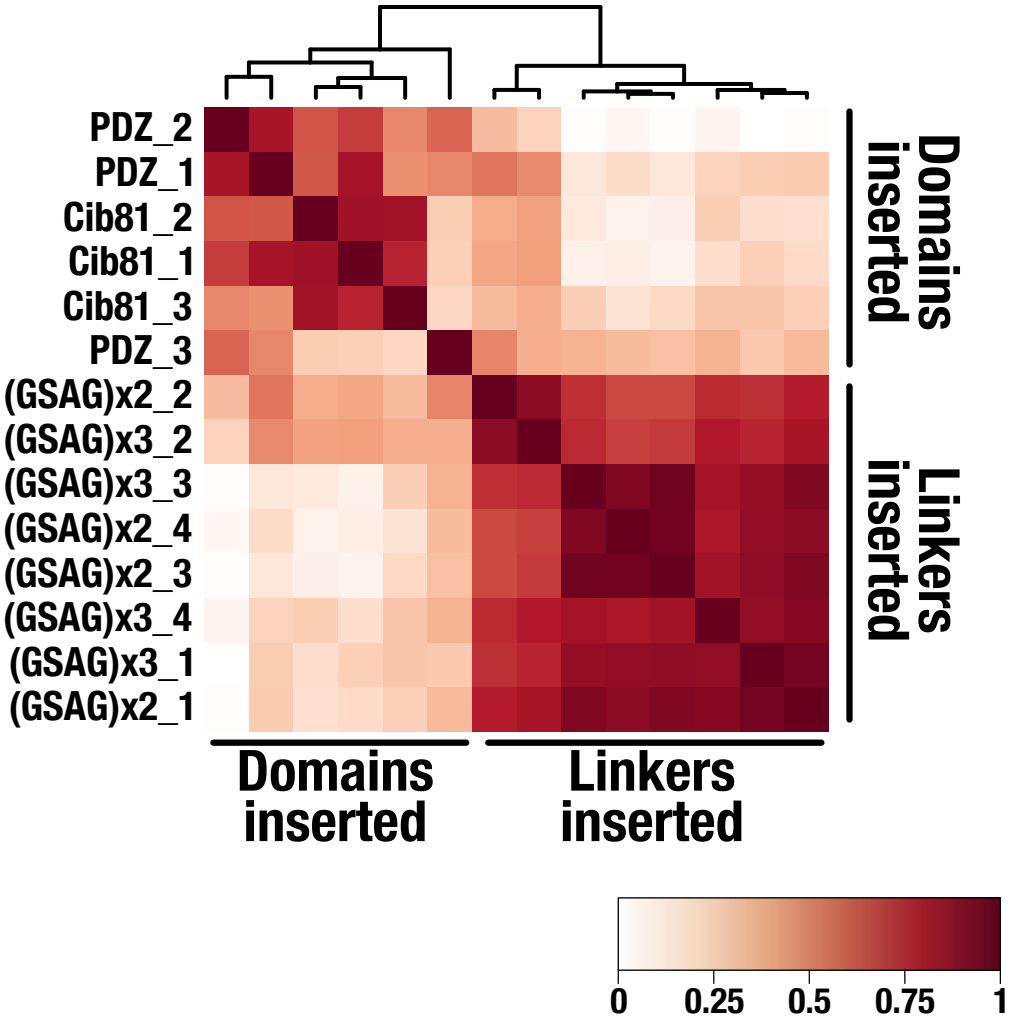
Supplemental Figure 1: Insertion Library Construction. (A) Libraries were generated in three cloning and selection steps. (1) Select for a MuA transposase delivered engineered MuA transposon with a chloramphenicol antibiotic cassette in a plasmid carrying Kir2.1. Flanking the antibiotic cassette are the beginnings and ends of flexible linkers with golden gate compatible BsaI type IIS restriction sites. (2) Reduced cycle PCR amplify and add on golden gate compatible BsmBI type IIS restriction sites and size separate channel genes with inserted transposons from those without transposons. Insert channel gene into a mammalian expression vector in-frame with a P2A-EGFP cassette. (3) Replace the transposon with a PCR amplified domain of interest with complementary BsaI sites and linkers using BsaI mediated golden gate cloning. **(B)** Architecture of a domain insertion position: At the position of the domain insertion the five positions upstream are replicated on the other side of the transpositions; domain insertion positions are identified as the last full codon coding for an amino acid or in other words that corresponding to the amino acid coded by bp1-3 of the replicated sequence. Domains are inserted with linkers to bring it into frame after insertion at +2 reading frame relative to the coding sequence.

Supplemental Figure 2



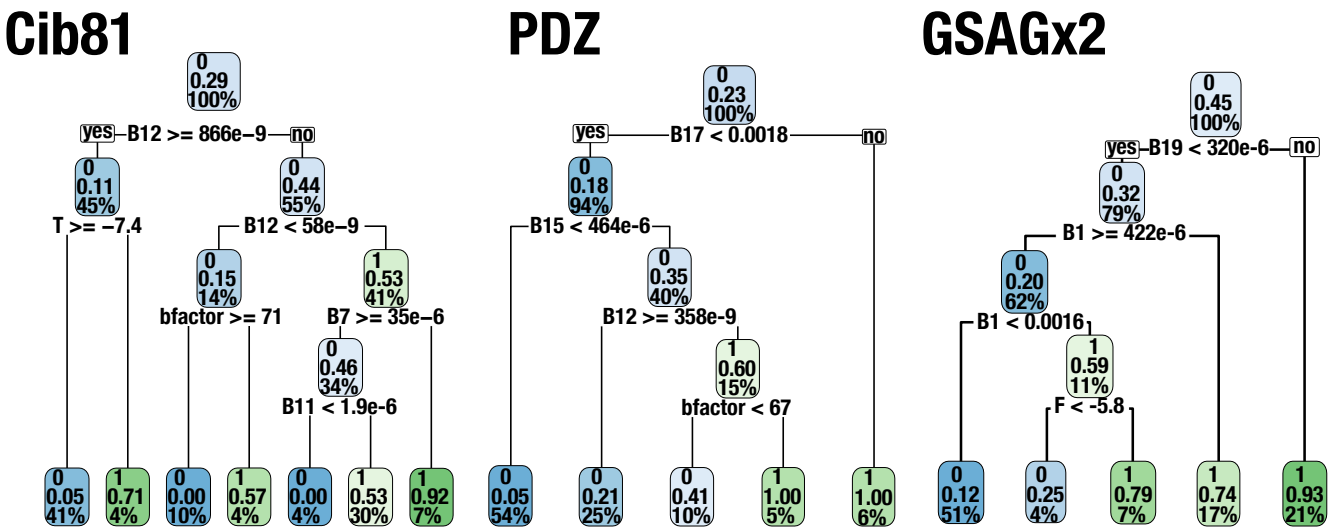
Supplemental Figure 2: Site-specific permissibility correlations. In Kir2.1, permissibility is highly correlated in (1) M1-IF – the PIP_2 binding site, (2) the βB loop where ATP binds in Kir6.2, (3 & 5) the D-E and L-M loops where Gβγ binds in GIRK, as well as (4) the G loop involved in channel gating. White indicated sites with incomplete data.

Supplemental Figure 3



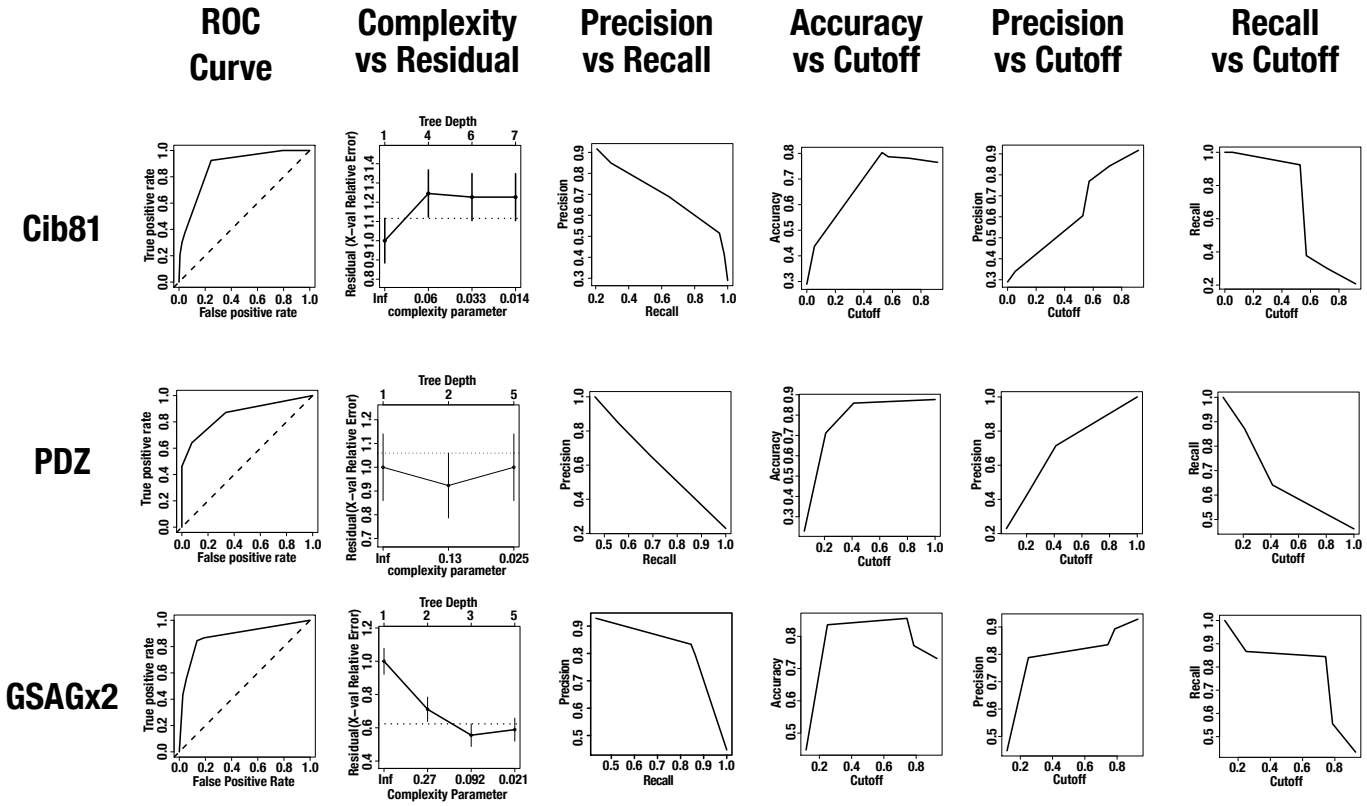
Supplemental Figure 3: Pearson correlation of biological replicates. Hierarchical clustering of all individual permissibility datasets based on domain structure and length. Biological replicates show a high degree of reproducibility. As expected, datasets for structured domains (Cib81 and PDZ) cluster discretely with themselves than with each other whereas flexible linkers cluster indiscriminately between each other.

Supplemental Figure 4



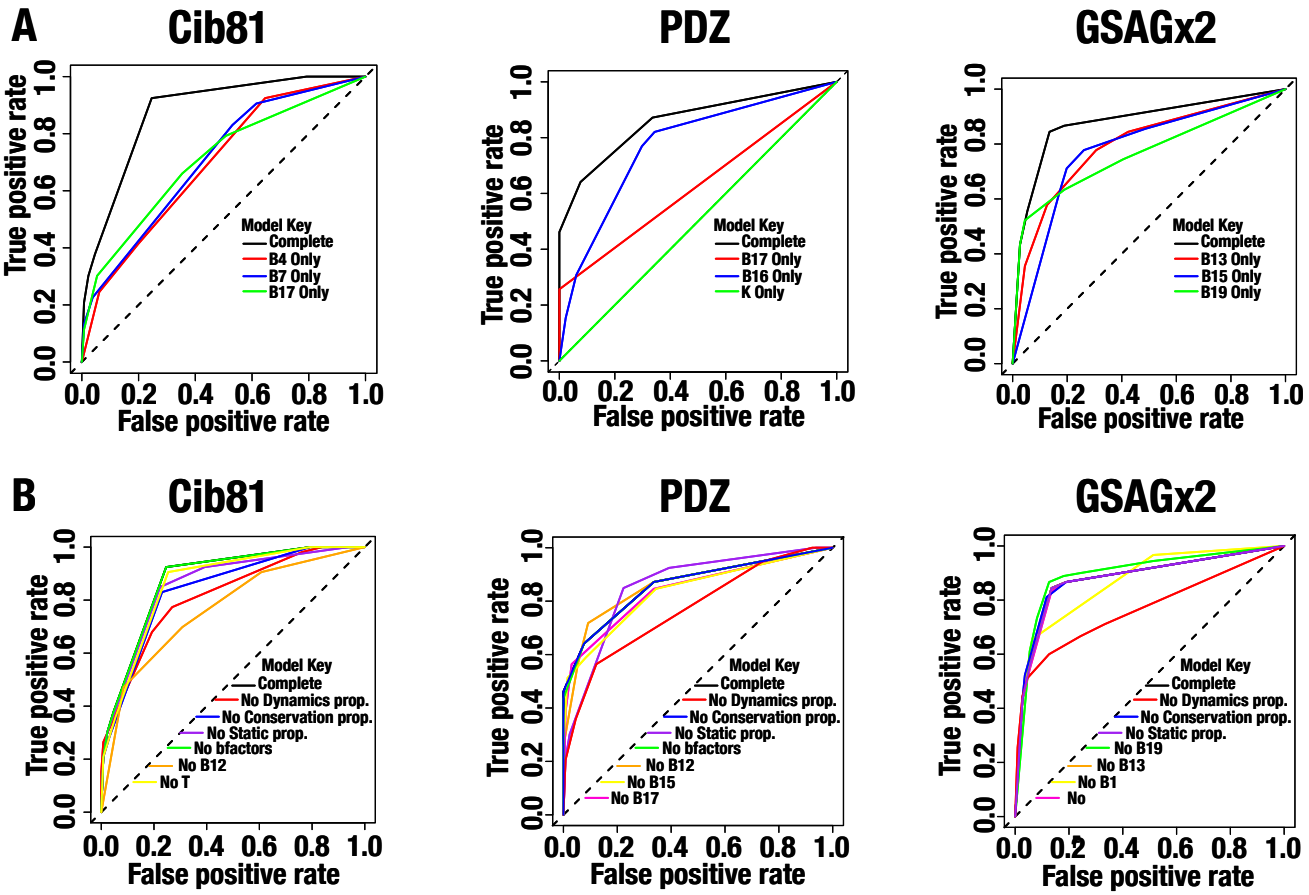
Supplemental Figure 4: Decision Trees. Decision trees that were trained on calculated conservation, static structural and dynamic protein properties to predict binarized Cib81, PDZ and $GSAG_{x2}$ permissibility. Models overwhelmingly use dynamic properties (normal modes) over all other categories of computed properties. All trees were restricted to a maximum depth of four and cross-validated ten times to minimize overfitting and enable comparison between models. Decision tree leaves can be read as: the topmost number refers to the leaf class (0- not permissible and 1-permissible), percentage of permissible samples within the leaf that fall within the class, and percentage of all data that are within the leaf. Color intensity refers to how pure the leaf is not permissive (blue) or permissive (green).

Supplemental Figure 5



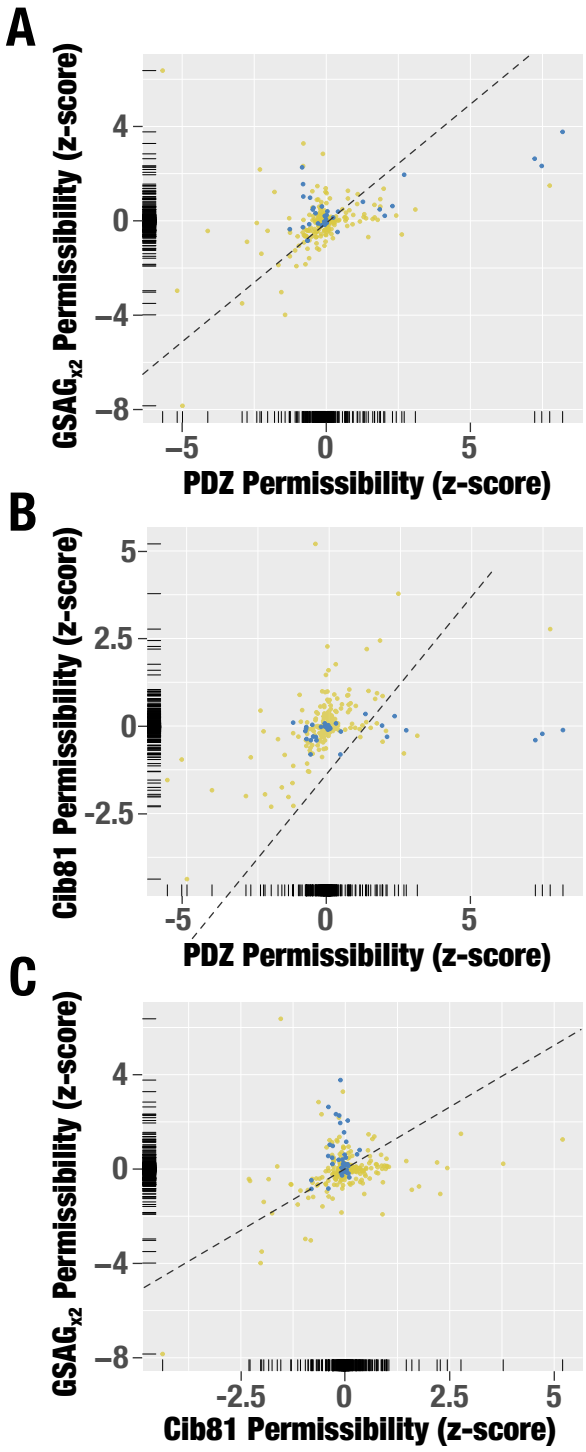
Supplemental Figure 5: Decision Tree Model Performance. Cib81, PDZ, and $GSAG_{x2}$ decision tree performance using different criteria. All models based on all parameters perform far better than random.

Supplemental Figure 6



Supplemental Figure 6: Decision trees trained on limited numbers of properties. (A) Decision trees were trained using the top three Spearman co-efficient correlated properties individually using the same model parameters as before (a maximum depth of four and cross-validated ten times). Models' performance was compared using receiver operator characteristic (ROC) curves. All models perform worse than models able to train on single properties showing that multiple interacting properties determine permissibility. Interestingly, the predicted effect of any amino acid mutated to lysine was highly linearly correlated with PDZ permissibility, however, using it alone was not able to predict PDZ permissibility at all whereas normal modes were able to predict PDZ permissibility individually. **(B)** Decision trees were trained with the decision tree determined predictive properties withheld individually and withholding each entire class of properties (conservation-, static structural- and dynamics based properties) using the same model parameters as before (a maximum depth of four and cross-validated ten times). Models' performance was compared using ROC curves. In every case, removing all computed conservation and static structural properties had little effect on model performance, however, removing dynamics based properties was detrimental to permissibility predictions. Apart from two examples (Cib81~B12 and $GSAG_{x2}$~B1) removing individual properties were not substantially impactful on model performance meaning there is redundancy in computed properties. Overall, all these decision trees reinforce that (1) dynamics are the most informative and closest correlated protein properties to permissibility and (2) permissibility is non-linearly made up of multiple interacting protein properties.

Supplemental Figure 7



Supplemental Figure 7: Differential Permissibility is common in functionally important sites. Z-scored permissibility comparisons between PDZ and GSAG$_{x2}$ linkers (**A**), PDZ and Cib81 (**B**), and Cib81 and GSAG$_{x2}$ linkers (**C**). Each dot represents a site in Kir2.1, color-coded by function (yellow, not involved in allosteric regulation; blue, involved in allosteric regulation).