

## Molecular Modeling and In Vitro Functional Analysis of the RGS12 PDZ Domain Variant Associated with High-Penetrance Familial Bipolar Disorder

Percy S. Agogo-Mawuli <sup>1</sup>, Joseph Mendez <sup>1</sup>, Emily A. Oestreich <sup>2</sup>,  
Dustin E. Bosch <sup>3</sup> and David P. Siderovski <sup>1,\*</sup>

\* Correspondence:

David P. Siderovski

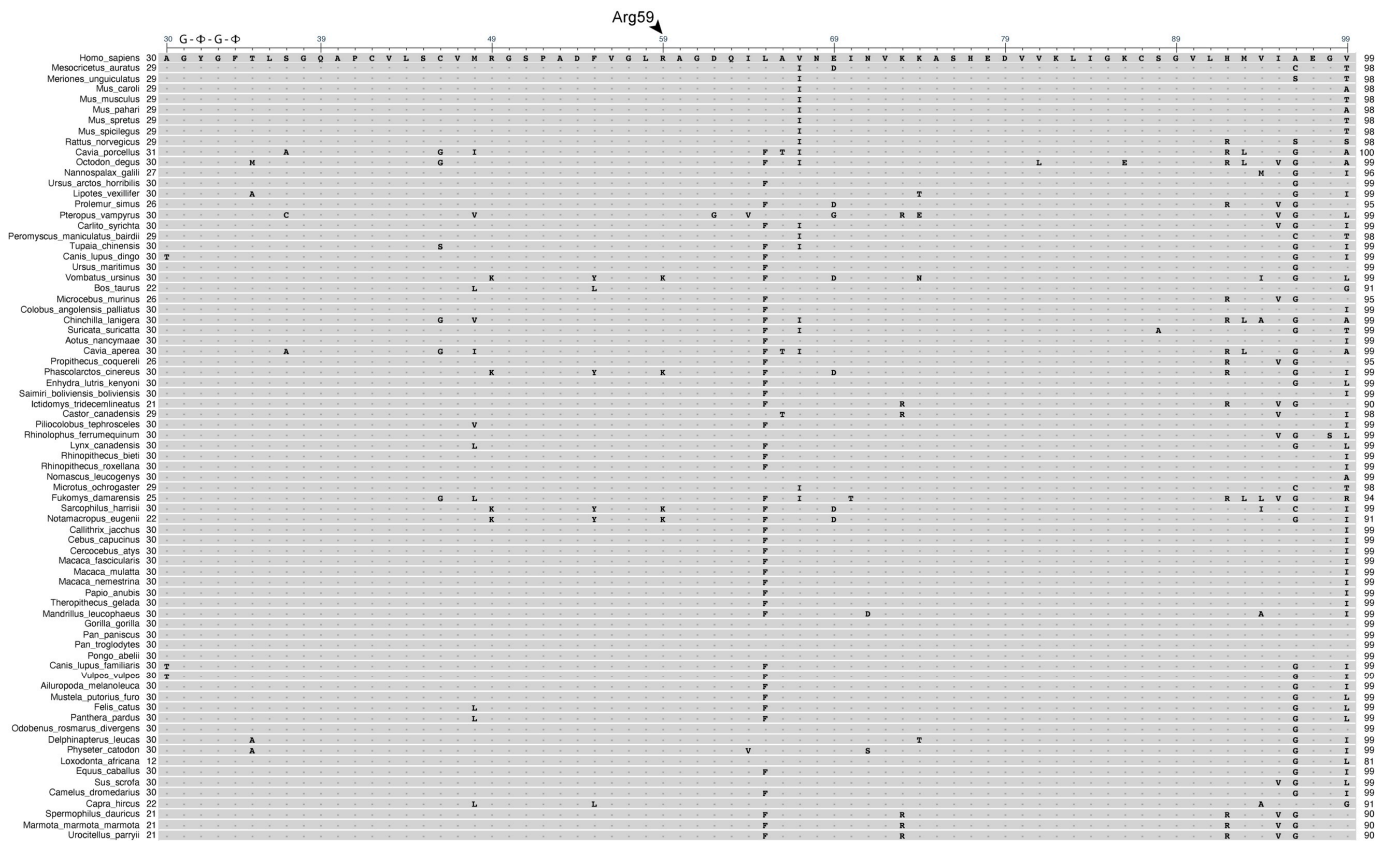
david.siderovski@unthsc.edu

<sup>1</sup> Department of Pharmacology & Neuroscience, University of North Texas Health Science Center,  
Fort Worth, TX 76107, USA

<sup>2</sup> Department of Biomedical Sciences, Pacific Northwest University of Health Sciences, Yakima,  
WA 98901, USA

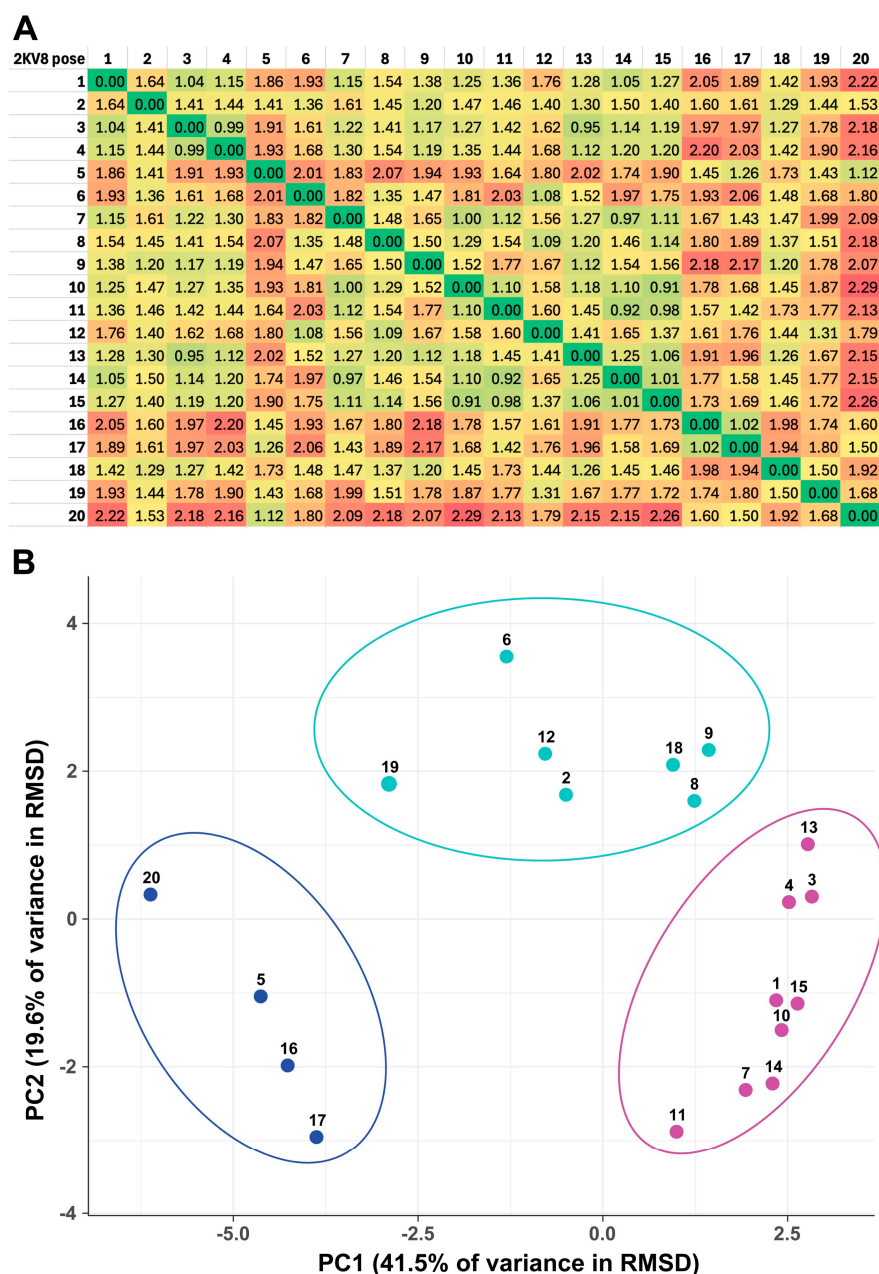
<sup>3</sup> Department of Pathology, Carver College of Medicine, University of Iowa, Iowa City,  
IA 52242, USA

Figure S1



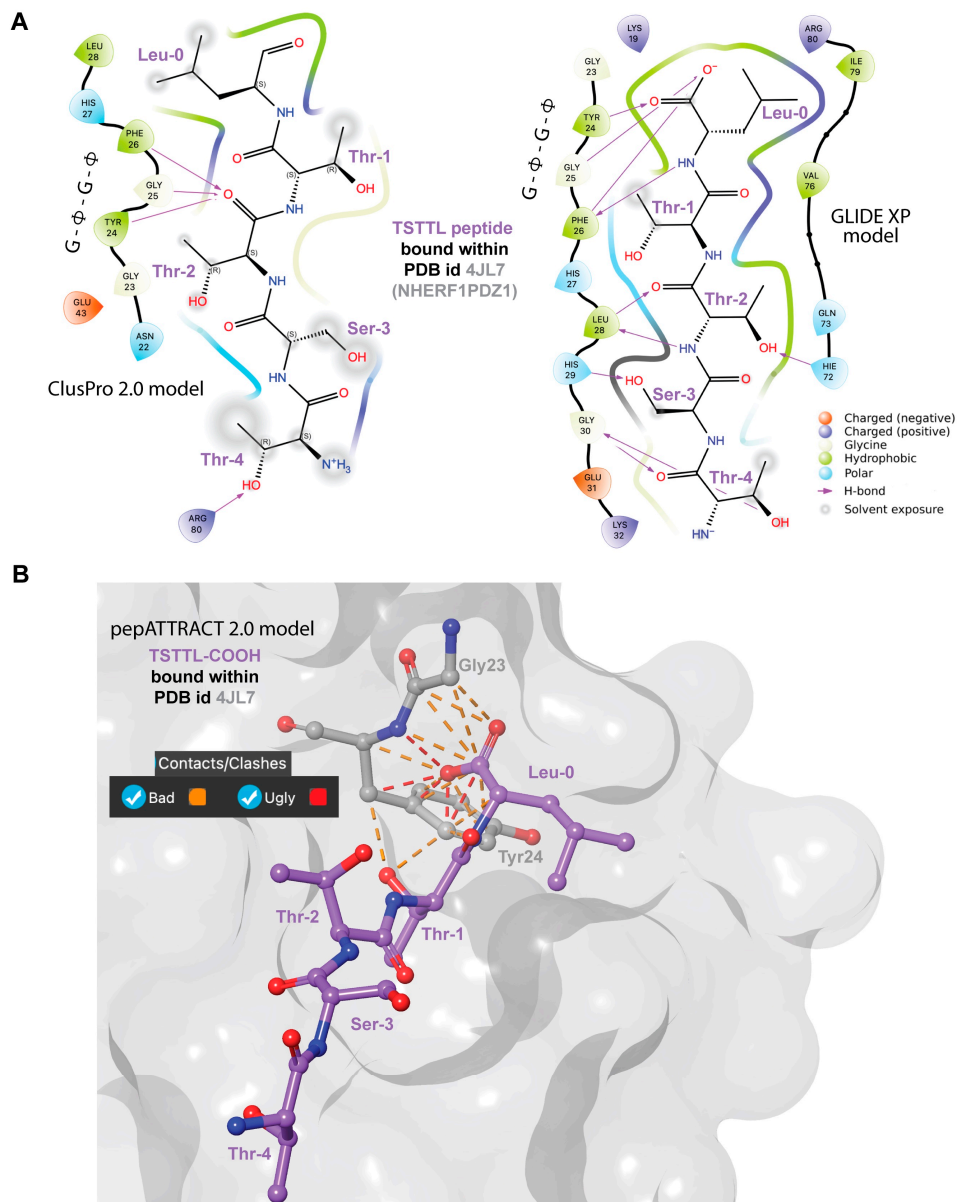
**Fig. S1.** Multiple sequence alignment (MSA) of PDZ domain sequences from 75 mammalian RGS12 orthologues. The PDZ domain sequence of human RGS12 (amino acids 30 to 99 of UniProt O14924) is displayed on the top row of the MSA. Subsequent rows contain the remaining 74 mammalian RGS12 orthologues as indicated by their species names. Identical amino acid residues within the non-human RGS12 orthologues are indicated by gray dots; only variations in their PDZ domain sequences are provided within the MSA. The G-Φ-G-Φ motif present in the human RGS12 PDZ domain sequence (Gly31 to Phe34) is completely conserved amongst all mammalian RGS12 orthologues. Similarly, the arginine residue (Arg59) that corresponds to the BD-associated R59Q variation in human RGS12 is highly conserved amongst most mammalian orthologues; however, four mammalian species diverge: a clade of Australian marsupials (*Vombatus ursinus*, *Phascogale carolinensis*, *Sarcophilus harrisii*, and *Notamacropus eugenii*) have a different basic amino acid (K, lysine) present at the position corresponding to Arg59 of human RGS12.

Figure S2



**Fig. S2. (A)** Heatmap displaying pairwise backbone root mean square deviation (RMSD) values between each of the 20 NMR-derived structural models of the unliganded RGS12 PDZ domain (PDB id: 2KV8). The RMSD values, measured in angstroms ( $\text{\AA}$ ), were calculated to assess structural variability among the models. Each cell in the heatmap represents the RMSD between two specific models, color-coded from green (low RMSD, indicating greater structural similarity) to red (high RMSD, indicating greater structural divergence). Visual Molecular Dynamics (VMD) software (<https://www.ks.uiuc.edu/Research/vmd/>) was used for RMSD calculations, with the diagonal showing each model's self-RMSD (0.00  $\text{\AA}$ ; dark green). **(B)** Principal Component Analysis (PCA) plot illustrating the clustering of the 20 RGS12 PDZ domain structures based on their pairwise RMSD values. Unit variance scaling was applied to normalize the data, ensuring equal weighting of each variable. The analysis was conducted using singular value decomposition (SVD) to handle any missing data. The x-axis (PC1) explains 41.5% of the total variance, while the y-axis (PC2) accounts for an additional 19.6% of the variance. The 20 models are represented as individual points, with clusters of conformations visually differentiated by distinct colors (royal blue, teal, and magenta), reflecting structural similarity. ClustVis (<https://biit.cs.ut.ee/clustvis/>) was used for PCA and plotting.

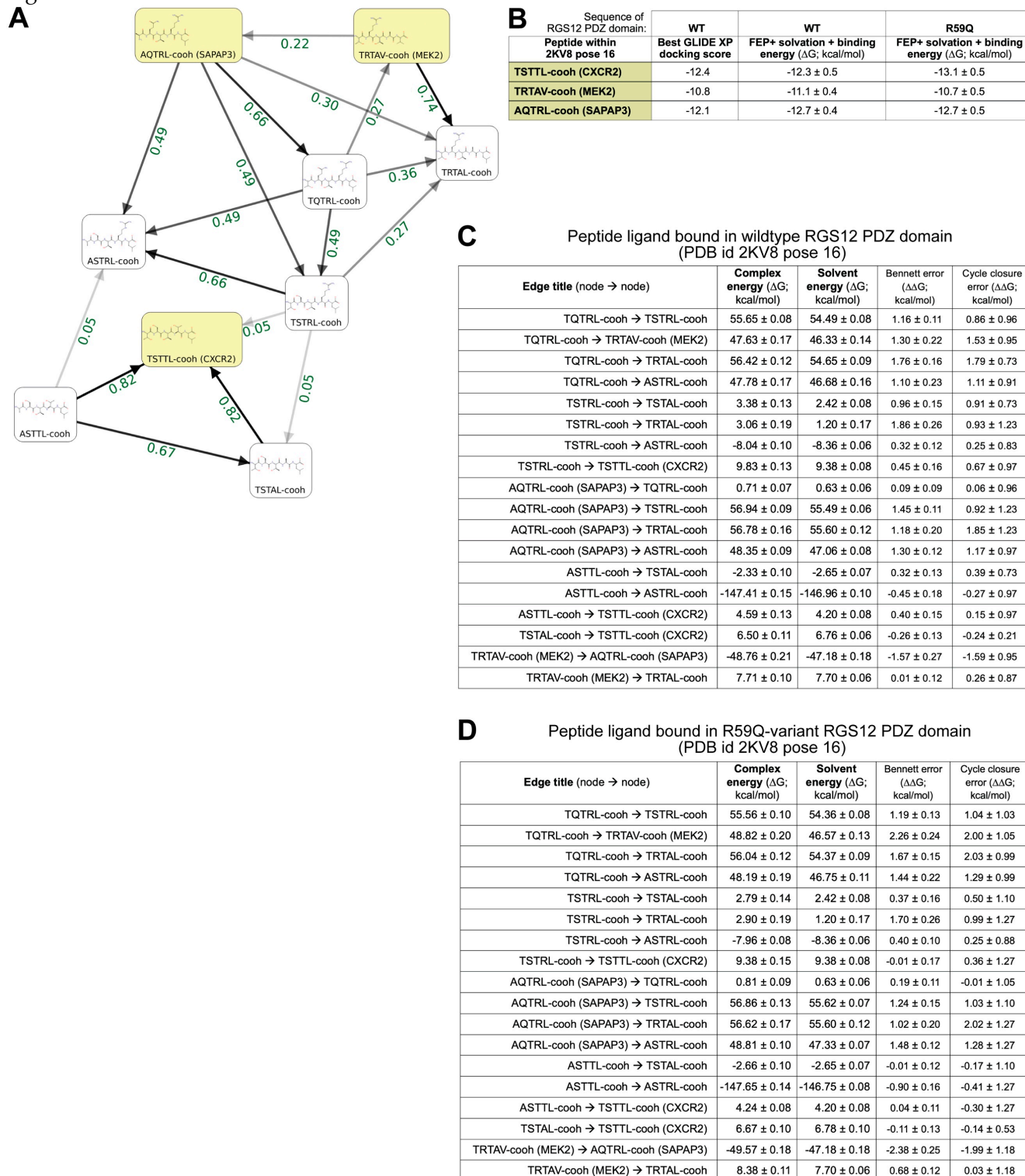
Figure S3



**Fig. S3.** Comparison of three protein/peptide docking algorithms -- ClusPro 2.0, GLIDE XP, and pepATTRACT 2.0 -- demonstrating differing predictions for the known NHERF1 PDZ1 domain interaction with the TSTTL-COOH terminal peptide from CXCR2 (PDB id 4JL7). **(A)** Schematic representations of predictions from ClusPro 2.0 (*left*) and GLIDE XP (*right*). In the ClusPro model, the TSTTL-COOH peptide adopts an orientation violating known PDZ domain/peptide interaction principles, particularly the improper alignment of the G-Φ-G-Φ motif with the peptide bond between Thr-2 and Thr-1. By contrast, the GLIDE XP model adheres to established PDZ domain/C-terminal peptide interaction principles, particularly the G-Φ-G-Φ motif's correct engagement with the peptide C-terminus and Leu-0. **(B)** Structural model of the TSTTL-COOH peptide bound to the NHERF1 PDZ1 domain (gray surface and residues Gly-23 and Tyr-24 of the G-Φ-G-Φ motif) as predicted by pepATTRACT 2.0. The molecular visualization highlights many steric clashes around the Leu-0 and Thr-1 residues of the bound peptide, as identified by Schrödinger's Maestro that quantitatively categorizes "bad" contacts (orange) and "ugly" contacts (red) based on the degree of atomic overlap between the peptide and the binding pocket. "Bad" contacts indicate mild steric clashes with less severe atomic overlaps, while "ugly" contacts represent more significant steric hindrances, where the overlap is too large for favorable interaction or binding. The GLIDE XP model, in contrast, shows no such clashes and aligns well with the canonical interaction framework for PDZ domains engaging C-tail peptides.

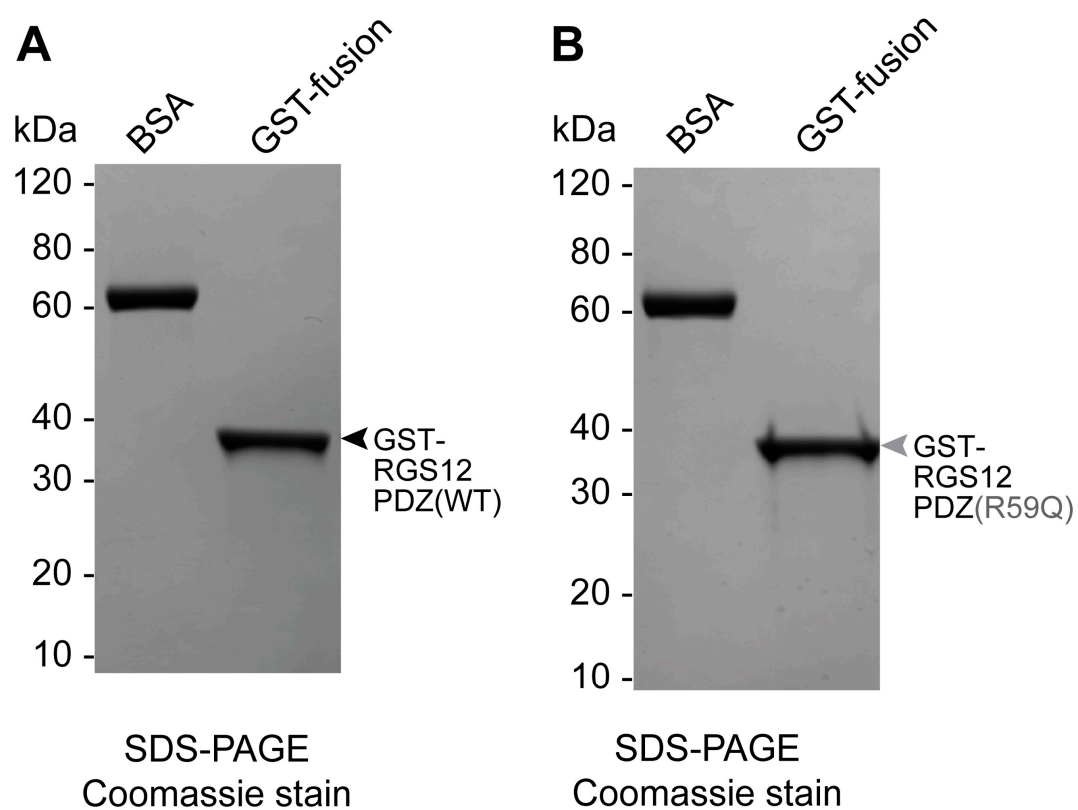


Figure S4



**Fig. S4.** FEP+ MD simulations of transit between nodes of individual pentamers bound to RGS12 PDZ domain. **(A)** The 9-node/18-edge permutation map networking together the three candidate C-tail ligands (highlighted in yellow, with name of originating protein in parentheses) and six different single amino-acid transformations thereof. Similarity index between two nodes is highlighted on each edge in green. Minimal edges per node (3) and minimal similarities ( $>0.04$ ) were maintained throughout the network prior to launching FEP+ MD simulations. **(B)** Summary of results obtained for wildtype (WT) and R59Q-variant versions of the RGS12 PDZ domain model (pose 16 of PDB id 2KV8), as compared to GLIDE XP docking scores for the same three C-tail ligands docking into pose 16 of PDB id 2KV8 (Figure 4). **(C, D)** Energy changes (and error estimations) for each node-to-node transit as reported by the FEP+ MD simulations.

Figure S5



**Fig. S5.** Regulator of G protein Signaling-12 (RGS12) PDZ domain proteins purified as glutathione-S-transferase (GST) fusions from *Escherichia coli*. Two microgram aliquots of wildtype (panel **A**) or R59Q variant (panel **B**) PDZ domain fusion protein, post purification by high-pressure/high-performance liquid chromatography (HPLC), were separated by sodium dodecyl sulphate (SDS)-polyacrylamide gel electrophoresis, along with molecular weight markers and a control aliquot of two micrograms of bovine serum albumin, and then visualized by Coomassie blue staining.