Activation of Polycystin-1 Signaling by Binding of Stalk-derived Peptide Agonists

Shristi Pawnikar[1], Brenda S. Magenheimer[3,4], Keya Joshi[6], Ericka Nevarez Munoz[3], Allan Haldane[5], Robin L. Maser[2,3,4,*] and Yinglong Miao[6,*]

[1]Center for Computational Biology and Department of Molecular Biosciences, University of Kansas, Lawrence, KS 66047; [2]Departments of Biochemistry and Molecular Biology, [3]Clinical Laboratory Sciences, [4]The Jared Grantham Kidney Institute, University of Kansas Medical Center, Kansas City, KS 66160, [5]Dept of Physics, and Center for Biophysics and Computational Biology, Temple University, Philadelphia, PA 19122, and [6]Department of Pharmacology and Computational Medicine Program, University of North Carolina – Chapel Hill, Chapel Hill, NC 27599.

* Corresponding emails: Yinglong_Miao@med.unc.edu; rmaser@kumc.edu

ABSTRACT

Polycystin-1 (PC1) is the membrane protein product of the PKD1 gene whose mutation is responsible for 85% of the cases of autosomal dominant polycystic kidney disease (ADPKD). ADPKD is primarily characterized by the formation of renal cysts and potential kidney failure. PC1 is an atypical G protein-coupled receptor (GPCR) consisting of 11 transmembrane helices and an autocatalytic GAIN domain that cleaves PC1 into extracellular N-terminal (NTF) and membrane-embedded C-terminal (CTF) fragments. Recently, signaling activation of the PC1 CTF was shown to be regulated by a stalk tethered agonist (TA), a distinct mechanism observed in the adhesion GPCR family. A novel allosteric activation pathway was elucidated for the PC1 CTF through a combination of Gaussian accelerated molecular dynamics (GaMD), mutagenesis and cellular signaling experiments. Here, we show that synthetic, soluble peptides with 7 to 21 residues derived from the stalk TA, in particular, peptides including the first 9 residues (p9), 17 residues (p17) and 21 residues (p21) exhibited the ability to re-activate signaling by a stalkless PC1 CTF mutant in cellular assays. To reveal molecular mechanisms of stalk peptide-mediated signaling activation, we have applied a novel Peptide GaMD (Pep-GaMD) algorithm to elucidate binding conformations of selected stalk peptide agonists p9, p17 and p21 to the stalkless PC1 CTF. The simulations revealed multiple specific binding regions of the stalk peptide agonists to the PC1 protein including an “intermediate” bound yet inactive state. Our Pep-GaMD simulation findings were consistent with the cellular assay experimental data. Binding of peptide agonists to the TOP domain of PC1 induced close TOP-putative pore loop interactions, a characteristic feature of the PC1 CTF signaling activation mechanism. Using sequence covariation analysis of PC1 homologs, we further showed that the peptide binding regions were consistent with covarying residue pairs identified between the TOP domain and the stalk TA. Therefore, structural dynamic insights into

the mechanisms of PC1 activation by stalk-derived peptide agonists have enabled an in-depth understanding of PC1 signaling. They will form a foundation for development of PC1 as a therapeutic target for the treatment of ADPKD.

Keywords: Polycystin-1, Peptide Gaussian accelerated Molecular Dynamics (Pep-GaMD), Peptide Agonists, Autosomal dominant polycystic kidney disease (ADPKD), Cellular Signaling, Sequence Covariation.

INTRODUCTION

Polycystin-1 (PC1) is the protein product of the PKD1 gene that is mutated in the majority of cases (~85%) of autosomal dominant polycystic kidney disease (ADPKD)[1]. ADPKD is a potentially lethal disease, affecting >0.6 million individuals in the US. It causes renal cyst formation that could consequently lead to kidney failure. Currently, the only approved treatment for ADPKD is Jynarque™, a small-molecule antagonist of the arginine vasopressin receptor 2, V2R, whose signaling, and production of cAMP has been shown to be increased in PKD. This drug targets one of the aberrant pathways downstream from the PKD gene mutation but is inadequate due to its limitations in only slowing disease progression and causing adverse side effects[2]. ADPKD severity is dependent on the functional level of PC1, and as such, therapies designed to increase the level of PC1 protein, and its functionality are currently being pursued[3-7]. Approximately one-third of PKD1 mutations are non-truncating and could encode partially functional PC1 protein[3, 8-10]. As such, therapeutic treatments that directly target and activate PC1 may represent a promising approach for the treatment of ADPKD. However, this approach remains difficult due to incomplete knowledge of the proximal-most functions of PC1.

PC1 shares characteristics with the Adhesion class of GPCRs (ADGRs), including a conserved GPCR autoproteolysis inducing (GAIN) domain that directs autocatalytic cleavage at an embedded GPCR proteolysis site (GPS) motif[11]. Intramolecular cleavage at the GPS motif generates two non-covalently attached fragments - the extracellular N-terminal fragment (NTF) and the membrane-embedded C-terminal fragment (CTF)[12, 13]. Similar to ADGRs, the PC1 NTF consists of multiple adhesive domains that promote interactions between cells and with the extracellular matrix[14-18], while the PC1 CTF is composed of 11 transmembrane (TM) helices and a short C-terminal tail (C-tail)[19] that has been shown to interact with G proteins for signaling

activation or regulation[20, 21] and has thus led to description of PC1 as an atypical GPCR. Previous studies demonstrated the critical importance of cleavage at the PC1 GPS site to prevent renal cystogenesis in mouse models[4, 22]. For the ADGRs, a tethered agonist (TA) model has been proposed for activation of G protein signaling. After dissociation of the NTF, the N-terminal stalk of the ADGR CTF interacts with its membrane-embedded TM domains to induce conformational rearrangements that mediate activation of G protein signaling[23-26]. Exogenous synthetic peptides consisting of various lengths of the N-terminal sequence of the stalk have been shown to function as soluble agonists in activation of signaling by full-length and CTF mutants for numerous ADGRs[20, 27].

In previous studies of the PC1 CTF, we revealed a stalk TA-dependent molecular mechanism underlying CTF-mediated activation of an NFAT promoter luciferase reporter through complementary *in vitro* cell signaling experiments and all-atom Gaussian accelerated Molecular Dynamics (GaMD) simulations[28]. GaMD is an unconstrained enhanced sampling method that works by adding a harmonic boost potential to reduce large biomolecular energy barriers[29] and has been used successfully to capture multiple complex biological processes[30-37,38-44,30,32] including GPCR activation[31]. Expression constructs encoding a stalkless PC1 CTF (a nonbiological mutant with deletion of the first 21 N-terminal residues of CTF) and three ADPKD-associated missense mutants within the stalk region (G3052R, R3063C, and R3063P) were shown to be defective in reporter activation as compared to wildtype PC1 CTF. GaMD simulations revealed a novel allosteric transduction pathway for activation of PC1 CTF signaling that involves initiation by the Stalk interacting with a large extracellular loop between TM segments S1/TM6 and S2/TM7, called the TOP domain, followed by close interactions between the TOP and a putative pore loop (PL) domain between the final 2 TM domains. GaMD simulations of the wildtype PC1 CTF also

identified a “Closed/Active” low-energy state related to the large number of Stalk-TOP contacts and the R3848-E4078 ionic interaction between the TOP and PL domains that was not present in the stalkless CTF[28].

Here, we have utilized *in vitro* cell signaling assays to identify peptide agonists targeting PC1 in combination with *in silico* studies to investigate their binding mechanisms for activation of PC1 signaling. Synthetic peptides of 7-21 residues in length derived from the N-terminus of the PC1 CTF stalk sequence were tested for their ability to re-activate signaling of the stalkless CTF expression construct. Peptide docking and simulations with the recently developed Peptide GaMD (Pep-GaMD), which is able to characterize peptide-protein binding processes more efficiently[45], were combined for selected peptide agonists p9, p17 and p21 to gain insight into their binding mechanism to the stalkless PC1 CTF. Pep-GaMD was able to successfully refine the docking conformations of the peptides bound to the extracellular TOP domain of PC1. In further Pep-GaMD simulations, the key salt bridge interaction between R3848 and E4078 from the TOP domain and PL, respectively, was observed upon binding of the peptides to stalkless PC1 CTF. Using Potts covariation analysis, in which a protein fitness model is inferred based on observed mutational covariation patterns in multiple sequence alignments (MSAs) of homologous proteins[46], we identified residues in the PC1 stalk with direct mutational covariation with residues in the TOP domain, which were strikingly consistent with the binding interfaces identified in docking and simulation studies. Overall, these analyses yielded mechanistic insights underlying the stalk peptide agonist mediated signal re-activation of stalkless PC1 CTF. Such insights provide significant contributions toward the future design and development of peptide modulators targeting PC1 for an effective ADPKD therapeutic treatment.

RESULTS

Synthetic, stalk-derived peptides re-activate NFAT reporter by CTF$^{\Delta st}$ *in trans*

Our previous study utilized expression constructs of human PC1 CTF. However, in order to prepare for eventual *in vivo* experiments in mouse models, we generated expression constructs of mouse (m) PC1 consisting of the signal peptide sequence of the T cell surface glycoprotein CD5 (MPMGSLQPLATLYLLGMLVASVLG) fused in frame with the stalk sequence of wildtype mCTF beginning with residue T3041, or with a 'stalkless' CTF lacking the first 21 residues of the stalk (mCTF$^{\Delta st}$) beginning with residue S3062 (**Fig. 1A**). The CD5 signal peptide coding sequence was added to the wildtype mCTF and stalkless mCTF$^{\Delta st}$ expression constructs in order to ensure their translation at the endoplasmic reticulum for plasma membrane localization.

Transient transfection of HEK293T cells with either empty expression vector (ev), CTF or CTF$^{\Delta st}$ showed that the CTF$^{\Delta st}$ mutant exhibited a dramatic loss of NFAT reporter activation that was essentially reduced to ev control levels (**Fig. 1B**). Both total (**Fig. 1C-D**) and cell surface (**Fig. S1A-B**) expression levels of CTF$^{\Delta st}$ were comparable to CTF, which suggests that neither protein stability nor membrane trafficking was responsible for the inability of CTF$^{\Delta st}$ to activate the NFAT reporter. These results are consistent with those obtained using expression constructs of human PC1 that demonstrated the stalk region of PC1 CTF acts as a tethered peptide agonist[28].

To further investigate the agonistic property of the CTF stalk, we synthesized peptides (p) consisting of the N-terminal 7, 9, 11, 13, 15, 17, 19 or 21 residues from the stalk sequence of mPC1. All peptides were appended with a C-terminal, 7-residue hydrophilic sequence (GGKKKKK) to increase solubility. HEK293T cells were transiently transfected with empty expression vector or mCTF$^{\Delta st}$ plasmids along with the NFAT luciferase reporter and then treated with stalk peptides p7 through p21 or with addition of culture medium only ('no peptide' control).

The NFAT reporter was significantly activated in $CTF^{\Delta st}$-transfected cells by treatment with p7, p9 or p17 as compared to their corresponding ev + peptide treatment controls. These stalk peptides also significantly increased reporter activity in comparison to the $CTF^{\Delta st}$ with no peptide treatment control (**Fig. 1E**). Treatment of $CTF^{\Delta st}$-transfected cells with p19 or p21 also significantly increased reporter activation in comparison to the $CTF^{\Delta st}$ + no peptide control; however, reporter activation occurred in both ev- and $CTF^{\Delta st}$-transfected cells treated with either p19 or p21, suggesting that p19- and p21-mediated activation was not dependent on exogenous expression of mouse $CTF^{\Delta st}$ and could be activating the endogenous human PC1 protein. Treatment of ev- and $CTF^{\Delta st}$-transfected cells with peptide consisting of the hydrophilic sequence alone (i.e., solubility peptide) showed that the solubility tag was not responsible for the rescue of $CTF^{\Delta st}$-mediated reporter activation by stalk peptides such as p17 (**Fig. S2**). Altogether, these results were consistent with soluble stalk-derived peptides acting as PC1 CTF agonists *in trans*, and provided additional support for the PC1 CTF stalk region harboring tethered agonist (TA) activity[28]. We hypothesized that the soluble, activating peptides bind to the TOP domain of PC1 in a manner mimicking the tethered stalk in order to reactivate signaling of the stalkless $CTF^{\Delta st}$ mutant. From among the active stalk-derived peptides, we selected p9, p17 and p21 that exhibited the highest agonist activity in $CTF^{\Delta st}$ reporter activation (**Fig. 1E**) for *in silico* simulation studies.

Docking and Pep-GaMD simulations of peptide agonist binding to stalkless PC1 CTF

We chose to use the stalkless CTF (ΔStalk CTF) as representing the least complex system in which the binding of exogenous peptides could be studied. ΔStalk CTF is not a biological form or a mutant protein of PC1 observed in ADPKD. However, in our previous study, it mimicked the ADPKD-associated stalk mutants by being defective in cell signaling assays and rarely formed the

R3848-E4078 salt bridge that was frequently seen in GaMD simulations with wild type CTF[28]. Therefore ΔStalk CTF served as the negative control for the following studies.

The cryo-EM structure of human PC1-PC2 complex (PDB: 6A70)[47] was used to build the computational model for ΔStalk PC1 CTF after deleting the first 21 residues (3049-3069) from the CTF[28]. We successfully docked the p9, p17 and p21 stalk peptides to the ΔStalk CTF model with HPEPDOCK[48] (**See SI**). The peptides all bound to the TOP domain and the interface between the TOP domain and extracellular loop 1 (ECL1) of CTF (**Fig. S3A-B**). In particular, peptide p21 occupied a closely similar binding region as the stalk in wildtype CTF as observed in the previous study[28]. We then performed five independent 500 ns Pep-GaMD simulations on each of the three stalk peptide agonists p9, p17 and p21 bound to ΔStalk CTF to refine their HPEPDOCK docking conformations (**See SI**).

With the Pep-GaMD simulation frames, we performed structural clustering of each peptide using the hierarchical agglomerative algorithm in CPPTRAJ[49]. The top-ranked conformations of each peptide bound to ΔStalk CTF were compared to their initial docking conformations. Next, we calculated 2D free energy profiles of the peptides-bound ΔStalk CTF by reweighting the Pep-GaMD simulations. The R3848-E4078 residue distance and the number of contacts between the peptides (p9, p17 and p21) and TOP domains were selected as the reaction coordinates. The number of contacts was calculated between any atom pairs within 4 Å distance of the peptide and extracellular domains of PC1 protein. In the subsequent analyses, stalk and peptide residues are numbered relative to the N terminus of the stalk as starting from 1, while residues of the Δstalk CTF are numbered according to the human PC1 protein sequence.

Active conformation of peptide p9-bound PC1 CTF

From the free energy profile of the p9-bound ΔStalk CTF, we identified "Unbound" and "Bound" low-energy states (**Fig. 2A**). In the docking conformation, peptide p9 bound to the interface between the TOP and ECL1 of ΔStalk CTF (**Fig. S3A-B**). In Pep-GaMD simulations, the p9 peptide dissociated from the TOP-ECL1 binding pocket and rebound to the TOP domain in a slightly different region (**Fig. 2B-C**). The p9 peptide sequence is mostly composed of hydrophobic residues. Polar interactions between the main chain atoms of peptide-protein residues were observed in the top-ranked representative conformation of the p9-bound ΔStalk CTF. Protein residues R3892 and H3864 formed hydrogen bonds with p9 residues A2 and A5, respectively (**Fig. 2D**). These interactions were also highlighted in the protein contact map between peptide p9 and the extracellular domains of CTF in the representative "Bound" state **(Fig. S4)**. The distance between the TOP domain residue R3848 and PL residue E4078 was 3.9 Å (**Fig. 2E**), suggesting that the top-ranked representative conformation of the p9-bound ΔStalk CTF was in the "Closed/Active" low-energy state. In addition, the 2D free energy profile of each individual simulation was calculated and the "Bound" low-energy state was consistently identified in the 2D free energy profiles of peptide p9 **(Fig. S5)**.

Active and Intermediate conformational states of peptide p17-bound PC1 CTF

The free energy profile of the p17-bound ΔStalk system allowed us to identify three low-energy states - "Unbound", "Intermediate", and "Bound" (**Fig. 3A**). In the docking conformation, peptide p17 bound to the interface between the TOP and ECL1 of ΔStalk CTF (**Fig. S3A-B**). In the Pep-GaMD refined "Bound" state, a folded antiparallel ß-strand conformation was observed for the peptide p17 at the interface of ECL1 and the TOP domain (**Fig. 3B-C**). Peptide residues T1, F3, A5, F8, F16 and V17 formed hydrophobic interactions with the protein residues H3311, R3314

and Y3307 from ECL1, and E3708, S3711, Q3707, A3704, R3700 and L3701 from the TOP domain (**Fig. 3D**). These interactions were also highlighted in the protein contact map between peptide p17 and the extracellular domains of CTF in the representative "Bound" state **(Fig. S4)**. The distance between the TOP domain residue R3848 and PL residue E4078 was 4.1 Å (**Fig. 3E**), suggesting that the top-ranked representative conformation of the p17 bound ΔStalk CTF was in the "Closed/Active" low-energy state.

In the "Intermediate" state, p17 with a short helical turn was also observed to bind the TOP domain of ΔStalk CTF (**Fig. 3B-C**). Hydrophobic residue interactions were also formed between the peptide and protein. In particular, peptide residues T1, F3, P10, P11, H13, R15, F16 and V17 formed hydrophobic interactions with the protein residues P3859, A3704, S3741, Q3739, Y3734, P3733, H3729, W3726, R3712 and R3856 from the TOP domain (**Fig. 3F**). The distance between the TOP domain residue R3848 and PL residue E4078 was 14.6 Å (**Fig. 3G**), suggesting that this representative conformation (ranked the second among the Pep-GaMD structural clusters) of the p17-bound ΔStalk CTF was in the "Intermediate" low-energy state. In addition, the 2D free energy profile of each individual simulation was calculated. Pep-GaMD simulations were able to refine the peptide conformation from the "Unbound" to "Intermediate" and "Bound" states in Sim1 and Sim5, while the peptide reached only the "Intermediate" state in the other three simulations **(Fig. S6)**. The free energy values of 2D PMF minima shown in **Fig. 3A** could differ from those in the 1D PMF minima of peptide structural clusters, especially with the usage of distinct reaction coordinates.

Active conformational state of peptide p21-bound PC1 CTF

Finally, the free energy profile of the p21-bound ΔStalk CTF allowed us to identify only a broad low-energy well corresponding to the "Bound" state (**Fig. 4A**). The docking conformation of p21-

bound ΔStalk CTF was refined through Pep-GaMD simulations, where folding of the peptide was observed on the protein surface of the TOP domain (**Fig. 4B-C**). The p21 peptide occupied a similar binding region as the stalk in wildtype CTF as observed in the previous study[28]. Hydrophobic contacts were observed between peptide residues L7, F8, P10, S12, H13, V14, V17, P19, E20 and P21 and protein residues L3863, L3701, I3705, L3709, E3708, R3712, F3714, H3729, W3726, V3730, L3732, P3733, N3738, R3856 and S3741 (**Fig. 4D**). These interactions were also highlighted in the protein contact map between peptide p21 and the extracellular domains of CTF in the representative "Bound" state **(Fig. S4)**. The distance calculated from the top-ranked structural cluster of the system between the TOP domain residue R3848 and PL residue E4078 was 3.8 Å, corresponding to the "Closed/Active" low-energy state **(Fig. 4E)**. Furthermore, time courses of the radius of gyration (*Rg*) and root-mean square deviation (RMSD) of p21 relative to the starting HPEPDOCK conformation showed large conformational changes of the peptide during Pep-GaMD simulations (**Fig. S7)**. In addition, the 2D free energy profile of each individual simulation showed that Pep-GaMD was able to refine the peptide docking conformation to the "Bound" state in all the five simulations **(Fig. S8)**.

Peptide binding regions correlated with covarying residue pairs identified between the TOP domain and stalk TA

To provide an independent basis of evidence supporting the observation of "Bound" and "Intermediate" states of agonist peptide binding, we constructed a multiple sequence alignment (MSA) with an effective count of 1022 evolutionarily diverged PC1 homologs (illustrated in **Fig. S9)** from which we inferred a Potts statistical model (**Fig. 5**). Columns of the MSA with "direct" statistical interactions, as detected using the Potts inference method, reflect compensatory mutation pairs maintained through evolution supporting a conserved function. We limited our MSA to 394

residues on the extracellular side of PC1 because of the computational challenge of fitting the entire PC1 sequence (**Fig. 5B** and **Fig. S9**).

Fig. 5A shows the pairs of positions with strong Potts interaction scores where one position is either in the stalk, the nearby GAIN domain, or the TM1 helix. Some predicted interaction pairs recapitulated beta-sheet contacts within the GAIN domain observed in the homolog rat latrophilin-1[11] as well as predicted by Alphafold[50] (**Fig. S10**) or between the extracellular ends of the TM2/4 and TM5/6 alpha helices known from cryo-EM structures[47] or predicted by Alphafold, validating that our model detected biologically functional interactions.

We identified strong interactions between the stalk and other residues from the Potts model. They were not observed in the cryo-EM structure, in which the flexible stalk is missing. For interactions with the TOP domain, out of the 4875 possible pairs (25 stalk residues by 195 TOP domain residues in our Potts model), this analysis detected a stringent set of 6 strongly interacting pairs. Remarkably, multiple positions in this small set were among those relevant to the "Intermediate" binding conformation of p17 and "Bound" conformation of p21 as identified from the Pep-GaMD simulations. These were W3726 and S3741 in the TOP domain, both interacting with T1 of the stalk, and P3859 interacting with N25 at the end of the stalk **(Fig. 5A).** Additionally, we identified E3743 to be strongly interacting with F16 in the stalk, and it was also near the observed binding region in the TOP domain for the peptide p17 in the "Intermediate" state near S3741 (**Fig. 3F**). The remaining two strong interactions between the stalk and TOP domain involved Q3821 with T1 and L3893 with G4. L3893 is adjacent to R3892 that was identified to interact with the peptide p9 in the "Bound" state (**Fig. 2D**) and mutating it may affect its neighbor's positioning. Besides the interactions between the TOP domain and the stalk TA, we also found a set of interactions between the stalk and the extracellular ends of TM2-TM3 helices and TM4-

TM5 helices, in which stalk residues G4, P10, F16 and E20 interact with W3298, A3296, and S3579, respectively, as well as a strong interaction between V3077, three residues past the end of the stalk, and T3856 in the TOP domain. TOP domain residue T3856 was also identified as relevant to the binding region of peptide p17 in the “Intermediate” state (**Fig. 3F)** and peptide p21 in the “Bound” state (**Fig. 4D)**. These interactions could additionally play a role in stalk-TA activation or could be related to other functionality such as cleavage in the GPS motif.

To gain further insight and to validate that these detected "direct" interactions reflect biologically meaningful functional interactions and are not artifacts of the data, we examined the residue-specific covariation observed in the MSA (**Fig. 5C**), which measures the difference between the observed pairwise residue frequency and its null expectation under assumption of independent variation. Values greater than ~1% are commonly found to be indications of a statistically reliable mutational covariation (see **SI Computational Methods**), and many of the covarying pairs discovered between the stalk and TOP domain were significantly above this value. We validated that the covarying residue pairings were consistent with biophysical interaction. For example, for the position-pair 20-3579, here annotated such that the first index is the stalk residue numbered relative to the N terminus of the stalk and the second index is the TOP domain residue numbered according to the human PC1 protein sequence, there were excess residue-pair counts in the MSA consistent with opposite-charge or polar pairing such as K20-E3579, N20-Q3579, and others, and a dearth of repulsive like-charge pairs such as E20-E3579. Similarly, position-pair 1-3741 favored certain combinations of polar residues such as T1-S3741. Other position-pairs appeared consistent with hydrophobic packing interactions, such as F16-A3296, G4-W3298, and T1-W3726. A large residue F or W at position 3298 in the TOP domain was commonly paired

with a G at stalk position 4, while a smaller I or L residue at position 3298 was more commonly paired with T at stalk position 4.

Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA)[51] analysis was further performed to calculate the binding free energies of peptides p9, p17 and p21 to PC1 CTF and decompose the residue-wise energy contributions using the gmx_MMPBSA software[52]. The relative rank of the overall peptide binding free energies **(Table S1)** was consistent with the experimental signaling data, i.e., p21>p9>p17, for which p21 showed the largest binding free energy value of binding (-40.29±6.94 kcal/mol). Furthermore, we performed residue-wise energy decomposition analysis with MM/PBSA using gmx_MMPBSA software[52], which allowed us to identify key residues that contributed the most to the peptide binding energies. These included residues T1 and V9 in p9 **(Table S2)**, residues T1, R15 and V17 in p17 **(Table S3)**, and residues P10, P11, P19 and P21 in p21 and residue W3726 in the PC1 CTF **(Table S4)**. The energetic contributions of these residues apparently correlated to the sequence coevolution predicted from the Potts model.

DISCUSSION

In *in vitro*, cell-based signaling assays, PC1 CTF-mediated activation of the NFAT reporter is dependent on its N-terminal, extracellular stalk, as shown by the loss of reporter activity with the CTF stalk-deletion expression construct, $CTF^{\Delta st\ 28}$ (**Fig. 1B**) and by the ability of synthetic, stalk sequence-derived peptides to reactivate signaling by $CTF^{\Delta st}$ *in trans* (**Fig. 1E**). A series of synthetic peptides derived from the N-terminal sequence of the mouse PC1 CTF stalk were used to determine their agonistic activity in PC1 $CTF^{\Delta st}$-transfected cells. Notably, treatment with stalk

peptides p7, p9, p17, p19 and p21 resulted in significant NFAT reporter activity over $CTF^{\Delta st}$ control (no peptide treatment), wherein the effects of p7, p9 and p17 were specific to mouse $CTF^{\Delta st}$-expressing cells. These data are consistent with the stalk peptides acting as soluble TA peptide agonists for PC1 and provide further evidence for the activation of PC1 signaling via an ADGR-like TA mechanism.

In ADPKD, numerous missense mutations reported within the GPCR autoproteolysis-inducing (GAIN) domain that have been shown to prevent or inhibit cleavage at the GPCR-coupled proteolysis site (GPS)[11]. Loss of PC1 GPS cleavage, which is known to cause ADPKD, would likely sequester the stalk tethered agonist within the interior of the GAIN domain, which would presumably interfere with interactions between stalk tethered agonist residues and the remainder of the CTF and thus loss of the stalk-mediated signaling mechanism. Furthermore, there are 10 single nucleotide polymorphisms reported within the stalk sequence (ADPKD Variant Database; https://pkdb.mayo.edu/welcome), most of which were found to significantly reduce CTF-mediated activation of the NFAT reporter[53]. In particular, the ADPKD-associated G3052R stalk mutation that was previously analyzed along with the stalkless CTF by GaMD simulations[28] has the same reduction in activity as the stalkless CTF in the cellular signaling reporter assays and the same loss of active/closed conformation interactions in GaMD analyses. Therefore, the stalkless CTF was used in our docking and simulation studies as representative of a biologically relevant mutant form of PC1.

To reveal the molecular mechanisms of the soluble stalk-derived peptides, we chose to perform HPEPDOCK docking and novel Pep-GaMD simulations to sample the peptide interactions with the ΔStalk PC1 CTF. Pep-GaMD simulations were able to refine the docking conformations of peptide agonists bound to the ΔStalk PC1 CTF. It is important to note that the

free energy profiles calculated from GaMD simulations of PC1 CTF were not fully converged since certain variations were observed among the individual simulations. Nevertheless, these calculations allowed us to identify representative low-energy binding conformations of the peptides. Pep-GaMD simulations sampled an antiparallel ß-strand and a short helical secondary structure of peptide p17 bound to the ΔStalk CTF. Furthermore, peptides p9 and p21 adopted a more folded structure as compared to their disordered loop conformations in the docking poses. We also observed TOP-PL interactions, particularly the salt bridge between residues R3848-E4078 that is a key feature of the stalk TA-mediated activation of signaling for PC1 CTF[28]. Signal transduction was initiated upon binding of the stalk (TA) to the TOP domain, which was transmitted to the PL via a salt bridge formation between residue R3848 in the TOP domain and residue E4078 in the PL. The bound peptide agonists p9, p17 and p21 maintained the ΔStalk CTF in its "Closed/Active" conformation as observed in the wildtype PC1 CTF simulations[28].

The interacting pairs identified using sequence-based covariation analysis matched the pairs identified by Pep-GaMD simulations, providing complementary evidence of the importance of these interactions and of the existence of the "Bound" and "Intermediate" binding states of the stalk TA and stalk-derived peptide agonist. This suggests that such stalk TA binding states are evolutionarily conserved across PC1 orthologs. Covariation analysis identifies interactions important during any part of the protein lifecycle, and alone cannot be used to distinguish which conformational state an interaction arises in. By comparison to the conformations found in the Pep-GaMD simulations, we found that most of the identified interactions between the stalk TA and TOP domain were consistent with either the "Intermediate" or "Bound" binding states of the stalk-derived peptides, which are related to CTF inactive and active signaling states, however, it remained possible that other interactions, such as between the start of the stalk-TA and TM2/TM3,

may be related to conformational states necessary for cleavage of the GAIN/GPS domain. Additionally, structural contacts may be incompletely detected at some positions when the statistical signal of covariation is masked by high conservation, subfamily specialization, or misalignment. This can explain why some interactions identified in the binding interface through docking are not detected using covariation analysis. Despite this, the specific subset of interactions detected using covariation analysis suggest broader peptide binding interfaces, and we found these to be consistent with the peptide binding interactions observed in the Pep-GaMD simulations and the MM/PBSA binding free energy analysis, and the covarying residue pairings were consistent with functional biophysical interactions.

The proposed binding interactions of the PC1 stalk peptides shares some similarity with those observed for the ADGRs. Specifically, Xiao et al. resolved cryo-EM structures of active ADGRG2 and ADGRG4 in complex with tethered Stachel sequences[27]. The structures showed that the 15 residue Stachel sequence inserts into the TM bundle to form intense hydrophobic interactions. A hydrophobic F/Y/LXφφφXφ motif identified in the ADGR tethered sequences formed five finger-like projections in the hydrophobic pits of the TM bundle[27]. In our study, we observed a similar pattern of intense hydrophobic interactions between the peptide agonists p9, p17 and p21 and the hydrophobic pockets in the TOP domain of PC1 CTF. Notably, a closely similar TOP binding pocket was identified for interaction of the tethered agonist (Stalk) in our previous study[28] and for binding of peptide agonist p21 in this study. The TOP domain hydrophobic pocket may serve as a significant candidate binding site for designing new synthetic peptides or small molecules to aid in rescue of PC1 function levels. Moreover, the shorter peptide agonists' (p9 and p17) binding sites also serve as novel pockets for design and development of therapeutic approaches for treating ADPKD. While the present study is focused on identification

of initial peptides that are able to activate the PC1 CTF, we shall include further mutation experiments and simulations, peptide SAR and optimization of the lead peptides in future studies. It is also important to note that we have not tested the selectivity of the peptides for PC1 versus PC2 in the present study primarily because transfection of PC2 does not activate the NFAT reporter. However, it is possible that co-transfection of PC2 with the PC1 CTF could alter the stalk peptide binding. This will be important to consider in future studies.

Acknowledgments

We thank Dr. Yan Zhang for her valuable discussions. This work used supercomputing resources with allocation award TG-MCB180049 through the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296, and project M2874 through the National Energy Research Scientific Computing Center (NERSC), which is a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231, and the Research Computing Cluster at the University of Kansas. This work was supported in part by National Institutes of Health (R01DK123590 and R56DK135824), Department of Defense CDMRP PRMRP Discovery Award (PR160710/W81XWH-17-1-0301), and Pilot Grant funding from the School of Health Professions at KU Medical Center (to R.L.M.), Pilot award 1004015 from Jared Grantham Kidney Institute at KU Medical Center (to Y.M. and R.L.M.) and startup project 27110 at University of North Carolina - Chapel Hill (to Y.M.), and A.H acknowledges support from National Institutes of Health (R35-GM132090 and OD020095). This research includes calculations carried out on HPC resources supported in part by the National

Science Foundation through major research instrumentation grant number 1625061 and by the US Army Research Laboratory under contract number W911NF-16-2-0189.

Author Contributions

R.M. and Y.M. designed research; S.P., B. S. M., E. N. M. and A.H. performed research; S.P., B. S. M., K.J; E. N. M., A.H., R.M. and Y.M. analyzed data; and S.P., K.J; A.H., R.M. and Y.M. wrote the paper.

Competing Interest Statement

No competing interests.

REFERENCES

(1) Harris, P. C.; Torres, V. E. Genetic mechanisms and signaling pathways in autosomal dominant polycystic kidney disease. *The Journal of clinical investigation* **2014**, *124* (6), 2315-2324.

(2) Ingelfinger, J. R. Tolvaptan and autosomal dominant polycystic kidney disease. Mass Medical Soc: 2017; Vol. 377, pp 1988-1989.

(3) Hopp, K.; Ward, C. J.; Hommerding, C. J.; Nasr, S. H.; Tuan, H.-F.; Gainullin, V. G.; Rossetti, S.; Torres, V. E.; Harris, P. C. Functional polycystin-1 dosage governs autosomal dominant polycystic kidney disease severity. *The Journal of clinical investigation* **2012**, *122* (11), 4257-4273.

(4) Cai, Y.; Fedeles, S. V.; Dong, K.; Anyatonwu, G.; Onoe, T.; Mitobe, M.; Gao, J.-D.; Okuhara, D.; Tian, X.; Gallagher, A.-R. Altered trafficking and stability of polycystins underlie polycystic kidney disease. *The Journal of clinical investigation* **2014**, *124* (12), 5129-5144.

(5) Hofherr, A.; Wagner, C. J.; Watnick, T.; Köttgen, M. Targeted rescue of a polycystic kidney disease mutation by lysosomal inhibition. *Kidney international* **2016**, *89* (4), 949-955.

(6) Krappitz, M.; Gallagher, A.-R.; Fedeles, S. Is it time to fold the cysts away? *Trends in molecular medicine* **2016**, *22* (12), 997-999.

(7) Lakhia, R.; Ramalingam, H.; Chang, C.-M.; Cobo-Stark, P.; Biggers, L.; Flaten, A.; Alvarez, J.; Valencia, T.; Wallace, D. P.; Lee, E. C. PKD1 and PKD2 mRNA cis-inhibition drives polycystic kidney disease progression. *Nature communications* **2022**, *13* (1), 4765.

(8) Tan, Y.-C.; Blumenfeld, J.; Rennert, H. Autosomal dominant polycystic kidney disease: genetics, mutations and microRNAs. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* **2011**, *1812* (10), 1202-1212.

(9) Rossetti, S.; Kubly, V. J.; Consugar, M. B.; Hopp, K.; Roy, S.; Horsley, S. W.; Chauveau, D.; Rees, L.; Barratt, T. M.; Van't Hoff, W. G. Incompletely penetrant PKD1 alleles suggest a role for gene dosage in cyst initiation in polycystic kidney disease. *Kidney international* **2009**, *75* (8), 848-855.

(10) Heyer, C. M.; Sundsbak, J. L.; Abebe, K. Z.; Chapman, A. B.; Torres, V. E.; Grantham, J. J.; Bae, K. T.; Schrier, R. W.; Perrone, R. D.; Braun, W. E. Predicted mutation strength of nontruncating PKD1 mutations aids genotype-phenotype correlations in autosomal dominant polycystic kidney disease. *Journal of the American Society of Nephrology* **2016**, *27* (9), 2872-2884.

(11) Araç, D.; Boucard, A. A.; Bolliger, M. F.; Nguyen, J.; Soltis, S. M.; Südhof, T. C.; Brunger, A. T. A novel evolutionarily conserved domain of cell-adhesion GPCRs mediates autoproteolysis. *The EMBO journal* **2012**, *31* (6), 1364-1378.

(12) Qian, F.; Boletta, A.; Bhunia, A. K.; Xu, H.; Liu, L.; Ahrabi, A. K.; Watnick, T. J.; Zhou, F.; Germino, G. G. Cleavage of polycystin-1 requires the receptor for egg jelly domain and is disrupted by human autosomal-dominant polycystic kidney disease 1-associated mutations. *Proceedings of the National Academy of Sciences* **2002**, *99* (26), 16981-16986.

(13) Kurbegovic, A.; Kim, H.; Xu, H.; Yu, S.; Cruanès, J.; Maser, R. L.; Boletta, A.; Trudel, M.; Qian, F. Novel functional complexity of polycystin-1 by GPS cleavage in vivo: role in polycystic kidney disease. *Molecular and cellular biology* **2014**, *34* (17), 3341-3353.

(14) Kim, S.; Nie, H.; Nesin, V.; Tran, U.; Outeda, P.; Bai, C.-X.; Keeling, J.; Maskey, D.; Watnick, T.; Wessely, O. The polycystin complex mediates Wnt/Ca2+ signalling. *Nature cell biology* **2016**, *18* (7), 752-764.

(15) Weston, B. S.; Bagnéris, C.; Price, R. G.; Stirling, J. L. The polycystin-1 C-type lectin domain binds carbohydrate in a calcium-dependent manner, and interacts with extracellular matrix proteins

in vitro. *Biochimica Et Biophysica Acta (BBA)-Molecular Basis of Disease* **2001**, *1536* (2-3), 161-176.
(16) Sandford, R.; Sgotto, B.; Aparicio, S.; Brenner, S.; Vaudin, M.; Wilson, R. K.; Chissoe, S.; Pepin, K.; Bateman, A.; Chothia, C. Comparative analysis of the polycystic kidney disease 1 (PKD1) gene reveals an integral membrane glycoprotein with multiple evolutionary conserved domains. *Human molecular genetics* **1997**, *6* (9), 1483-1489.
(17) Ibraghimov-Beskrovnaya, O.; Bukanov, N. O.; Donohue, L. C.; Dackowski, W. R.; Klinger, K. W.; Landes, G. M. Strong homophilic interactions of the Ig-like domains of polycystin-1, the protein product of an autosomal dominant polycystic kidney disease gene, PKD1. *Human molecular genetics* **2000**, *9* (11), 1641-1649.
(18) Weston, B. S.; Malhas, A. N.; Price, R. G. Structure–function relationships of the extracellular domain of the autosomal dominant polycystic kidney disease-associated protein, polycystin-1. *FEBS letters* **2003**, *538* (1-3), 8-13.
(19) Nims, N.; Vassmer, D.; Maser, R. L. Transmembrane domain analysis of polycystin-1, the product of the polycystic kidney disease-1 (PKD1) gene: evidence for 11 membrane-spanning domains. *Biochemistry* **2003**, *42* (44), 13035-13048.
(20) Maser, R. L.; Calvet, J. P. Adhesion GPCRs as a paradigm for understanding polycystin-1 G protein regulation. *Cellular signalling* **2020**, *72*, 109637.
(21) Maser, R. L.; Calvet, J. P.; Parnell, S. C. The GPCR properties of polycystin-1-A new paradigm. *Frontiers in Molecular Biosciences* **2022**, *9*, 1035507.
(22) Yu, S.; Hackmann, K.; Gao, J.; He, X.; Piontek, K.; García-González, M. A.; Menezes, L. F.; Xu, H.; Germino, G. G.; Zuo, J. Essential role of cleavage of Polycystin-1 at G protein-coupled receptor proteolytic site for kidney tubular structure. *Proceedings of the National Academy of Sciences* **2007**, *104* (47), 18688-18693.
(23) Liebscher, I.; Schön, J.; Petersen, S. C.; Fischer, L.; Auerbach, N.; Demberg, L. M.; Mogha, A.; Cöster, M.; Simon, K.-U.; Rothemund, S. A tethered agonist within the ectodomain activates the adhesion G protein-coupled receptors GPR126 and GPR133. *Cell reports* **2014**, *9* (6), 2018-2026.
(24) Schöneberg, T.; Liebscher, I.; Luo, R.; Monk, K. R.; Piao, X. Tethered agonists: a new mechanism underlying adhesion G protein-coupled receptor activation. *Journal of Receptors and Signal Transduction* **2015**, *35* (3), 220-223.
(25) Demberg, L. M.; Rothemund, S.; Schöneberg, T.; Liebscher, I. Identification of the tethered peptide agonist of the adhesion G protein-coupled receptor GPR64/ADGRG2. *Biochemical and biophysical research communications* **2015**, *464* (3), 743-747.
(26) Stoveken, H. M.; Hajduczok, A. G.; Xu, L.; Tall, G. G. Adhesion G protein-coupled receptors are activated by exposure of a cryptic tethered agonist. *Proceedings of the National Academy of Sciences* **2015**, *112* (19), 6194-6199.
(27) Xiao, P.; Guo, S.; Wen, X.; He, Q.-T.; Lin, H.; Huang, S.-M.; Gou, L.; Zhang, C.; Yang, Z.; Zhong, Y.-N. Tethered peptide activation mechanism of the adhesion GPCRs ADGRG2 and ADGRG4. *Nature* **2022**, *604* (7907), 771-778.
(28) Pawnikar, S.; Magenheimer, B. S.; Munoz, E. N.; Maser, R. L.; Miao, Y. Mechanism of tethered agonist-mediated signaling by polycystin-1. *Proceedings of the National Academy of Sciences* **2022**, *119* (19), e2113786119.
(29) Miao, Y.; Sinko, W.; Pierce, L.; Bucher, D.; Walker, R. C.; McCammon, J. A. Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. *Journal of chemical theory and computation* **2014**, *10* (7), 2677-2689.

(30) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian accelerated molecular dynamics: unconstrained enhanced sampling and free energy calculation. *Journal of chemical theory and computation* **2015**, *11* (8), 3584-3595.
(31) Miao, Y.; McCammon, J. A. Graded activation and free energy landscapes of a muscarinic G-protein–coupled receptor. *Proceedings of the National Academy of Sciences* **2016**, *113* (43), 12162-12167.
(32) Pang, Y. T.; Miao, Y.; Wang, Y.; McCammon, J. A. Gaussian accelerated molecular dynamics in NAMD. *Journal of chemical theory and computation* **2017**, *13* (1), 9-19.
(33) Miao, Y.; McCammon, J. A. Gaussian accelerated molecular dynamics: theory, implementation, and applications. In *Annual reports in computational chemistry*, Vol. 13; Elsevier, 2017; pp 231-278.
(34) Wang, Y.-T.; Chan, Y.-H. Understanding the molecular basis of agonist/antagonist mechanism of human mu opioid receptor through gaussian accelerated molecular dynamics method. *Scientific reports* **2017**, *7* (1), 1-11.
(35) Liao, J.-M.; Wang, Y.-T. In silico studies of conformational dynamics of Mu opioid receptor performed using gaussian accelerated molecular dynamics. *Journal of Biomolecular Structure and Dynamics* **2019**, *37* (1), 166-177.
(36) Miao, Y.; Bhattarai, A.; Nguyen, A. T.; Christopoulos, A.; May, L. T. Structural basis for binding of allosteric drug leads in the adenosine A1 receptor. *Scientific reports* **2018**, *8* (1), 16836.
(37) Chuang, C.-H.; Chiou, S.-j.; Cheng, T.-L.; Wang, Y.-T. A molecular dynamics simulation study decodes the Zika virus NS5 methyltransferase bound to SAH and RNA analogue. *Scientific reports* **2018**, *8* (1), 6336.
(38) Sibener, L. V.; Fernandes, R. A.; Kolawole, E. M.; Carbone, C. B.; Liu, F.; McAffee, D.; Birnbaum, M. E.; Yang, X.; Su, L. F.; Yu, W. Isolation and visualization of a structural trigger that uncouples TCR signaling from pMHC binding. *Cell* **2018**, *174* (3), 672.
(39) Park, J. B.; Kim, Y. H.; Yoo, Y.; Kim, J.; Jun, S.-H.; Cho, J. W.; El Qaidi, S.; Walpole, S.; Monaco, S.; García-García, A. A. Structural basis for arginine glycosylation of host substrates by bacterial effector proteins. *Nature communications* **2018**, *9* (1), 4283.
(40) Miao, Y.; McCammon, J. A. Mechanism of the G-protein mimetic nanobody binding to a muscarinic G-protein-coupled receptor. *Proceedings of the National Academy of Sciences* **2018**, *115* (12), 3036-3041.
(41) Ricci, C. G.; Chen, J. S.; Miao, Y.; Jinek, M.; Doudna, J. A.; McCammon, J. A.; Palermo, G. Deciphering off-target effects in CRISPR-Cas9 through accelerated molecular dynamics. *ACS central science* **2019**, *5* (4), 651-662.
(42) Palermo, G.; Miao, Y.; Walker, R. C.; Jinek, M.; McCammon, J. A. CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations. *Proceedings of the National Academy of Sciences* **2017**, *114* (28), 7260-7265.
(43) Bhattarai, A.; Wang, J.; Miao, Y. G-protein-coupled receptor–membrane interactions depend on the receptor activation state. *Journal of computational chemistry* **2020**, *41* (5), 460-471.
(44) Pawnikar, S.; Miao, Y. Pathway and mechanism of drug binding to chemokine receptors revealed by accelerated molecular simulations. *Future Medicinal Chemistry* **2020**, *12* (13), 1213-1225.
(45) Wang, J.; Miao, Y. Peptide Gaussian accelerated molecular dynamics (Pep-GaMD): Enhanced sampling and free energy and kinetics calculations of peptide binding. *The Journal of chemical physics* **2020**, *153* (15).

(46) Levy, R. M.; Haldane, A.; Flynn, W. F. Potts Hamiltonian models of protein co-variation, free energy landscapes, and evolutionary fitness. *Current opinion in structural biology* **2017**, *43*, 55-62.
(47) Su, Q.; Hu, F.; Ge, X.; Lei, J.; Yu, S.; Wang, T.; Zhou, Q.; Mei, C.; Shi, Y. Structure of the human PKD1-PKD2 complex. *Science* **2018**, *361* (6406), eaat9819.
(48) Zhou, P.; Jin, B.; Li, H.; Huang, S.-Y. HPEPDOCK: a web server for blind peptide–protein docking based on a hierarchical algorithm. *Nucleic acids research* **2018**, *46* (W1), W443-W450.
(49) Roe, D. R.; Cheatham III, T. E. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *Journal of chemical theory and computation* **2013**, *9* (7), 3084-3095.
(50) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A. Highly accurate protein structure prediction with AlphaFold. *nature* **2021**, *596* (7873), 583-589.
(51) Rastelli, G.; Rio, A. D.; Degliesposti, G.; Sgobba, M. Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *Journal of computational chemistry* **2010**, *31* (4), 797-810.
(52) Valdés-Tresanco, M. S.; Valdés-Tresanco, M. E.; Valiente, P. A.; Moreno, E. gmx_MMPBSA: a new tool to perform end-state free energy calculations with GROMACS. *Journal of chemical theory and computation* **2021**, *17* (10), 6281-6291.
(53) Magenheimer, B. S.; Munoz, E. N.; Ravichandran, J.; Maser, R. L. Constitutive signaling by the C-terminal fragment of polycystin-1 is mediated by a tethered peptide agonist. *bioRxiv* **2021**, 2021.2008. 2020.457171.

Figure Captions

Figure 1. Synthetic peptides derived from the stalk sequence of PC1 can stimulate signaling of stalkless PC1 CTF. (**A**) Alignment of CTF stalk sequences from human (h) and mouse (m) PC1. $CTF^{\Delta st}$ has a 21-residue deletion from the N-terminal end of the stalk region. Arrow, GPS cleavage site. Non-identical residues shown in bolded blue. **(B)** Activation of the NFAT-luc reporter by transfected mCTF or $mCTF^{\Delta st}$ expression constructs shown relative to empty expression vector (ev) as means (+ standard deviation, SD) of 3 wells/construct from each of 7 independent experiments. (**C**) Representative Western blot of total cell lysates from one of the experiments in (B), probed with antisera A19 against mouse PC1 C-tail. ns, non-specific. (**D**) Summary of the total expression levels (means +SD) of $CTF^{\Delta st}$ relative to CTF from the experiments in (B). (**E**) Stalk peptide treatment of ev- or $mCTF^{\Delta st}$-transfected cells. Sequences of stalk-derived peptides p7-p21 are shown. Graph represents the fold NFAT-luc activation for both eV- (gray bars) and $CTF^{\Delta st}$- (blue bars) transfected cells relative to the $CTF^{\Delta st}$ control after 24 hr treatment with or without peptide. Results are the means (+SD) of 3 separate experiments, each with 3 wells/condition. *, $p < 0.05$; ***, $p = 0.0001$; ****, $p < 0.0001$. Analysis by 1-way ANOVA with Tukey-Kramer post-test.

Figure 2. (**A**) Free energy profile of the p9-bound ΔStalk CTF regarding the number of atom contacts between p9 and extracellular domains of CTF and the distance between the CZ atom of R3848 and the CD atom of R4078 in CTF calculated from Pep-GaMD simulations. (**B-C**) Comparison of HPEPDOCK docking (cyan) and Pep-GaMD refined (magenta) conformations of peptide p9 in ***Δ***Stalk CTF. (**D**) Polar interactions between peptide-protein residues observed in the top-ranked representative conformations of p9. Peptide residues are numbered relative to the N terminus of the stalk with the peptide starting from 1, while residues within Δstalk CTF are

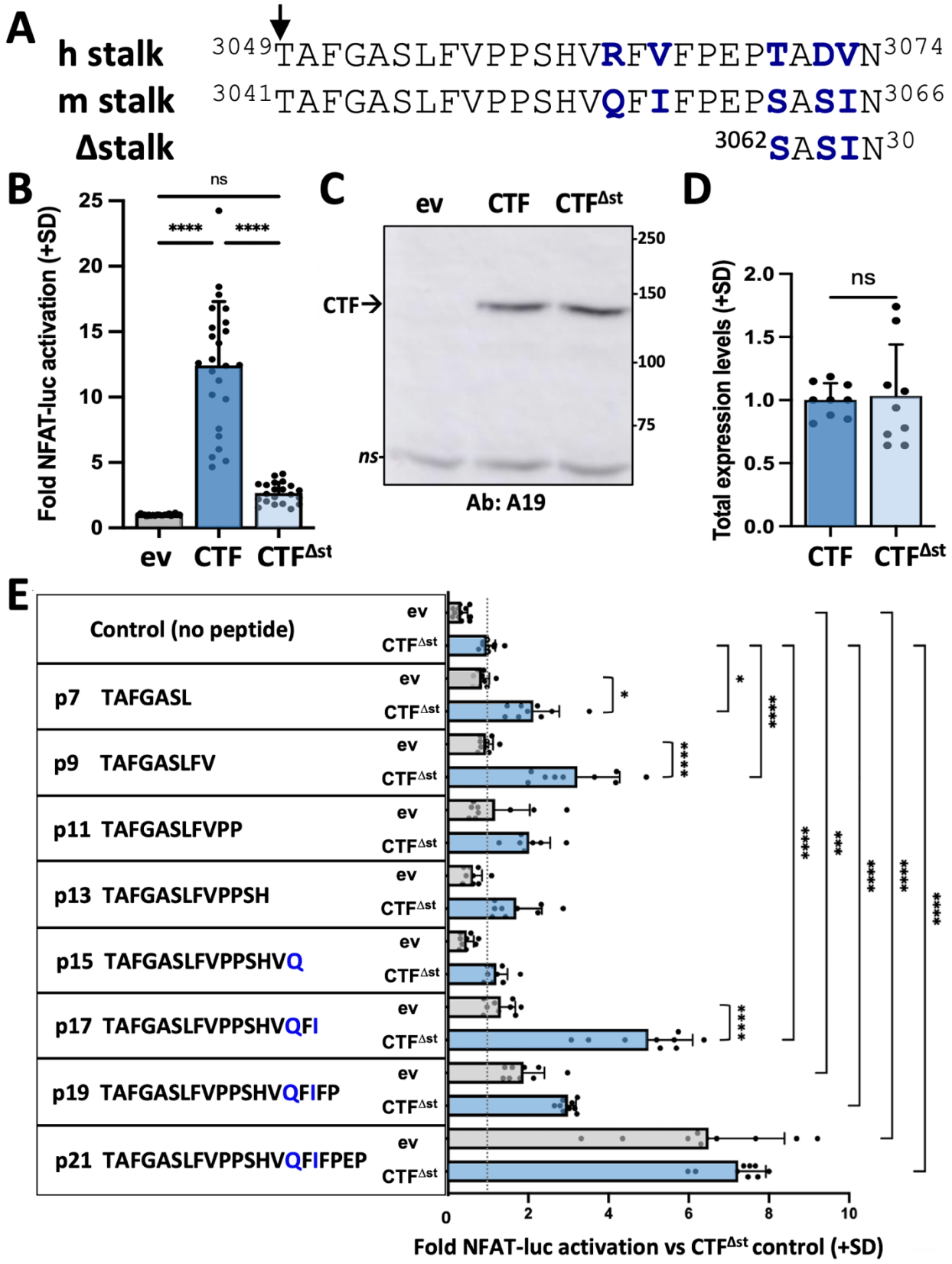
numbered according to the standard PC1 residue number. (**E**) Distance between TOP domain residue R3848 and PL residue E4078 observed in p9-bound ΔStalk CTF.

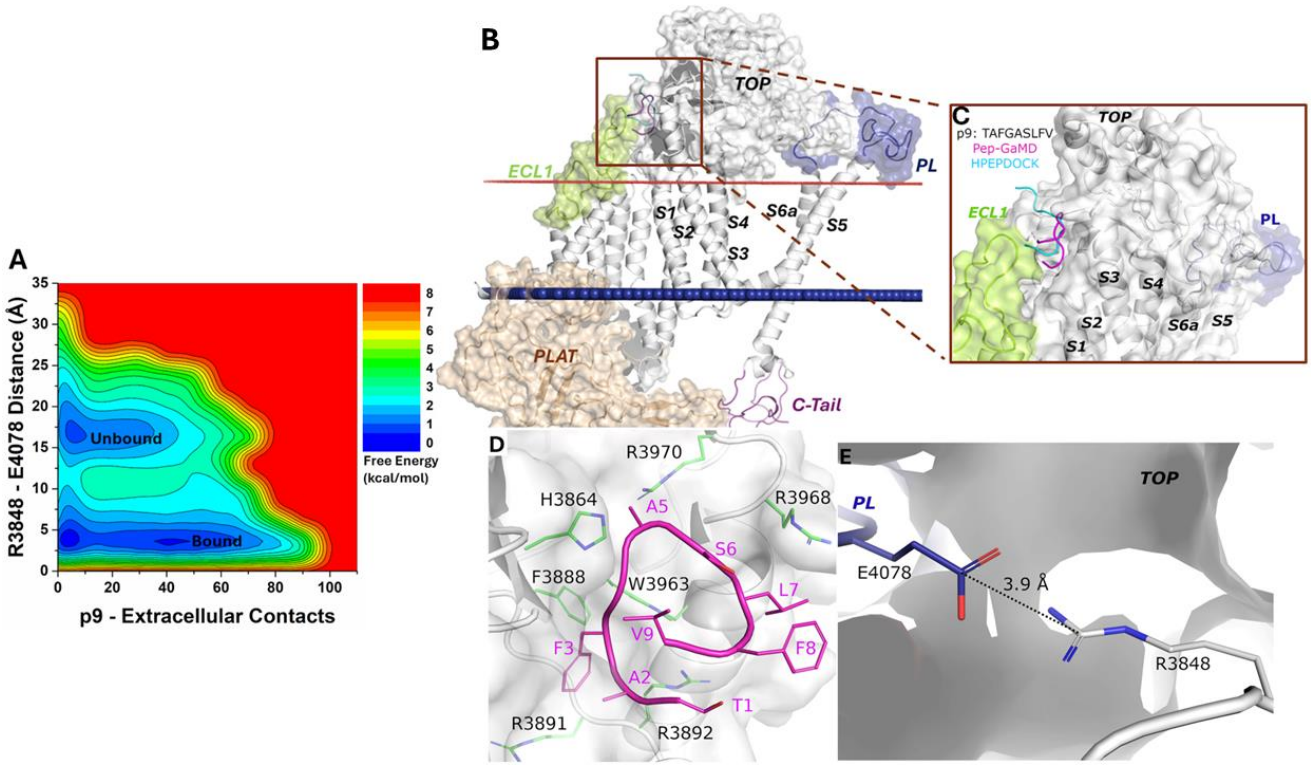
Figure 3. (**A**) Free energy profile of the p17-bound ΔStalk CTF regarding the number of atom contacts between p17 and extracellular domains of CTF and the distance between the CZ atom of R3848 and the CD atom of R4078 in CTF calculated from Pep-GaMD simulations. (**B-C**) Comparison of HPEPDOCK docking (cyan) and Pep-GaMD refined (magenta) conformations of peptide p17 in **Δ**Stalk CTF. Hydrophobic interactions (red dashed lines) between peptide-protein residues observed in the (**D**) "Bound" and (**F**) "Intermediate" low-energy conformations of p17-bound ΔStalk CTF. Distance between TOP domain residue R3848 and PL residue E4078 observed in the **(E)** "Bound" and **(G)** "Intermediate" low-energy conformations of p17-bound ΔStalk CTF.

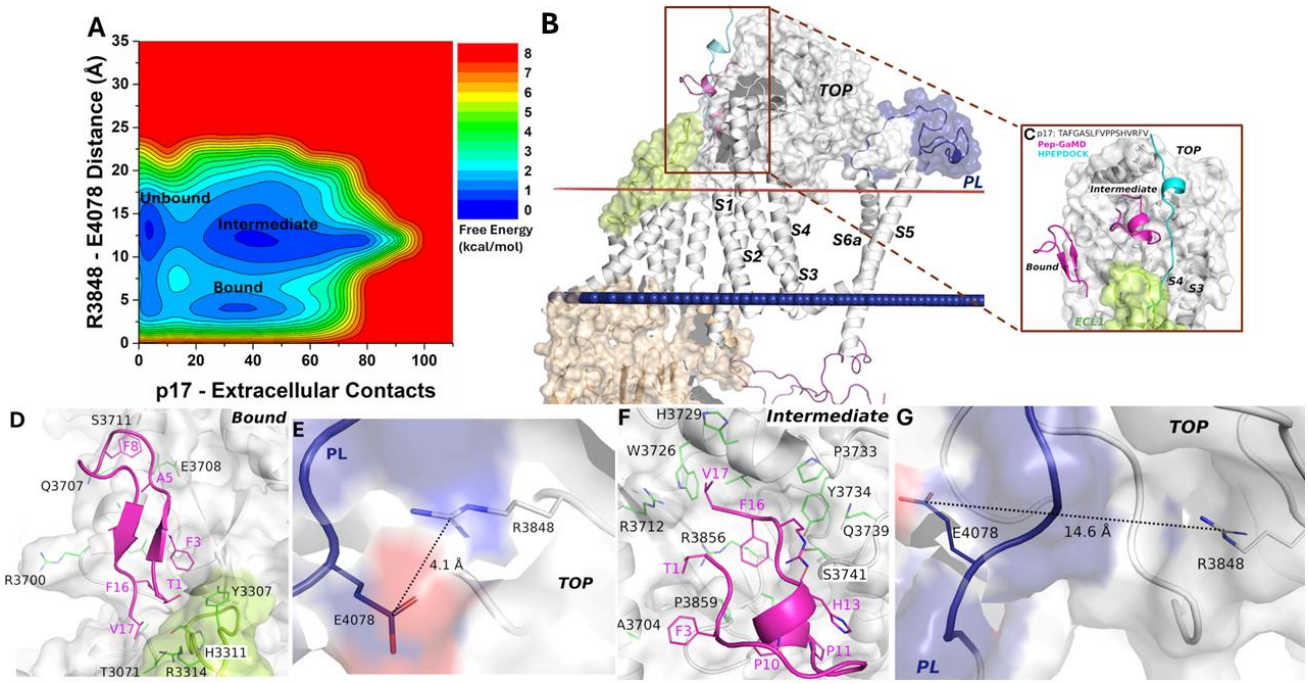
Figure 4. (**A**) Free energy profile of the p21-bound ΔStalk CTF regarding the number of atom contacts between p21 and extracellular domains of CTF and the distance between the CZ atom of R3848 and the CD atom of R4078 in CTF calculated from Pep-GaMD simulations. (**B-C**) Comparison of HPEPDOCK docking (cyan) and Pep-GaMD refined (magenta) conformations of peptide p21 in **Δ**Stalk CTF. (**D**) Polar interactions between peptide-protein residues observed in the top-ranked representative conformations of p21. (**E**) Distance between TOP domain residue R3848 and PL residue E4078 observed in p21-bound ΔStalk CTF.

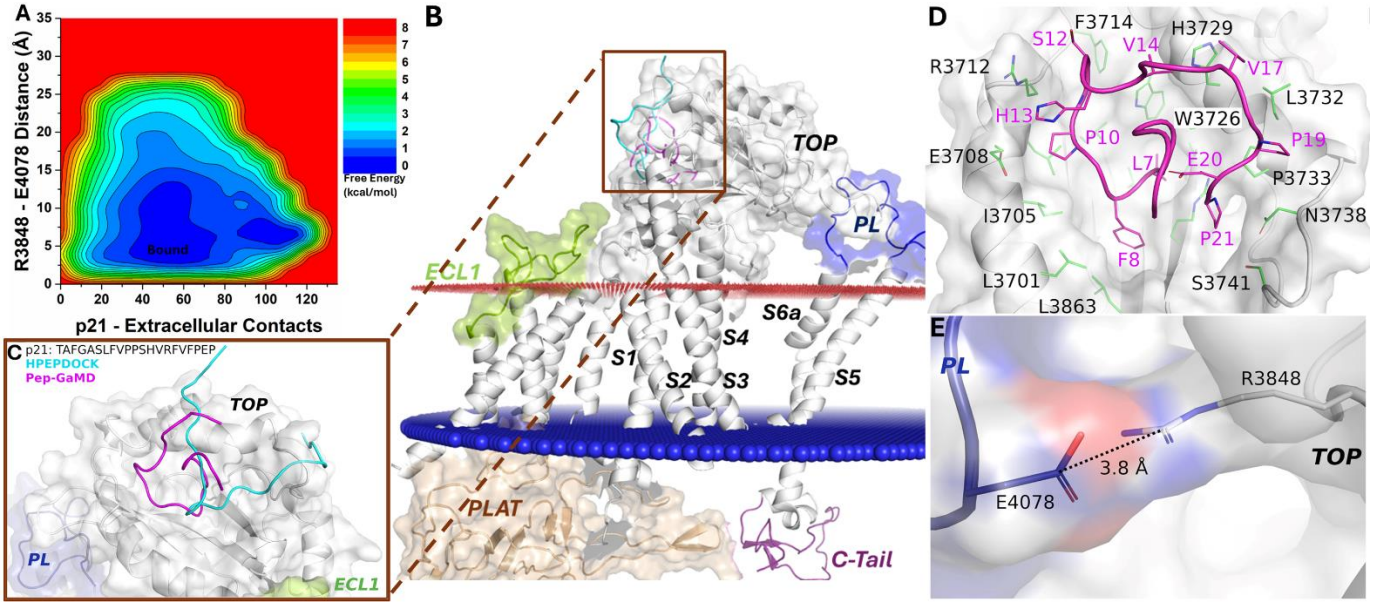
Figure 5. (**A**) Potts interaction map based on the PKD1 multiple-sequence-alignment illustrated in **Figure S4**, showing interactions with the stalk. Gray dots are shown for residue position-pairs with Potts covariation scores above a threshold, colored darker for higher scores, and selected interacting pairs are annotated with the stalk residue (horizontal, numbered from the stalk N-terminus) and other residue (vertical, standard PKD1 numbering) with the PKD1 residue at each position. The secondary structure as a function of position is annotated along the axes. (**B**) Cartoon

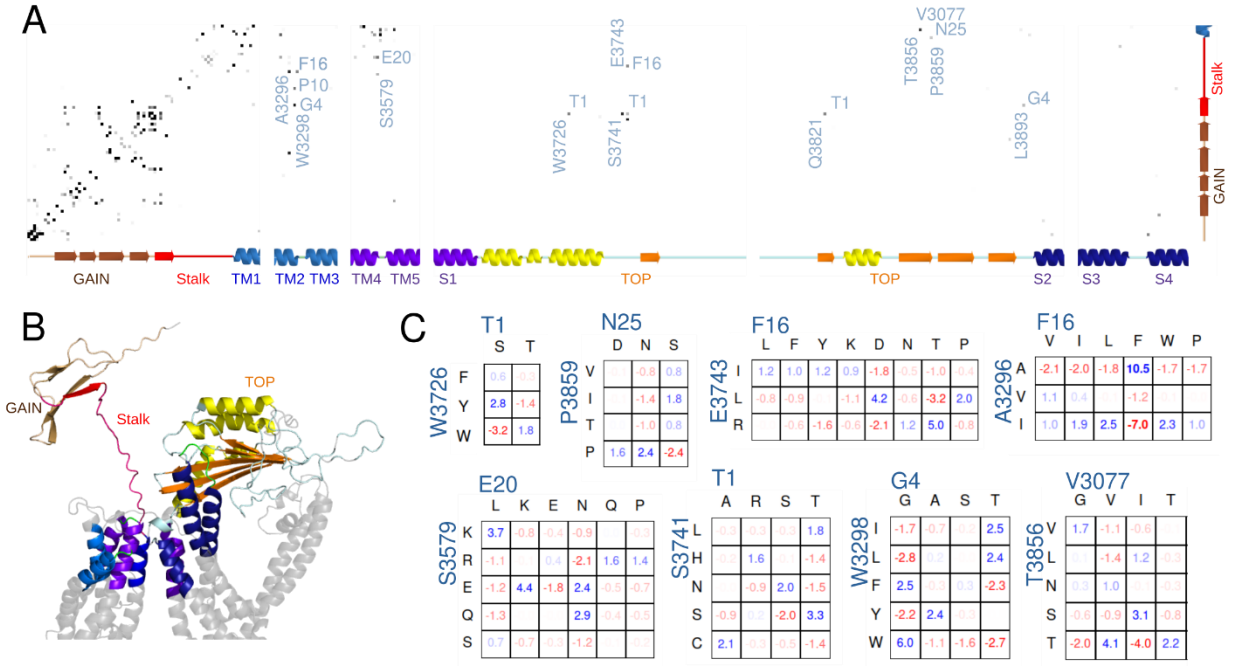
showing the subset of PC1 included in the Potts covariation analysis colored as in the secondary structure in panel A, using a structure predicted by AlphaFold. Gray regions were excluded from the Potts model. (**C**) Residue Covariation scores for selected position-pairs. The scores reflect the percentage excess frequency of the residue-pair relative to the null expected frequency if the MSA columns were uncorrelated, with blue values reflecting excess and red dearth. Only the most common residue types are shown.


A
h stalk 3049TAFGASLFVPPSHVRFVFPEPTADVN3074
m stalk 3041TAFGASLFVPPSHVQFIFPEPSASIN3066
Δstalk 3062SASIN30
B
Fold NFAT-luc activation (+SD)
ns
****
****
ev
CTF
CTFΔst
C
ev
CTF
CTFΔst
CTF
ns
250
150
100
75
Ab: A19
D
Total expression levels (+SD)
ns
CTF
CTFΔst
E
Control (no peptide)
p7 TAFGASL
p9 TAFGASLFV
p11 TAFGASLFVPP
p13 TAFGASLFVPPSH
p15 TAFGASLFVPPSHVQ
p17 TAFGASLFVPPSHVQFI
p19 TAFGASLFVPPSHVQFIFP
p21 TAFGASLFVPPSHVQFIFPEP
ev
CTFΔst
Fold NFAT-luc activation vs CTFΔst control (+SD)

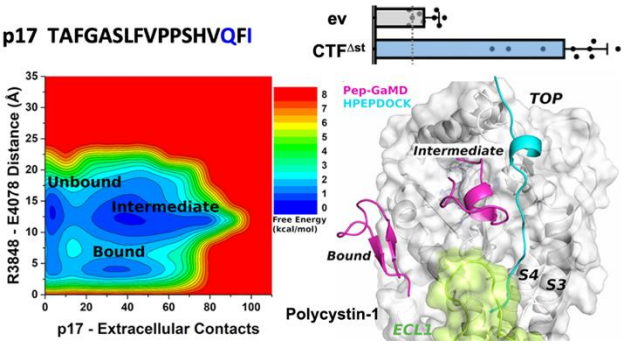

A
R3848 - E4078 Distance (Å)
p9 - Extracellular Contacts
Unbound
Bound
Free Energy (kcal/mol)
B
TOP
PL
ECL1
S1
S2
S3
S4
S6a
S5
PLAT
C-Tail
C
p9: TAFGASLFV
Pep-GaMD
HPEPDOCK
D
R3970
H3864
A5
R3968
S6
F3888
W3963
L7
V9
F3
F8
A2
T1
R3891
R3892
E
TOP
PL
E4078
3.9 Å
R3848


A
R3848 - E4078 Distance (Å)
p17 - Extracellular Contacts
Unbound
Intermediate
Bound
Free Energy (kcal/mol)
B
TOP
PL
S1
S2
S3
S4
S6a
S5
C
p17: TAFGASLFVPPSHVRFV
Pep-GaMD
HPEPDOCK
TOP
Intermediate
Bound
S4
S3
ECL1
D
Bound
S3711
F8
E3708
A5
Q3707
F3
R3700
F16
T1
Y3307
V17
H3311
T3071
R3314
E
PL
R3848
4.1 Å
E4078
TOP
F
Intermediate
H3729
W3726
P3733
V17
Y3734
F16
R3712
Q3739
R3856
T1
S3741
P3859
H13
A3704
F3
P11
P10
G
TOP
E4078
14.6 Å
R3848
PL

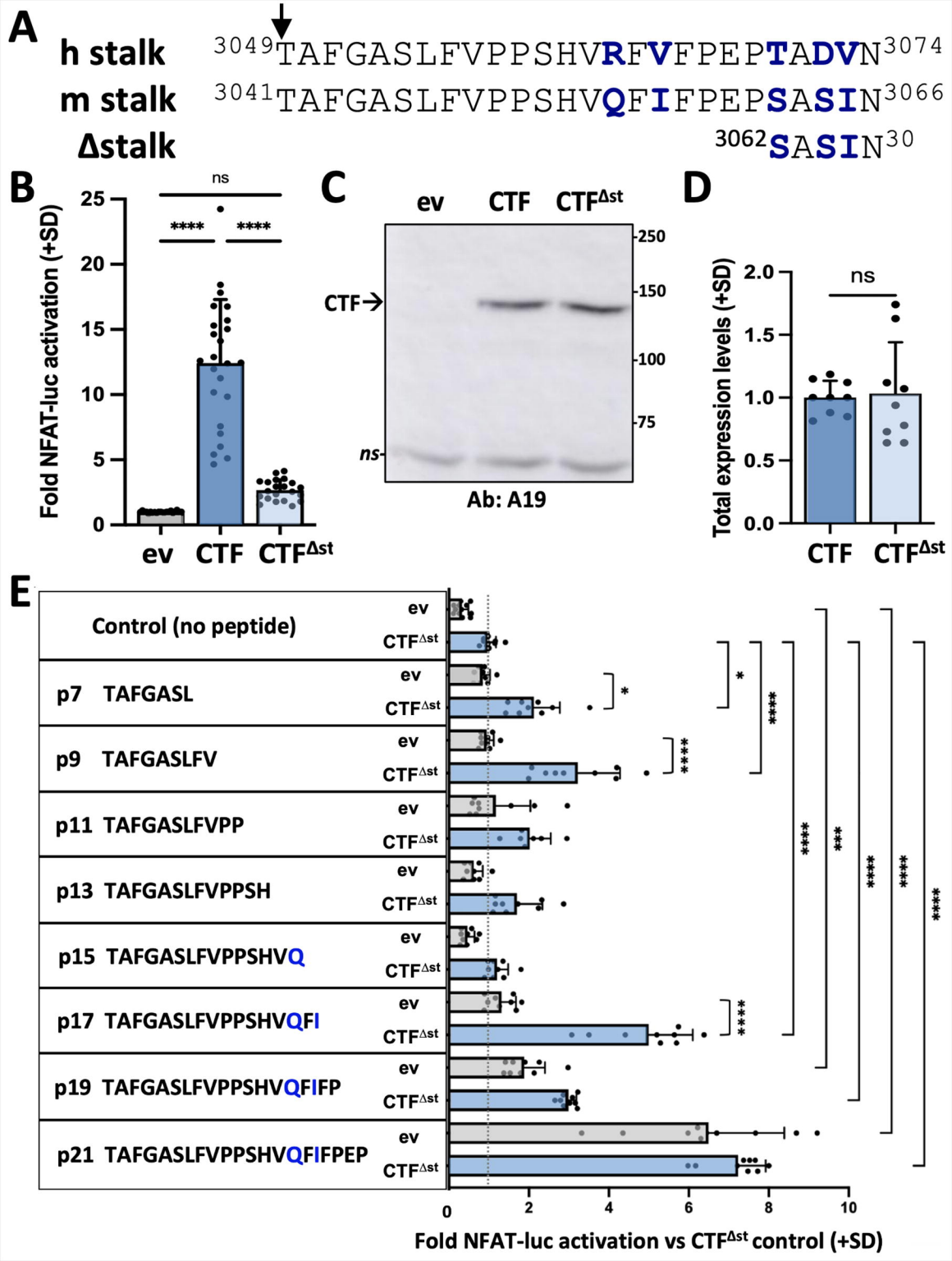

A
R3848 - E4078 Distance (Å)
p21 - Extracellular Contacts
Bound
Free Energy (kcal/mol)
B
TOP
PL
ECL1
S6a
S4
S1
S2
S3
S5
PLAT
C-Tail
C
p21: TAFGASLFVPPSHVRFVFPEP
HPEPDOCK
Pep-GaMD
TOP
PL
ECL1
D
F3714
H3729
S12
V14
V17
R3712
L3732
H13
W3726
P10
P19
E3708
L7
E20
P3733
I3705
N3738
P21
F8
L3701
S3741
L3863
E
PL
R3848
3.8 Å
TOP
E4078

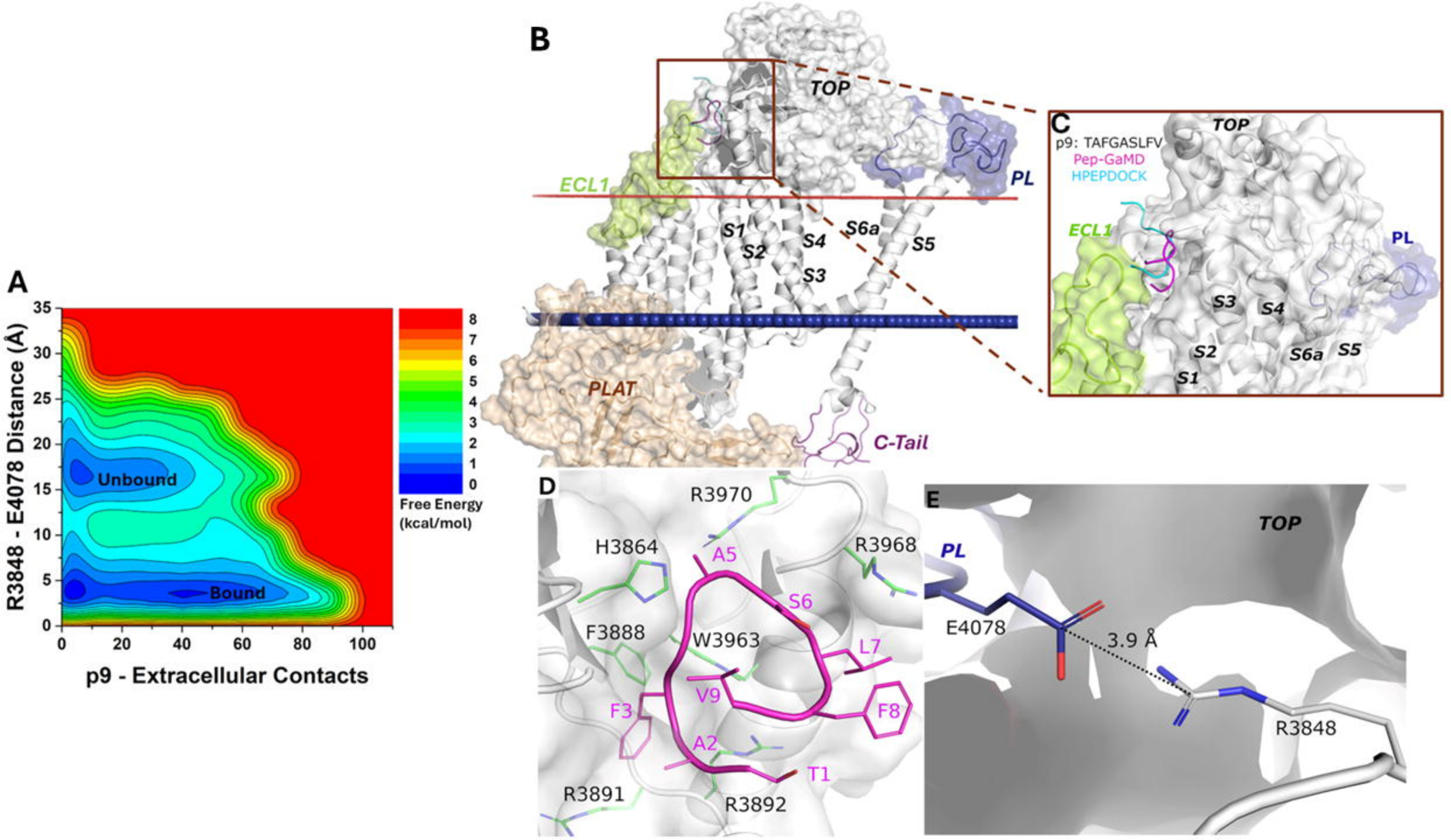

A
F16
P10
G4
A3296
W3298
E20
S3579
E3743
F16
T1
T1
W3726
S3741
T1
Q3821
V3077
N25
T3856
P3859
G4
L3893
Stalk
GAIN
GAIN
Stalk
TM1
TM2
TM3
TM4
TM5
S1
TOP
TOP
S2
S3
S4
B
GAIN
Stalk
TOP
C
T1
W3726
N25
P3859
F16
E3743
F16
A3296
E20
S3579
T1
S3741
G4
W3298
V3077
T3856

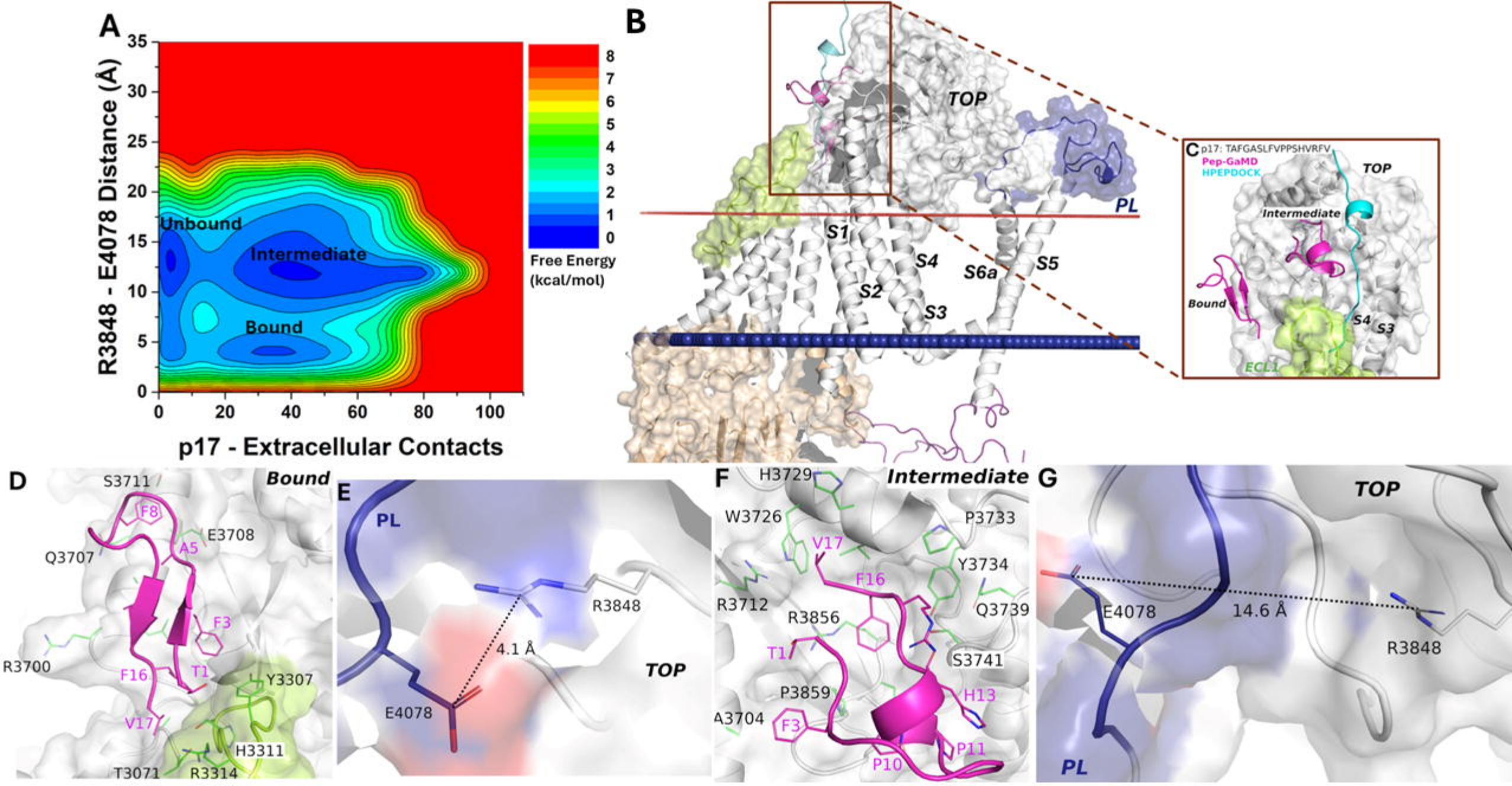
TOC Graphic

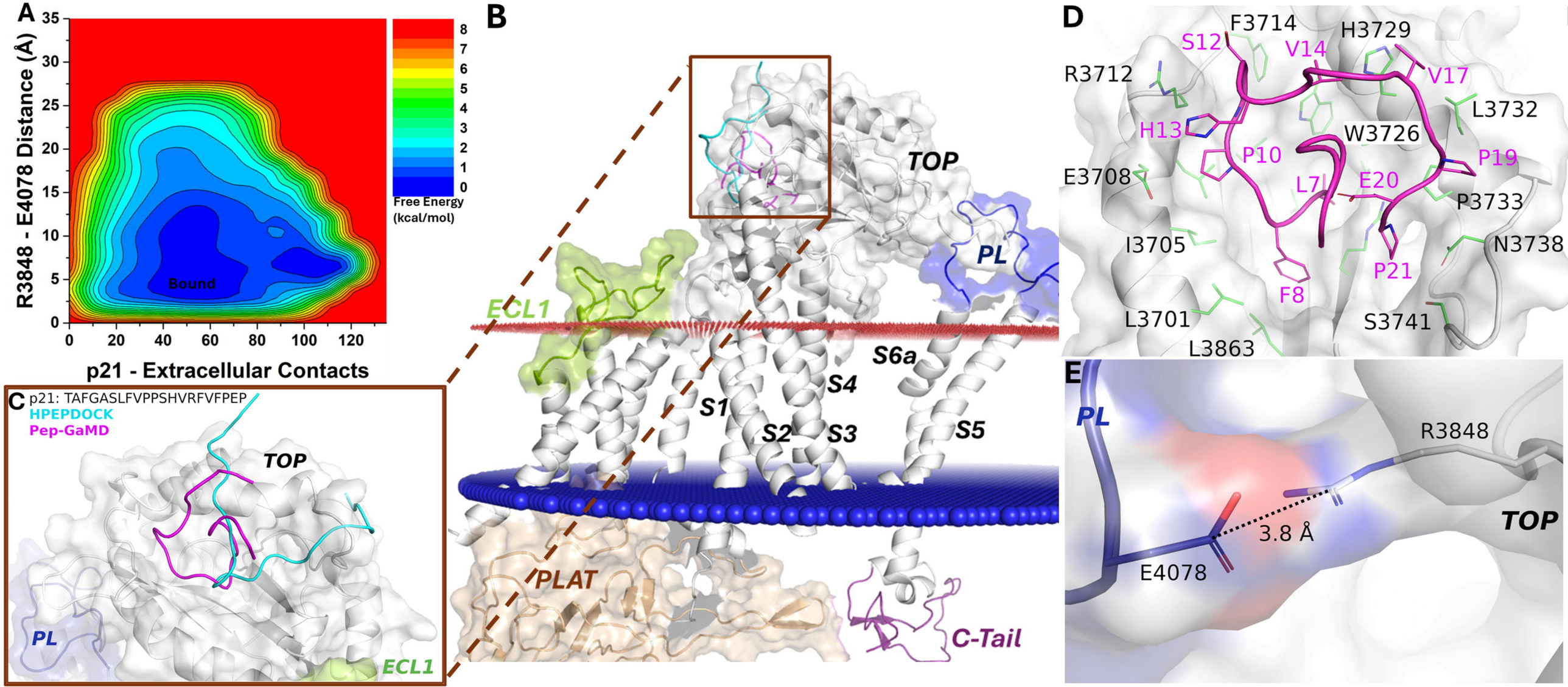


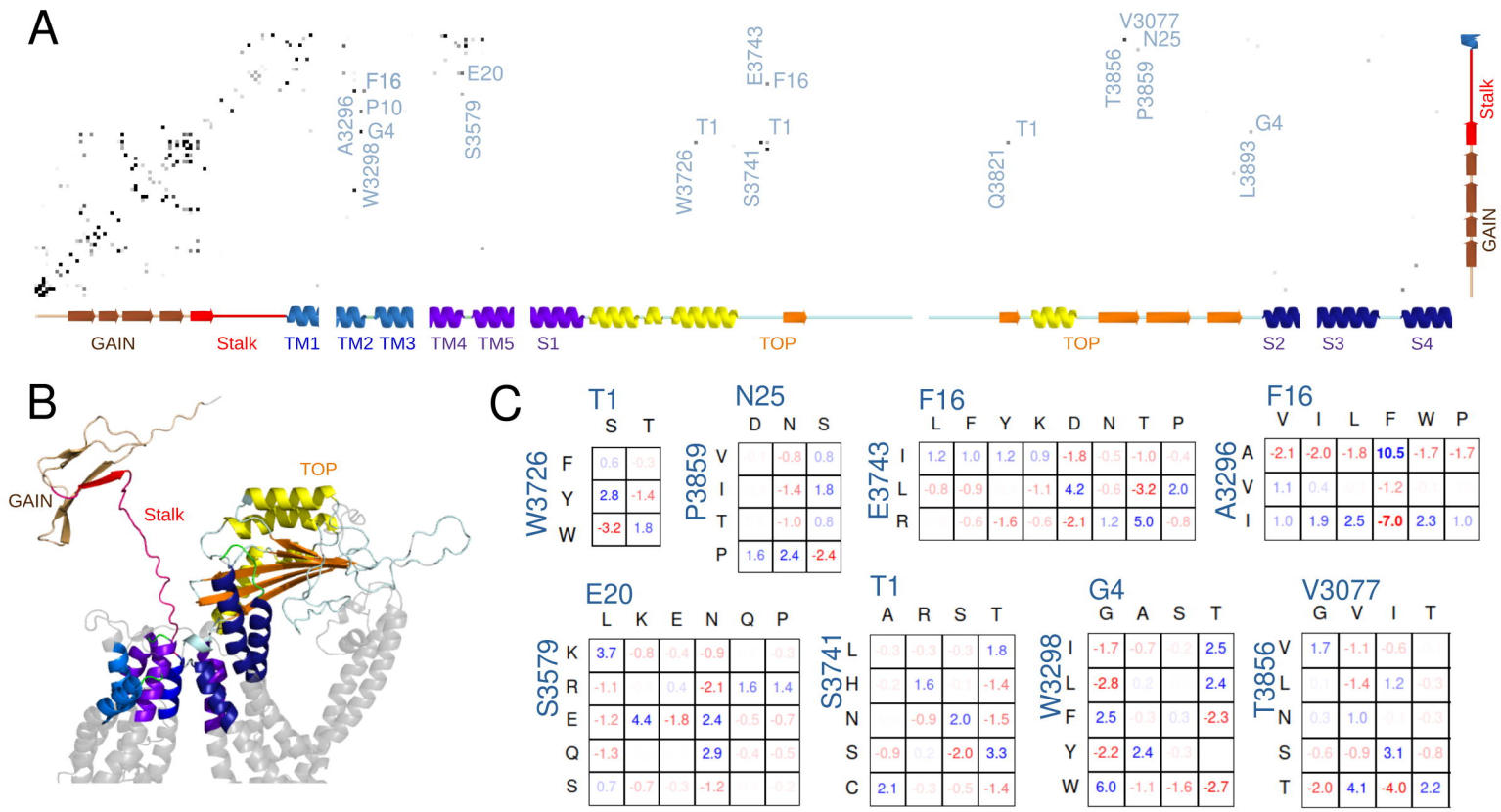
Structural dynamic models are presented for binding of novel synthetic, soluble peptide agonists and associated activation of Polycystin-1 through a combination of complementary cellular signaling assays, accelerated molecular simulations and sequence coevolutionary analysis. The p17 peptide derived from the first 17 residues of the protein stalk tethered agonist is shown here.

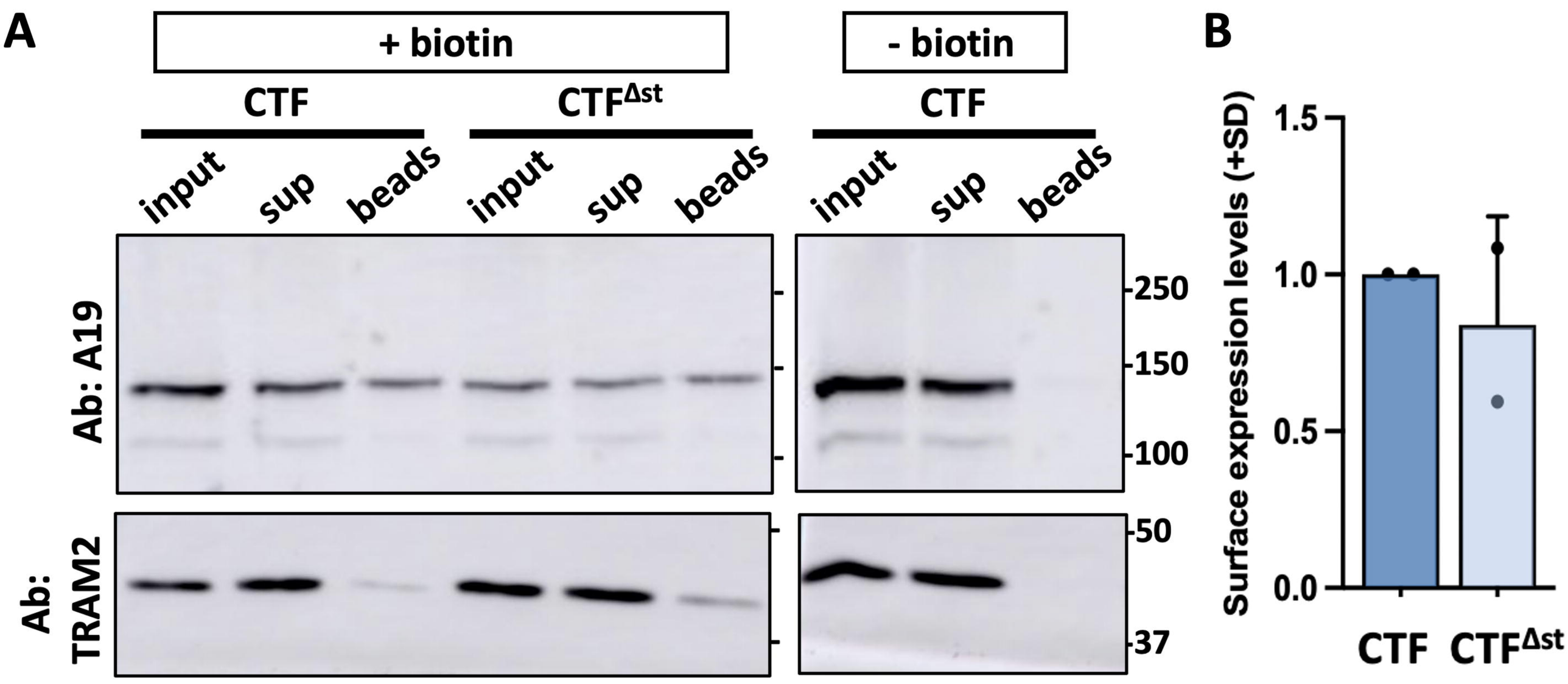

A
h stalk
3049TAFGASLFVPPSHVRFVFPEPTADVN3074
m stalk
3041TAFGASLFVPPSHVQFIFPEPSASIN3066
Δstalk
3062SASIN30
B
Fold NFAT-luc activation (+SD)
ns
****
****
ev
CTF
CTFΔst
C
ev
CTF
CTFΔst
CTF
ns
250
150
100
75
Ab: A19
D
Total expression levels (+SD)
ns
CTF
CTFΔst
E
Control (no peptide)
p7 TAFGASL
p9 TAFGASLFV
p11 TAFGASLFVPP
p13 TAFGASLFVPPSH
p15 TAFGASLFVPPSHVQ
p17 TAFGASLFVPPSHVQFI
p19 TAFGASLFVPPSHVQFIFP
p21 TAFGASLFVPPSHVQFIFPEP
ev
CTFΔst
Fold NFAT-luc activation vs CTFΔst control (+SD)

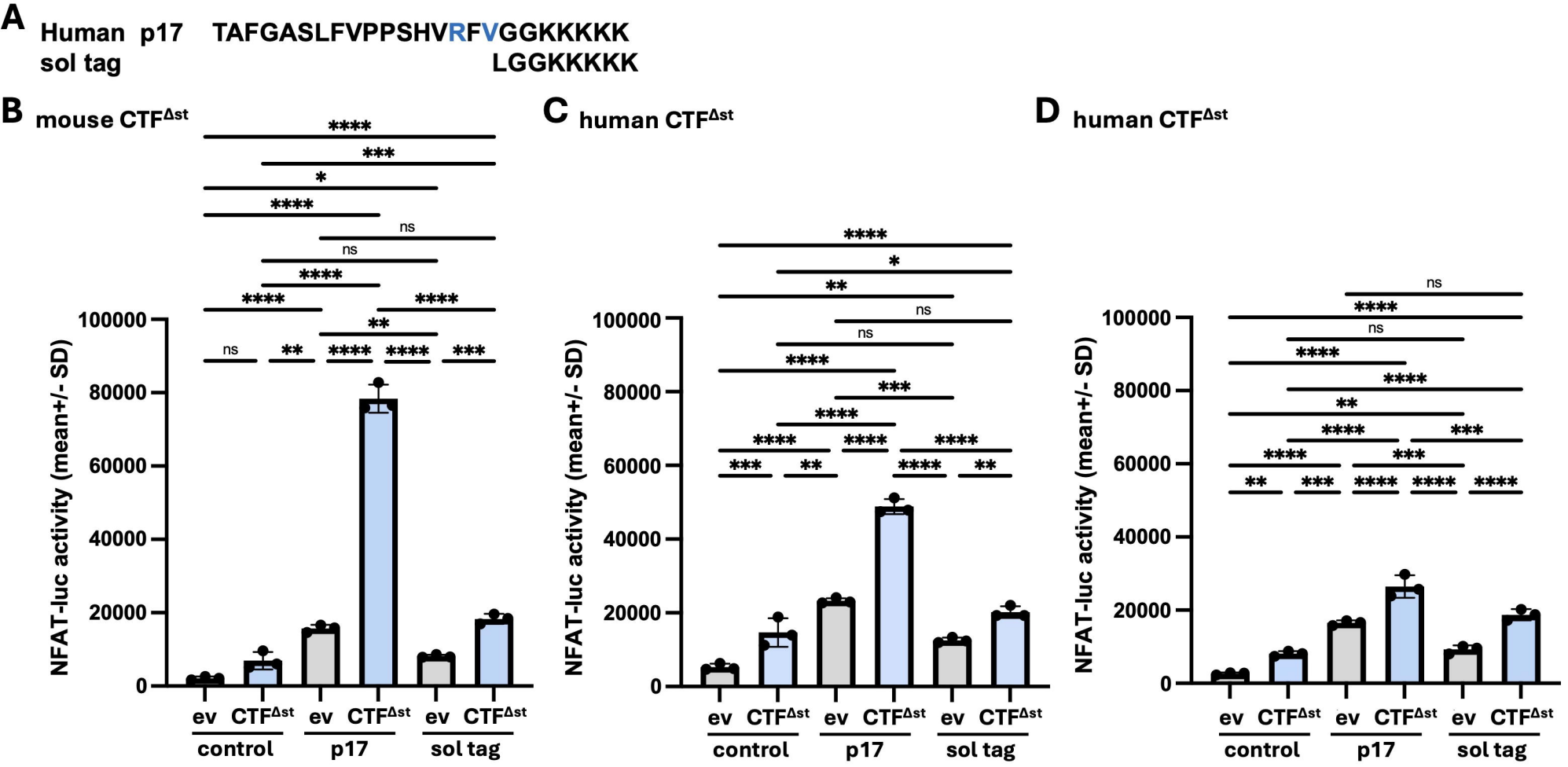

A
R3848 - E4078 Distance (Å)
p9 - Extracellular Contacts
Unbound
Bound
Free Energy (kcal/mol)
B
TOP
ECL1
PL
S1
S2
S3
S4
S6a
S5
PLAT
C-Tail
C
p9: TAFGASLFV
Pep-GaMD
HPEPDOCK
D
R3970
H3864
A5
R3968
S6
F3888
W3963
L7
V9
F3
F8
A2
T1
R3891
R3892
E
TOP
PL
E4078
3.9 Å
R3848

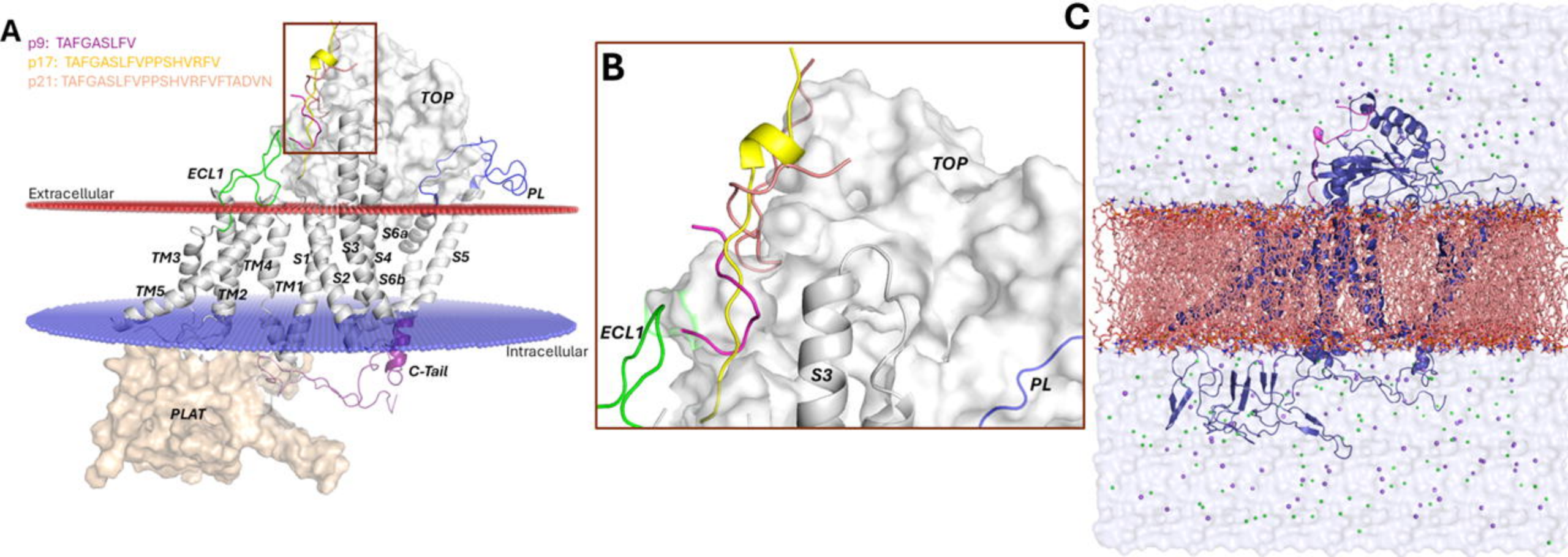

A
R3848 - E4078 Distance (Å)
p17 - Extracellular Contacts
Unbound
Intermediate
Bound
Free Energy (kcal/mol)
B
TOP
PL
S1
S2
S3
S4
S6a
S5
C
p17: TAFGASLFVPPSHVRFV
Pep-GaMD
HPEPDOCK
TOP
Intermediate
Bound
S4
S3
ECL1
D
Bound
S3711
F8
E3708
Q3707
A5
F3
R3700
F16
T1
Y3307
V17
H3311
T3071
R3314
E
PL
R3848
4.1 Å
TOP
E4078
F
H3729
Intermediate
W3726
P3733
V17
Y3734
F16
R3712
Q3739
R3856
T1
S3741
P3859
H13
A3704
F3
P11
P10
G
TOP
E4078
14.6 Å
R3848
PL

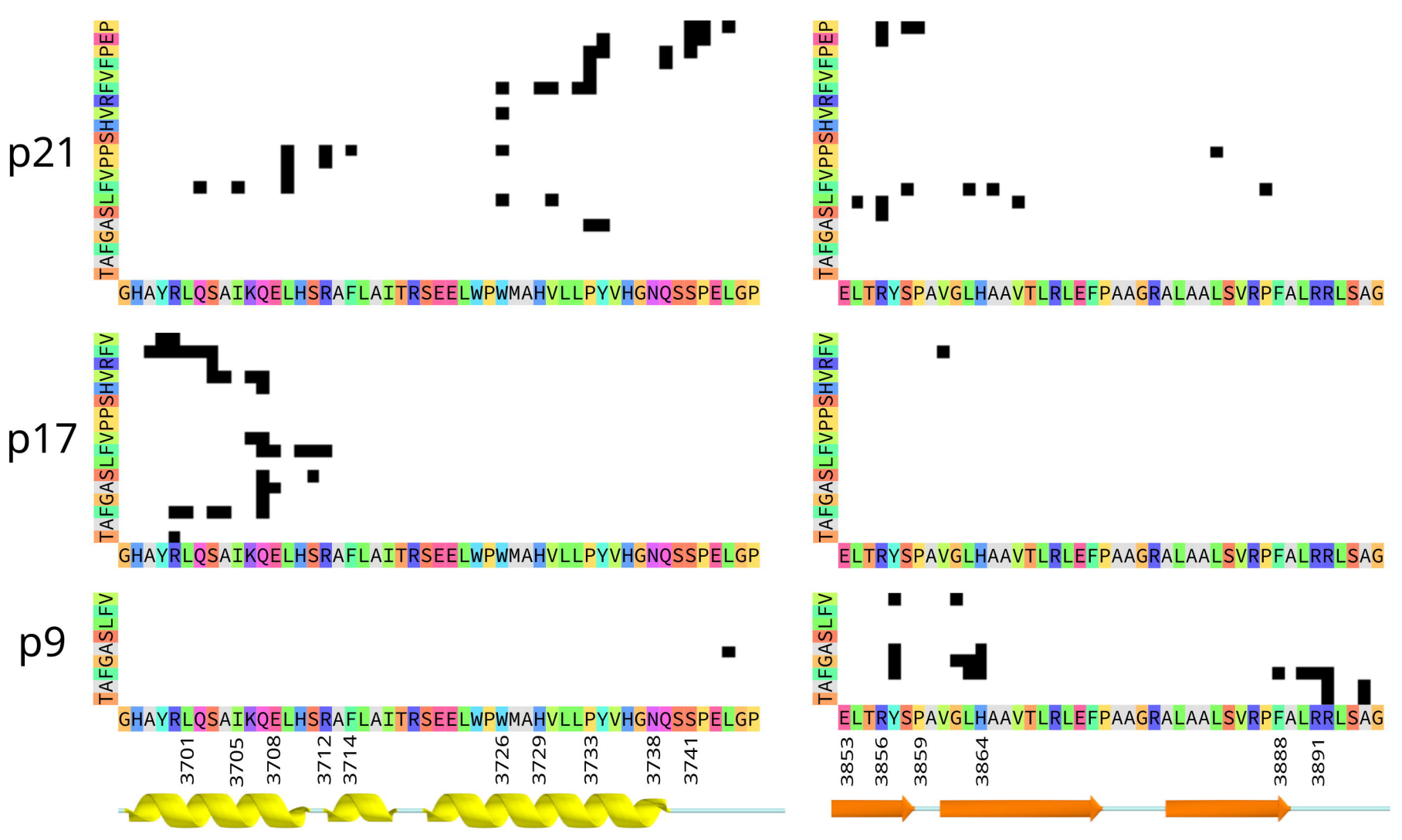

A
R3848 - E4078 Distance (Å)
p21 - Extracellular Contacts
Bound
Free Energy (kcal/mol)
B
TOP
PL
ECL1
S6a
S4
S1
S2
S3
S5
PLAT
C-Tail
C
p21: TAFGASLFVPPSHVRFVFPEP
HPEPDOCK
Pep-GaMD
TOP
PL
ECL1
D
F3714
H3729
S12
V14
V17
R3712
L3732
H13
W3726
P10
P19
E3708
L7
E20
P3733
I3705
N3738
P21
F8
L3701
S3741
L3863
E
PL
R3848
TOP
3.8 Å
E4078

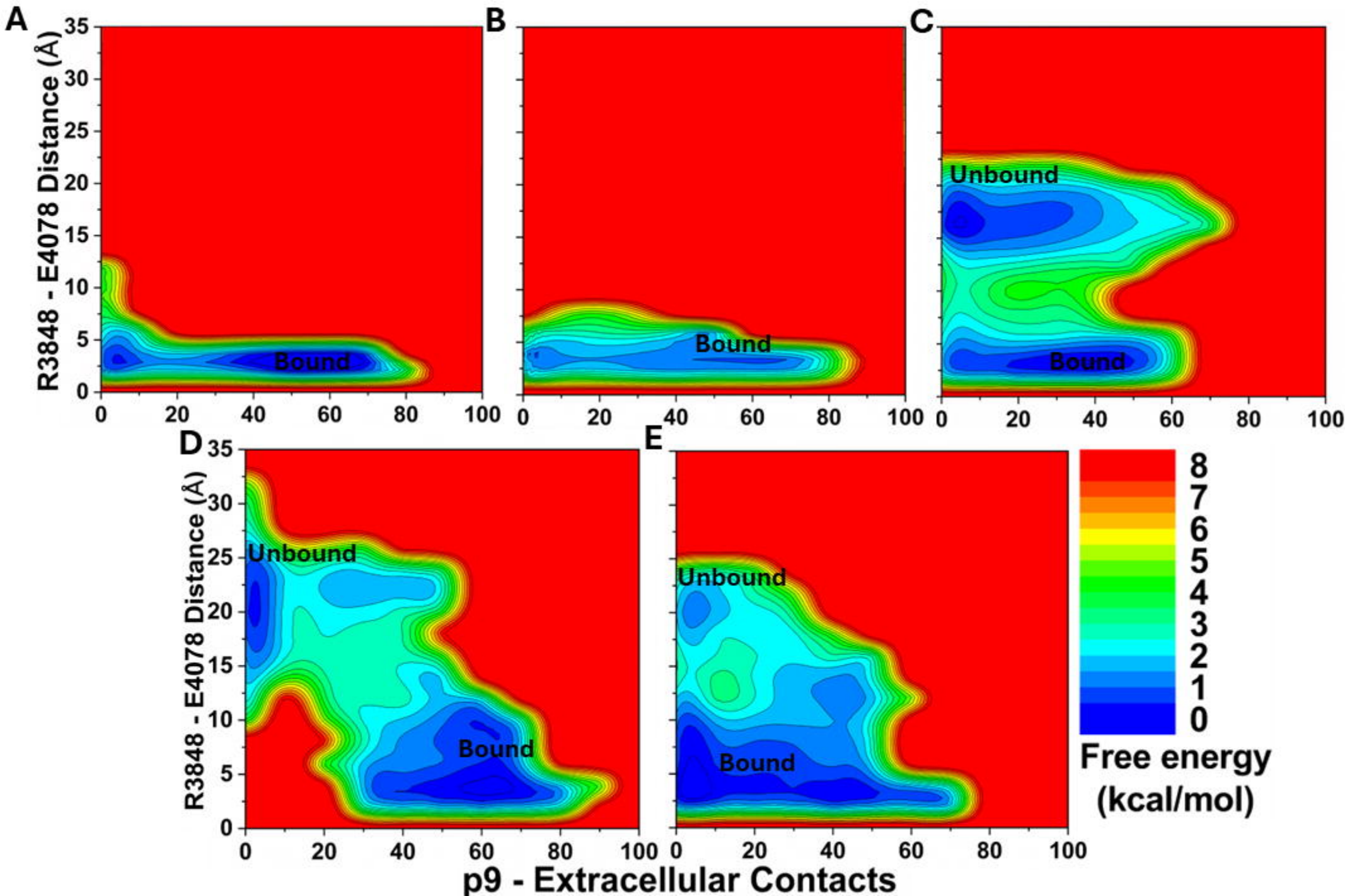

A
F16
P10
G4
A3296
W3298
E20
S3579
T1
W3726
E3743
F16
T1
S3741
T1
Q3821
V3077
N25
T3856
P3859
G4
L3893
Stalk
GAIN
GAIN
Stalk
TM1
TM2 TM3
TM4 TM5
S1
TOP
TOP
S2
S3
S4
B
GAIN
Stalk
TOP
C
T1
S T
W3726
F 0.6 -0.3
Y 2.8 -1.4
W -3.2 1.8
N25
D N S
P3859
V -0.1 -0.8 0.8
I -1.4 1.8
T -1.0 0.8
P 1.6 2.4 -2.4
F16
L F Y K D N T P
E3743
I 1.2 1.0 1.2 0.9 -1.8 -0.5 -1.0 -0.4
L -0.8 -0.9 -1.1 4.2 -0.6 -3.2 2.0
R -0.6 -1.6 -0.6 -2.1 1.2 5.0 -0.8
F16
V I L F W P
A3296
A -2.1 -2.0 -1.8 10.5 -1.7 -1.7
V 1.1 0.4 -1.2 0.1
I 1.0 1.9 2.5 -7.0 2.3 1.0
E20
L K E N Q P
S3579
K 3.7 -0.8 -0.4 -0.9 -0.3
R -1.1 0.4 -2.1 1.6 1.4
E -1.2 4.4 -1.8 2.4 -0.5 -0.7
Q -1.3 2.9 -0.4 -0.5
S 0.7 -0.7 0.2 -1.2 -0.2
T1
A R S T
S3741
L -0.3 0.2 -0.3 1.8
H -0.2 1.6 -0.1 -1.4
N -0.9 2.0 -1.5
S -0.9 0.2 -2.0 3.3
C 2.1 0.2 -0.5 -1.4
G4
G A S T
W3298
I -1.7 -0.7 -0.2 2.5
L -2.8 0.2 2.4
F 2.5 0.2 0.3 -2.3
Y -2.2 2.4 -0.3
W 6.0 -1.1 -1.6 -2.7
V3077
G V I T
T3856
V 1.7 -1.1 -0.6
L -1.4 1.2 -0.3
N 0.3 1.0 0.2
S -0.6 -0.9 3.1 -0.8
T -2.0 4.1 -4.0 2.2

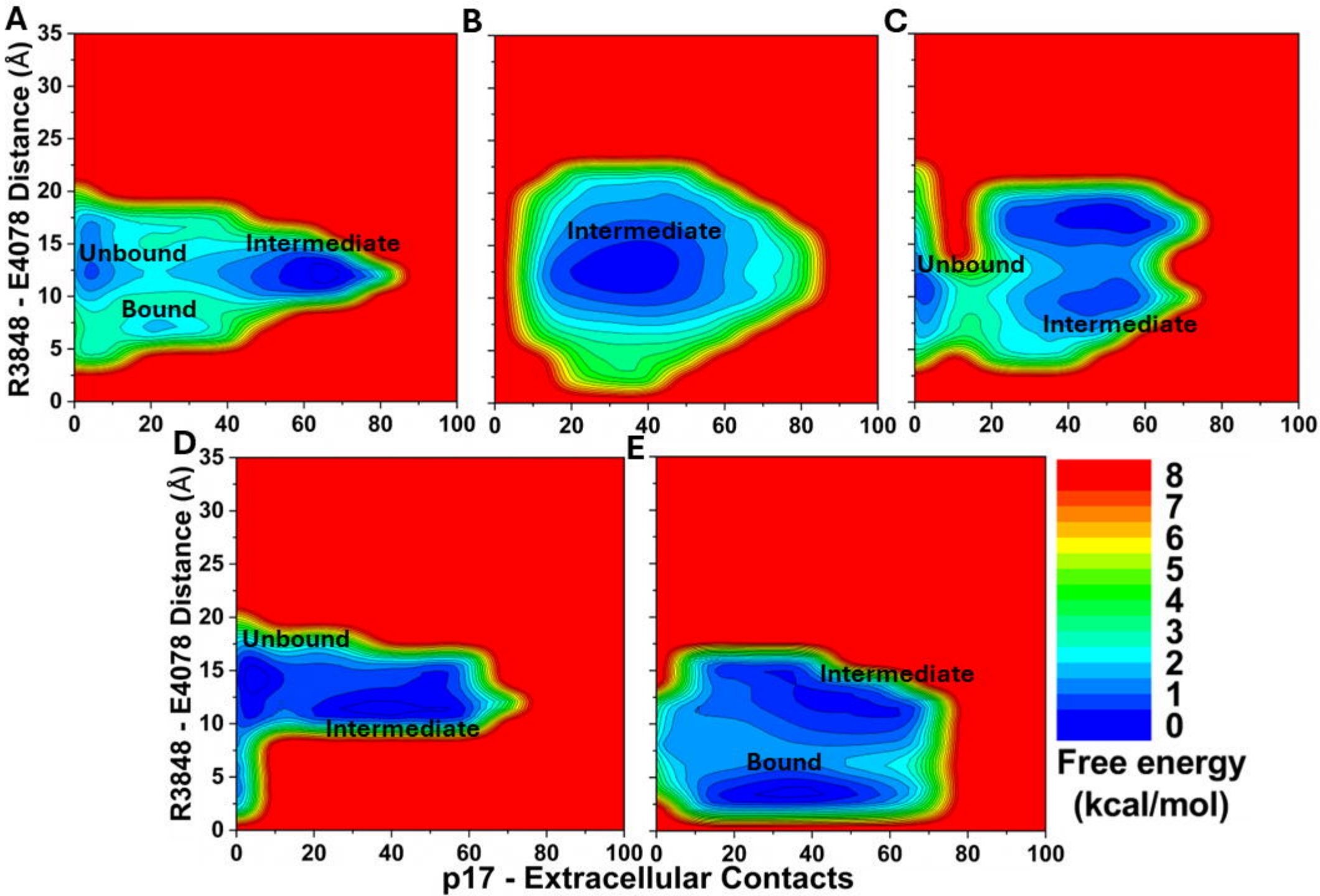

A
+ biotin
- biotin
CTF
CTFΔst
CTF
input
sup
beads
input
sup
beads
input
sup
beads
Ab: A19
Ab: TRAM2
250
150
100
50
37
B
Surface expression levels (+SD)
1.5
1.0
0.5
0.0
CTF
CTFΔst

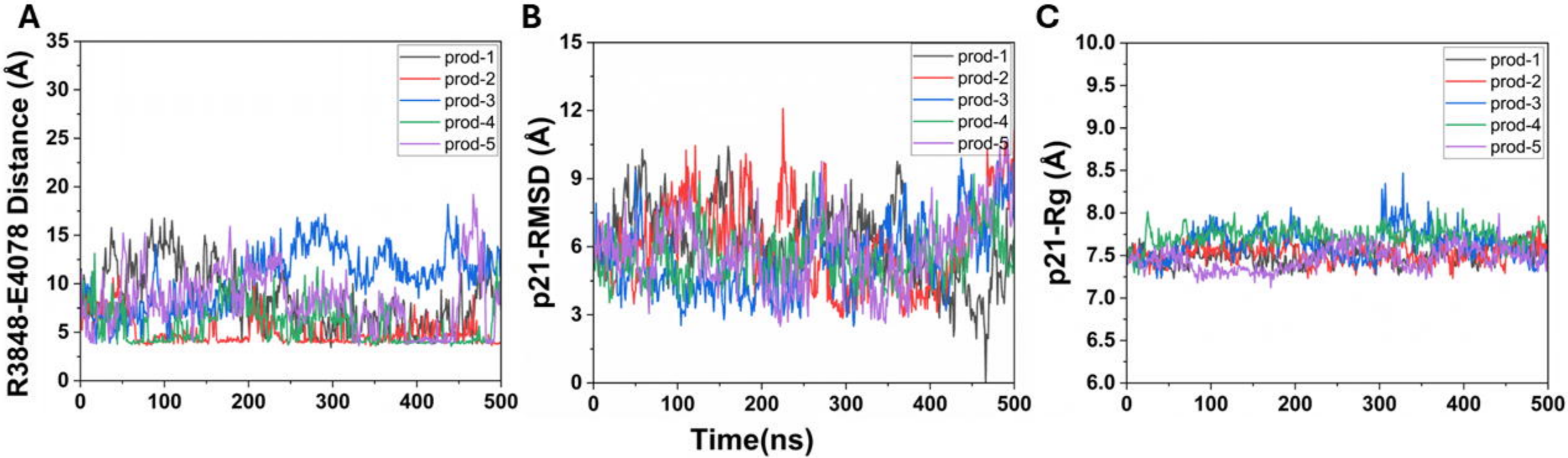

A
Human p17 TAFGASLFVPPSHVRFVGGKKKKK
sol tag LGGKKKKK
B mouse CTFΔst
C human CTFΔst
D human CTFΔst
NFAT-luc activity (mean+/- SD)
100000
80000
60000
40000
20000
0
ev
CTFΔst
control
p17
sol tag
ns
*
**
***
****

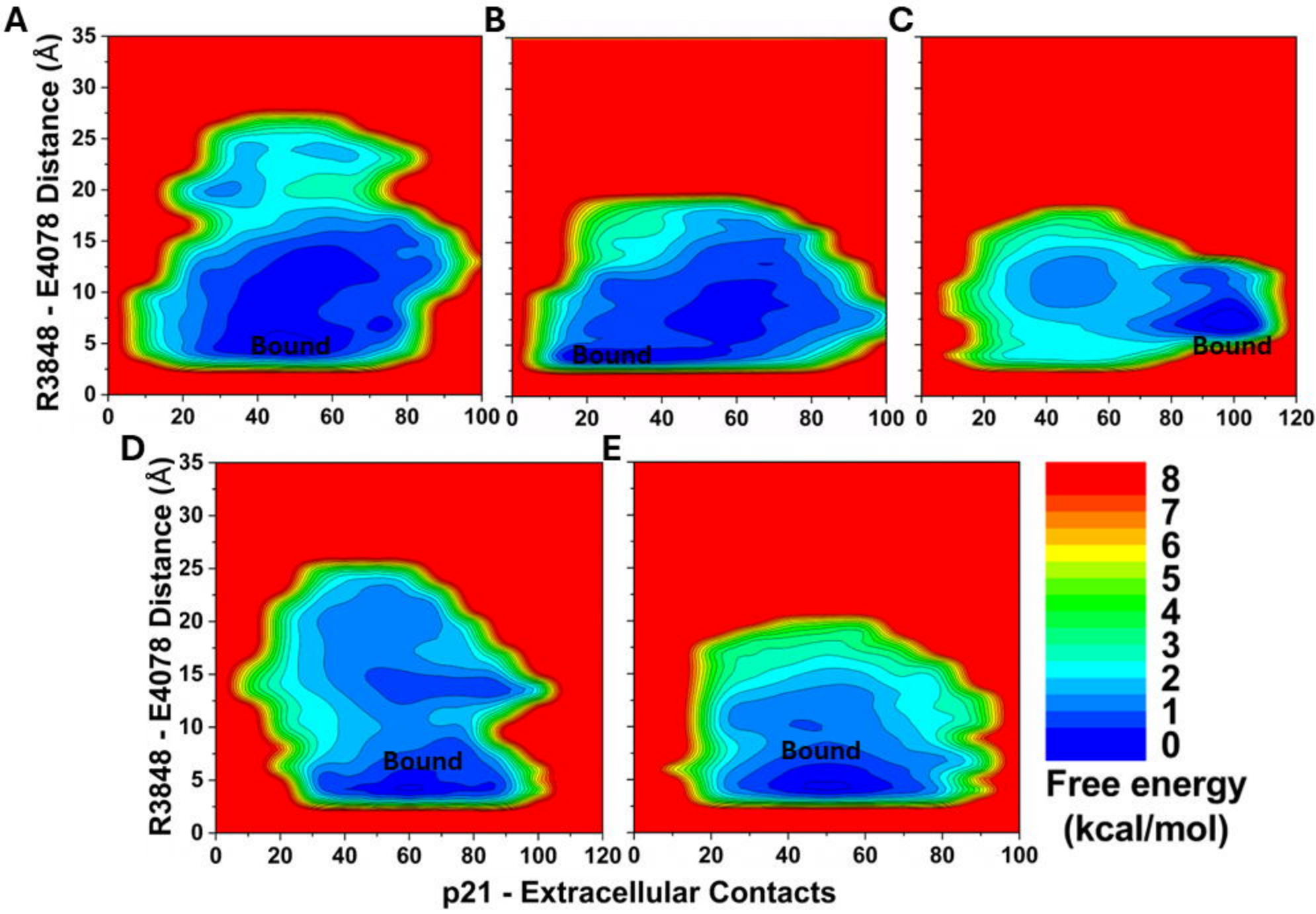

A
p9: TAFGASLFV
p17: TAFGASLFVPPSHVRFV
p21: TAFGASLFVPPSHVRFVFTADVN
TOP
ECL1
PL
Extracellular
S6a
S3
S4
S5
TM3
TM4
S1
S2
S6b
TM5
TM2
TM1
Intracellular
C-Tail
PLAT
B
TOP
ECL1
S3
PL
C

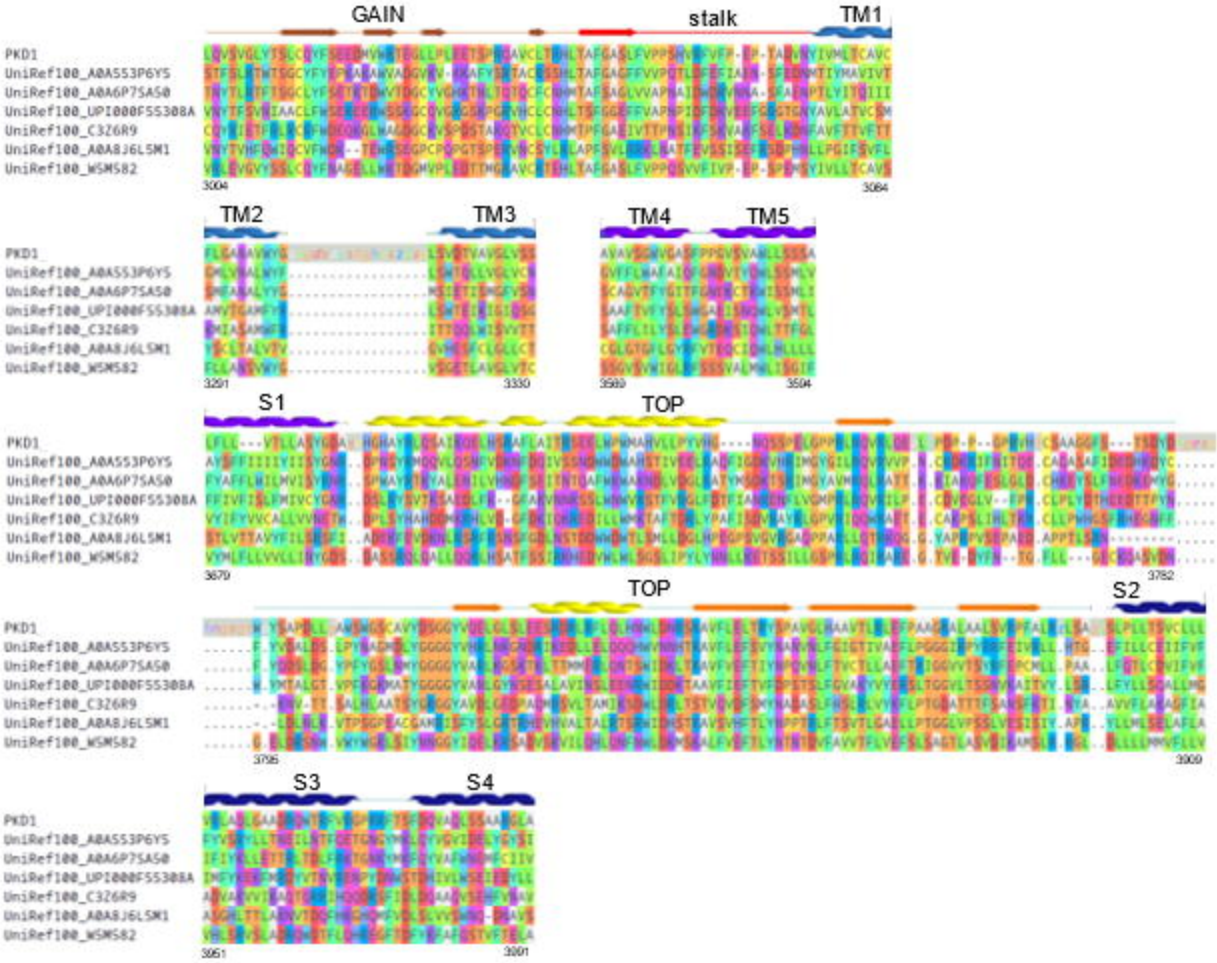

p21
p17
p9
TAFGASLFVPPSHVRFVFPEP
TAFGASLFVPPSHVRFV
TAFGASLFV
GHAYRLQSAIKQELHSRAFLAITRSEELWPWMAHVLLPYVHGNQSSPELGP
ELTRYSPAVGLHAAVTLRLEFPAAGRALAALSVRPFALRRLSAG
3701
3705
3708
3712
3714
3726
3729
3733
3738
3741
3853
3856
3859
3864
3888
3891

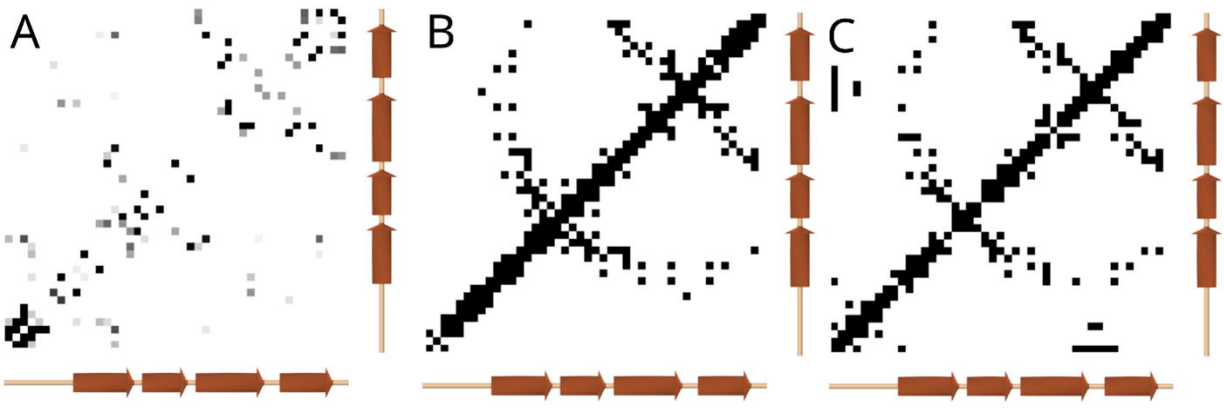

A
B
C
D
E
R3848 - E4078 Distance (Å)
p9 - Extracellular Contacts
Bound
Unbound
Free energy (kcal/mol)
0
1
2
3
4
5
6
7
8

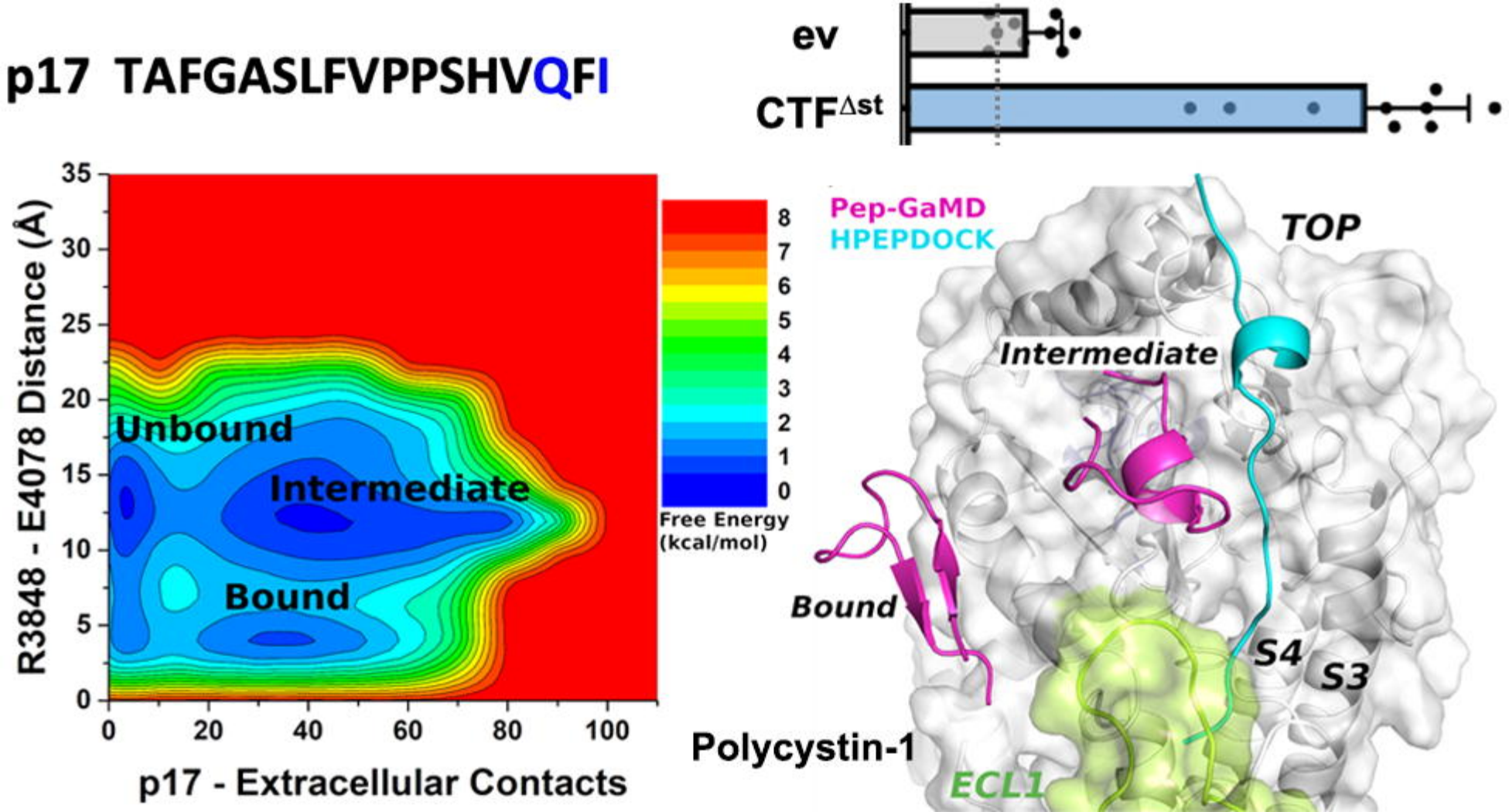

A
B
C
D
E
R3848 - E4078 Distance (Å)
p17 - Extracellular Contacts
Unbound
Intermediate
Bound
Free energy (kcal/mol)
0
1
2
3
4
5
6
7
8


A
R3848-E4078 Distance (Å)
35
30
25
20
15
10
5
0
0
100
200
300
400
500
prod-1
prod-2
prod-3
prod-4
prod-5
B
p21-RMSD (Å)
15
12
9
6
3
0
0
100
200
300
400
500
prod-1
prod-2
prod-3
prod-4
prod-5
C
p21-Rg (Å)
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
0
100
200
300
400
500
prod-1
prod-2
prod-3
prod-4
prod-5
Time(ns)


A
B
C
D
E
R3848 - E4078 Distance (Å)
p21 - Extracellular Contacts
Bound
Free energy (kcal/mol)
0
1
2
3
4
5
6
7
8


GAIN
stalk
TM1
PKD1
UniRef100_A0A553P6Y5
UniRef100_A0A6P75A50
UniRef100_UPI000F55308A
UniRef100_C3Z6R9
UniRef100_A0A8J6L5M1
UniRef100_W5M582
3004
3084
TM2
TM3
TM4
TM5
3291
3330
3589
3594
S1
TOP
3679
3782
TOP
S2
3795
3909
S3
S4
3951
3991


A
B
C


p17 TAFGASLFVPPSHVQFI
ev
CTF$^{\Delta st}$
R3848 - E4078 Distance (Å)
p17 - Extracellular Contacts
Unbound
Intermediate
Bound
Free Energy (kcal/mol)
Pep-GaMD
HPEPDOCK
TOP
Intermediate
Bound
S4
S3
Polycystin-1
ECL1