

## Electronic Supplementary Information

# Binding of the Plant-Derived Toxin Simplexin to Bovine Protein Kinase C: Insights from Molecular Dynamics

Yuchen Zhou,<sup>†</sup> Ras Baizureen Roseli,<sup>†</sup> Natasha L. Hungerford,<sup>‡,†</sup> Mary T. Fletcher,<sup>‡,†</sup>

Diane Ouwerkerk,<sup>§,#</sup> Rosalind A. Gilbert,<sup>§,#</sup> Elizabeth H. Krenske<sup>\*,†</sup>

<sup>†</sup> School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, QLD 4072, Australia

<sup>‡</sup> Queensland Alliance for Agriculture and Food Innovation (QAAFI), The University of Queensland, Health and Food Sciences Precinct, Coopers Plains, QLD 4108, Australia

<sup>§</sup> Agri-Science Queensland, Department of Agriculture and Fisheries (DAF), EcoSciences Precinct, Dutton Park, Queensland, 4102, Australia

<sup>#</sup> Queensland Alliance for Agriculture and Food Innovation (QAAFI), The University of Queensland, Queensland Bioscience Precinct, St Lucia, QLD 4072, Australia

\* Corresponding author. Email: e.krenske@uq.edu.au. Telephone: +61 7 3365 4632.

# Table of Contents

## Supplementary Methods

Structure Preparation	S3
Ligand Parameters	S5
Membrane Force Field and Ion Parameters	S5
System Construction and Simulation Details	S6
Binding Enthalpy Calculation	S8

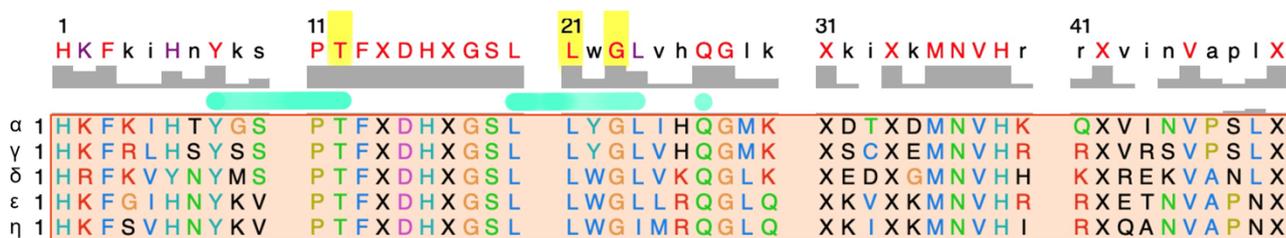
## Supplementary Figures

<b>Simulations of simplexin and its analogues with PKC<math>\alpha</math>-C1B in water:</b>	S12
RMSD of the heavy atoms of simplexin and its analogues	S13
RMSD of the backbone of PKC $\alpha$ -C1B complexed with simplexin and its analogues	S14
<b>Simulations of simplexin and its analogues with PKC<math>\alpha</math>-C1B in membrane:</b>	S15
Hydrogen bond occupancies of simplexin in PKC $\alpha$ -C1B	S16
RMSD of the heavy atoms of simplexin and its analogues	S17
RMSD of the backbone of PKC $\alpha$ -C1B complexed with simplexin and its analogues	S18
Number of hydrogen bonds between PKC $\alpha$ -C1B and simplexin analogues	S19
Enthalpy decomposition of PKC $\alpha$ -C1B complexed with simplexin and its analogues	S20
<b>Simulations of DAG and phorbol esters with PKC<math>\alpha</math>-C1B in water:</b>	S21
RMSD of the heavy atoms of ligands	S22
RMSD of the backbone of PKC $\alpha$ -C1B complexed with different ligands	S23
Number of hydrogen bonds between PKC $\alpha$ -C1B and ligands	S24
Enthalpy decomposition of PKC $\alpha$ -C1B complexed with different ligands	S25
<b>Simulations of DAG and phorbol esters with PKC<math>\alpha</math>-C1B in membrane:</b>	S26
RMSD of the heavy atoms of ligands vs. time	S27
RMSD of the backbone of PKC $\alpha$ -C1B complexed with different ligands	S28
Number of hydrogen bonds between PKC $\alpha$ -C1B and ligands	S29
Enthalpy decomposition of PKC $\alpha$ -C1B complexed with different ligands	S30
<b>Quantum mechanical calculations</b>	S31
<b>References</b>	S35

## Supplementary Methods

### Structure Preparation

Structures of PKC $\alpha$  (Uniprot ID: P04409), PKC $\gamma$  (Uniprot ID: P05128), PKC $\epsilon$  (Uniprot ID: A0A4W2CHB2), and PKC $\eta$  (Uniprot ID: F1MY82) were downloaded from the *AlphaFold Protein Structure Database*.<sup>1</sup> The source organism for PKC $\alpha$ , PKC $\gamma$ , and PKC $\eta$  was *Bos taurus*, while the source organism for PKC $\epsilon$  was a *Bos indicus x Bos taurus* hybrid. These PKC isoforms were chosen due to their sequence differences in the C1B domain. The C1B domains of the PKC isoforms were determined by sequence alignment with *Mus musculus* PKC $\delta$ -C1B (PDB ID: 1PTR).<sup>2</sup> The residues of C1B domains were renumbered starting from the common histidine residue as shown in **Figure S1**.

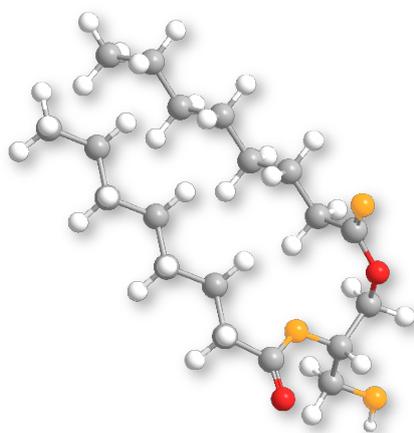


**Figure S1.** Sequence alignment of PKC-C1B isoforms utilized in this research. The GLY51 residue of the C1B domain is not shown. Residues that form major hydrogen bonds with simplexin are highlighted in yellow and the residues that form the activator binding site are highlighted in green.  $\delta$ : mouse PKC $\delta$ -C1B (1PTR);  $\alpha$ ,  $\gamma$ ,  $\epsilon$  and  $\eta$ : bovine PKC-C1Bs. X represents a deprotonated cysteine residue. The original residue number of these isoforms starts from  $\alpha$ : 102,  $\gamma$ : 102,  $\delta$ : 231,  $\epsilon$ : 243,  $\eta$ : 246.

Although the amino acid length of mouse PKC $\delta$ -C1B (PDB ID: 1PTR) is 50, an additional glycine residue was included at the C-terminus of the bovine C1B domains. This residue (GLY51) is highly conserved and was included because truncation at the 50<sup>th</sup> residue would have terminated the sequence at a deprotonated cysteine, which is not supported for terminal patching in *CHARMM-GUI*.<sup>3-5</sup> To prepare the resulting structures for simulations (e.g. adding any missing atoms), the PKC-C1B isoforms were then submitted to the PDB structure processor of Yinfo Cloud Computing Platform.<sup>6</sup> The processed 3D structures were subsequently aligned with the mouse PKC $\delta$ -C1B to obtain the correct ligand-binding posture, and the zinc ions were manually inserted into the zinc fingers of the proteins.

Simplexin was docked into the binding site of PKC $\delta$ -C1B using *AutoDock Vina*<sup>7</sup> to obtain a bound pose. In order to include the known binding site of P13A<sup>2</sup> (which is presumed also to represent the binding site for simplexin and the other ligands studied here) in the grid box, the centre of the box was set to 14.938, 19.797 and 24.716 Å (x, y, and z), respectively, and the x, y, and z dimensions were set to 24, 27 and 17.25 Å, respectively. The exhaustiveness was set to 40, and the top 20 conformations with the strongest affinities were saved.

To generate the ligand parameters and topologies required for MD simulations, the 3D structures of the ligands were initially optimized using B3LYP-D3(BJ)/6-311G(d,p) in *Gaussian16*.<sup>8-12</sup> To preserve the ligand binding pose and coordinates, the keyword “nosymm” was used. (The “nosymm” keyword was used to prevent *Gaussian* from reorientating the input coordinates into a “standard orientation” and producing an unwanted change of the ligand’s bound pose.) During the optimization of simplexin and its analogues, the dihedral angles of the hydroxyl C–O bonds were held fixed. During the optimization of DAG, three of the oxygens were held fixed in space, as highlighted in orange in **Figure S2**, to preserve the binding pose found in the crystal structure reported by Igumenova et al. (PDB ID: 7L92)<sup>13</sup>.



**Figure S2.** Structure of di-octanoyl-sn-1,2-glycerol (DAG) showing, in orange, the oxygen atoms that were frozen in geometry optimization.

## Ligand Parameters

Using the optimized structures of ligands obtained above, electrostatic potential calculations were performed at the B3LYP-D3(BJ)/def2-TZVP<sup>14</sup> level of theory in *Gaussian16* using the keywords “pop=mk” and “iop=(6/33=2, 6/42=6)”. Electrostatic potential calculations were performed in the gas phase and in aqueous phase separately, using the PCM implicit solvent model<sup>15</sup> for the latter, and using the keyword “nosymm” to preserve the ligand binding pose. The gas phase and aqueous outputs were independently processed using the *Antechamber* module of *AmberTools22*<sup>16</sup> to fit the restrained electrostatic potential (RESP) charges. *GROMACS*<sup>17</sup> compatible topology files, coordinates restraint files and *gro* formatted coordinates files were generated using the *parmchk2*, *tleap* modules of *AmberTools22* and the *ACPYPE*<sup>18</sup> program.

## Membrane Force Field and Ion Parameters

For simulations involving membrane, the membrane was simulated as a POPS bilayer, using the Lipid21 forcefield.<sup>19</sup> POPS is an anionic phospholipid. It has been found that simulations of anionic phospholipids using Lipid21 are sensitive to both cation type and force field parameters.<sup>19</sup> Previous work has explored the effect of dispersion corrections, cut-off distances, cation parameters on membrane properties such as area per lipid.<sup>19-21</sup> For anionic POPS and POPG lipids, the utilization of ion parameters from Amber14SB, which were generated by Joung and Cheatham<sup>22</sup>, resulted in lipid condensing and low area per lipid.<sup>19</sup> However, simulations of potassium cations with Åqvist<sup>23</sup> ion parameters from Amber99SB allowed a POPS membrane to demonstrate a more realistic area per lipid value ( $61.78 \pm 0.34 \text{ \AA}^2$ , close to the experimental value  $62.7 \text{ \AA}^2$ )<sup>24</sup>. In that work, a cut-off distance of 1.0 nm was used for both Lennard-Jones and Coulombic interactions, and a dispersion correction was used for both energy and pressure.

In the present work, K<sup>+</sup> ions with Åqvist ion parameters were used in the membrane simulations. To explore the influence of the cut-off distance and dispersion correction, MD simulations were performed on a system comprising the PKC $\delta$ -C1B–simplexin complex embedded in a POPS bilayer surrounded by water and neutralizing K<sup>+</sup> ions. Four simulations were performed, varying the cut-off distance and whether or not a dispersion correction was applied to energy and pressure. Following a 100 ns equilibration, a 300 ns production simulation was performed. The area per lipid in the upper

leaflet of the membrane (corresponding to the extracellular side) was calculated under each set of conditions. Results are given in **Table S1**.

**Table S1.** Area per lipid ( $\text{\AA}^2$ ) obtained under different simulation conditions.

Cut-off (nm)	Dispersion correction	
	ON	Off
1.0	60.2 (1.1)	60.8 (1.1)
1.4	58.5 (1.2)	58.7 (0.9)

<sup>a</sup> Values are the average  $\pm$  standard deviation from 30,000 snapshots taken over a 300 ns simulation.

The results show that the area per lipid decreases as the cut-off distance increases or when the dispersion correction is utilized, consistent with previous research.<sup>19-21</sup> The simulation with a 1.0 nm cut-off distance and no dispersion correction gave the area per lipid closest to the experimental value of  $62.7 \text{\AA}^2$ .<sup>24</sup> However, inclusion of a dispersion correction also gave a result considered acceptable. Because we aimed to perform simulations in both an isotropic water phase and in a heterogeneous membrane system, we decided that inclusion of a dispersion correction (with a cut-off distance of 1.0 nm) represented the best compromise between accuracy and consistency.

### System Construction and Simulation Details

For simulations of protein–ligand complexes in water, the following protocol was used for system preparation. The repaired and aligned protein structure was processed with the *pdb2gmx* module of *GROMACS*, generating the protein topology file as well as a *gro* formatted protein structure. The protein structure was then merged with the ligand structure file containing information about the binding pose. The ligand topology along with the necessary restraints were incorporated into the protein topology file. The protein–ligand complex was placed in a dodecahedral box, ensuring a minimum distance of 1 nm between the complex and the box edge. Finally, the complex was solvated with TIP3P<sup>25</sup> water and neutralized with potassium or chloride ions as appropriate.

The Amber14SB force field "amber14sb\_parmbsc1", downloaded from the official website of *GROMACS*<sup>26</sup>, was used for the protein (including zinc ions). Energy minimization was performed using the steepest descent algorithm with a convergence criterion of  $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ . A 100 ps

NVT equilibration was then performed, followed by a 100 ps NPT equilibration, in both cases applying position restraints to the protein and ligand with a force constant of  $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ . Next, a 200 ns production molecular dynamics simulation was conducted. The V-rescale thermostat was employed with a time constant of 0.1 ps and a reference temperature of 300 K. To facilitate temperature coupling, the system was divided into two groups: protein+ligand and water+ions. The C-rescale barostat and Parrinello–Rahman barostat were used in the NPT and production simulations. Pressure was controlled isotropically with a time constant of 2 ps and a reference pressure of 1 bar. Constraints on bonds containing hydrogen atoms were imposed using the LINC algorithm<sup>27</sup>. Particle Mesh Ewald (PME)<sup>28</sup> was applied for long-range electrostatics with a cut-off distance of 1.0 nm, and the same cut-off distance was also utilized for van der Waals interaction and short-range neighbour list. All simulations were performed using a timestep of 2 fs. A dispersion correction was applied to account for energy and pressure effects. The Verlet method was employed for the cut-off scheme of van der Waals interactions, and the potential-shift-Verlet approach was used as a modifier for both van der Waals and electrostatic interactions. Trajectories were saved every 10 ps. Velocities of atoms were generated randomly at the beginning of the NVT equilibration from a Maxwell–Boltzmann distribution, with a reference temperature of 300 K.

For simulations in which dihedral restraints were applied, the reference dihedral angles of the alkyl chain were determined from the *GROMACS* energy-minimized structures of the simplexin analogues. A force constant of  $100 \text{ kJ mol}^{-1} \text{ rad}^{-2}$  was applied to restrict the motion of the dihedral angles whenever they deviated by more than  $5^\circ$  from the reference values.

For simulations involving membrane, the following protocol was used for system preparation. The membrane was constructed using *Bilayer Builder* of the input generator module in *CHARMM-GUI*. Firstly, the insertion posture of mouse PKC $\delta$ -C1B (PDB ID: 1PTR) was obtained from the *OPM* database<sup>29</sup>, which contains information about the embedding depth and angle of PKC $\delta$ -C1B in the membrane. Next, the structures of the bovine PKC-C1B isoforms including PKC $\alpha$ -C1B, PKC $\gamma$ -C1B, PKC $\epsilon$ -C1B, and PKC $\eta$ -C1B, including any bound ligands, were aligned to the PKC $\delta$ -C1B structure from the *OPM* database. The protein was uploaded to *CHARMM-GUI* and processed with the following protocol. Both the C and N termini were patched with standard CO<sub>2</sub>H/NH<sub>2</sub> termini. The “PDB orientation” was selected to position the protein in the membrane based on the posture information from the *OPM*. The resulting protein was placed into a hexagonal prismatic box. The length of the box on X and Y axis was initially set to 64 Å, and the thickness of water layer was 10 Å. Next, a bilayer

membrane composed of POPS was generated based on the box size. The system was subsequently neutralized with potassium cations, and finally the *GROMACS* input files specifying the Amber14SB and Lipid21 force fields were generated. To properly incorporate the ligand into the system, the information from the “[atomtypes]” section of the ligand topology file was moved to the same section of the “*forcefield.itp*” file generated by *CHARMM-GUI*. Then the ligand structure was merged with the system generated by *CHARMM-GUI*. Finally, the Åqvist potassium ion parameters were specified.

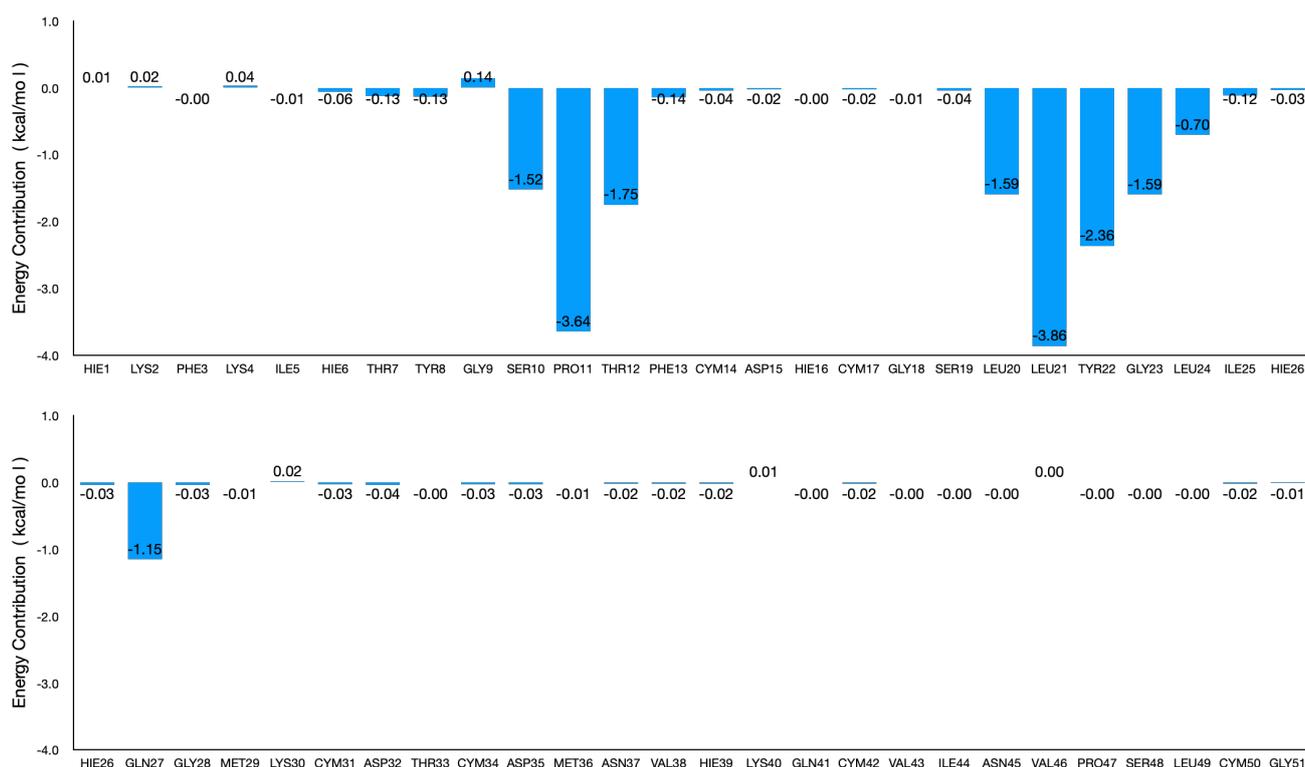
The simulation details of the membrane systems were mostly the same as for the simulations performed in water, but had the following differences. Firstly, after energy minimization, the equilibration stage was divided into six steps including two NVT steps followed by four NPT steps. For the first two NVT equilibrations, a timestep of 1 fs was employed; the duration of both simulations was 125 ps. The first NPT equilibration used a timestep of 1 fs and had a duration of 125ps, while the remaining three NPT equilibrations used a timestep of 2 fs and had durations of 500 ps. Over the six equilibrations, position and dihedral restraints for the protein and membrane were gradually relaxed following the default protocol of *CHARMM-GUI*, while the position restraints of ligand were maintained. The Parrinello–Rahman barostat was utilized throughout the simulations, with the system separated into three groups for pressure coupling, namely, solutes (protein, zinc ions, ligand), membrane, and solvent (water and potassium ions). The pressure couple type was semi-isotropic since the system was heterogeneous. A time constant of 5 ps was employed for pressure coupling, and a time constant of 1 ps was employed for temperature coupling. Finally, center-of-mass motions were removed between solvents and the solute plus membrane. After the equilibration steps, a production MD simulation of 200 ns was conducted.

### **Binding Enthalpy Calculations**

Binding enthalpy calculations followed the protocol described below. After the 200 ns production MD simulation, the last frame was utilized as the starting structure for subsequent short simulations. Then the starting structure underwent NVT and NPT equilibration (as described above for aqueous and membrane simulations), followed by a MD simulation of 5 ns. From the last 4 ns of the simulation, 400 frames were extracted at intervals of 10 ps. These frames were then utilized to calculate the binding enthalpy using *gmx\_MMPBSA*<sup>30</sup>. This process was repeated 50 times to obtain

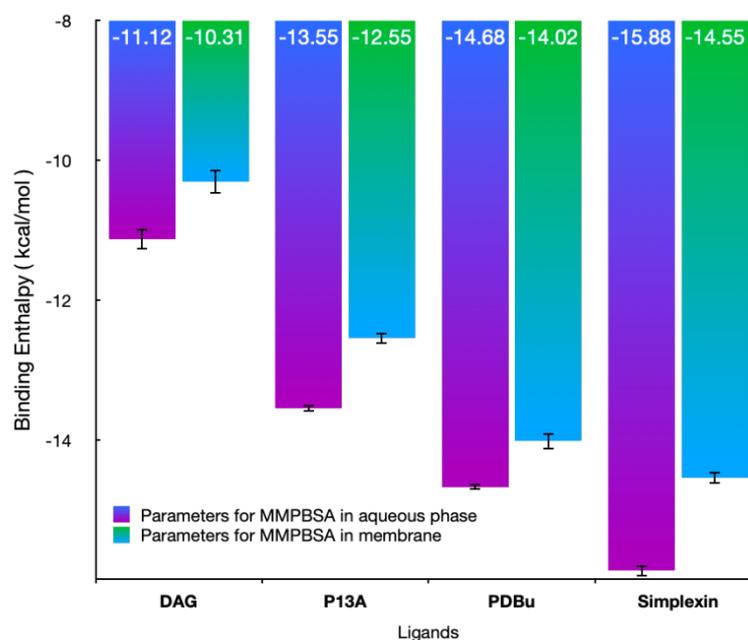
independent replicates. As the velocities were randomly generated during the initial NVT equilibration, each simulation can be considered statistically independent. Consequently, the final binding enthalpies are determined by averaging the MMPBSA results from each of the 50 replicates.

Per-residue enthalpy decomposition analyses were performed on the protein–ligand complexes. The decomposition analyses were performed on the entire protein–ligand complexes, but typically only the results for residues lying within 6 Å of the ligand are shown in our plots. The 6 Å value was sufficient to capture all the important interactions, as illustrated by **Figure S3**, which shows the binding enthalpy decomposition for all 51 residues of the simplexin–PKC $\alpha$ -C1B complex in water. Occasionally, the residues PHE13 or ILE25 also appeared in the list of interactions. These two residues are, however, omitted from our discussion as they contributed in only a minor way (e.g. the absolute value of their enthalpy contribution was  $\leq 0.2$  kcal/mol, or their interactions were detected in only a small number of the 50 replicates).



**Figure S3.** Per-residue decomposition analysis of the binding enthalpies of simplexin complexed with PKC $\alpha$ -C1B in water, showing all 51 amino acid residues.

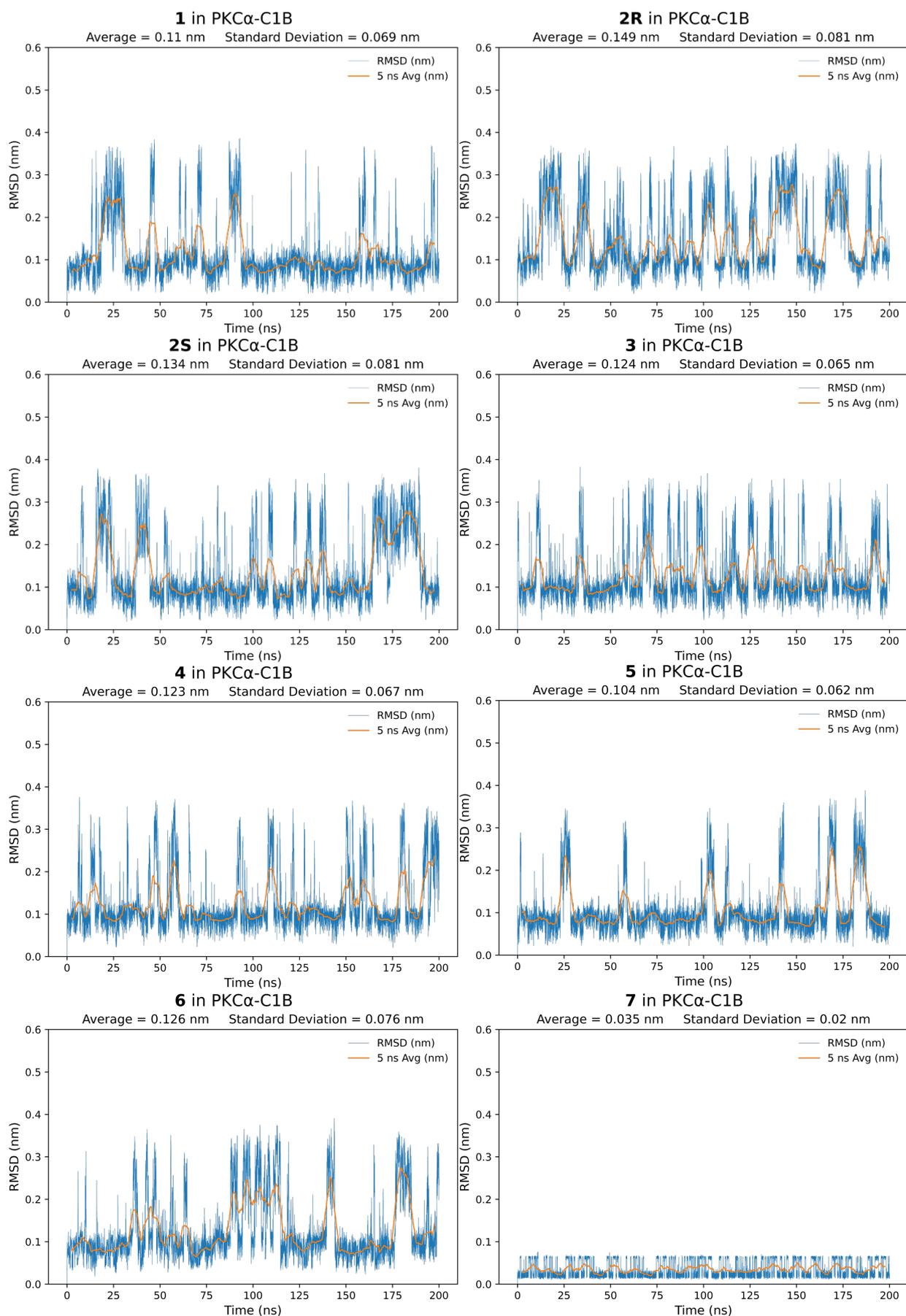
The parameters used with *gmx\_MMPBSA* are as follows. For the binding enthalpy calculation of systems simulated in aqueous phase, default Poisson–Boltzmann (PB) parameters were used, with the exception of setting the atomic radii to 0. This adjustment was made because GAFF<sup>31</sup> atom types were used for ligands. The temperature was set to 300 K. Different parameters were used for calculations of membrane systems. An implicit membrane model was used with a dielectric constant set to 7 F/m. The dielectric constants for the protein and water were set to 80 F/m and 1 F/m, respectively. The "ipb = 1" keyword was employed for the dielectric interface, as it is the only method supported in implicit membrane calculations. Instead of the default (eneopt = 2) method used in the aqueous phase simulations, the Particle-Particle Particle-Mesh (P3M) method<sup>32</sup> was utilized in the membrane simulations to compute the total electrostatic energy and forces. This difference means that the results of the MMPBSA calculations on the membrane-embedded systems are not directly comparable to those of the aqueous-phase systems; a systematic deviation between the two systems is present, amounting to ca. 1 kcal/mol. To quantify the deviation, we conducted several sets of calculations, as follows: (i) based on the trajectories from simulations in the aqueous phase, calculations were performed employing different MMPBSA parameters; and (ii) based on the trajectories from the membrane simulations, the implicit membrane model was turned off by setting the membrane thickness to 0 and moving the membrane 100 Å away from the protein. This configuration ensured that the protein and ligand were only surrounded by implicit water solvent. Comparison among the results of these calculations revealed the effects of different parameters, which are displayed in **Figure S3**. The four activators of PKC demonstrate relatively consistent differences in the calculated binding enthalpies between the two different parameter settings. Compared with the calculation with the parameters used in aqueous phase, the use of the parameters from the calculations involving membrane resulted in an underestimation (less negative) of the binding enthalpy by 0.7–1.3 kcal/mol. On this basis, we identify a ca. 1 kcal/mol systematic underestimation of binding affinity for the membrane simulations.



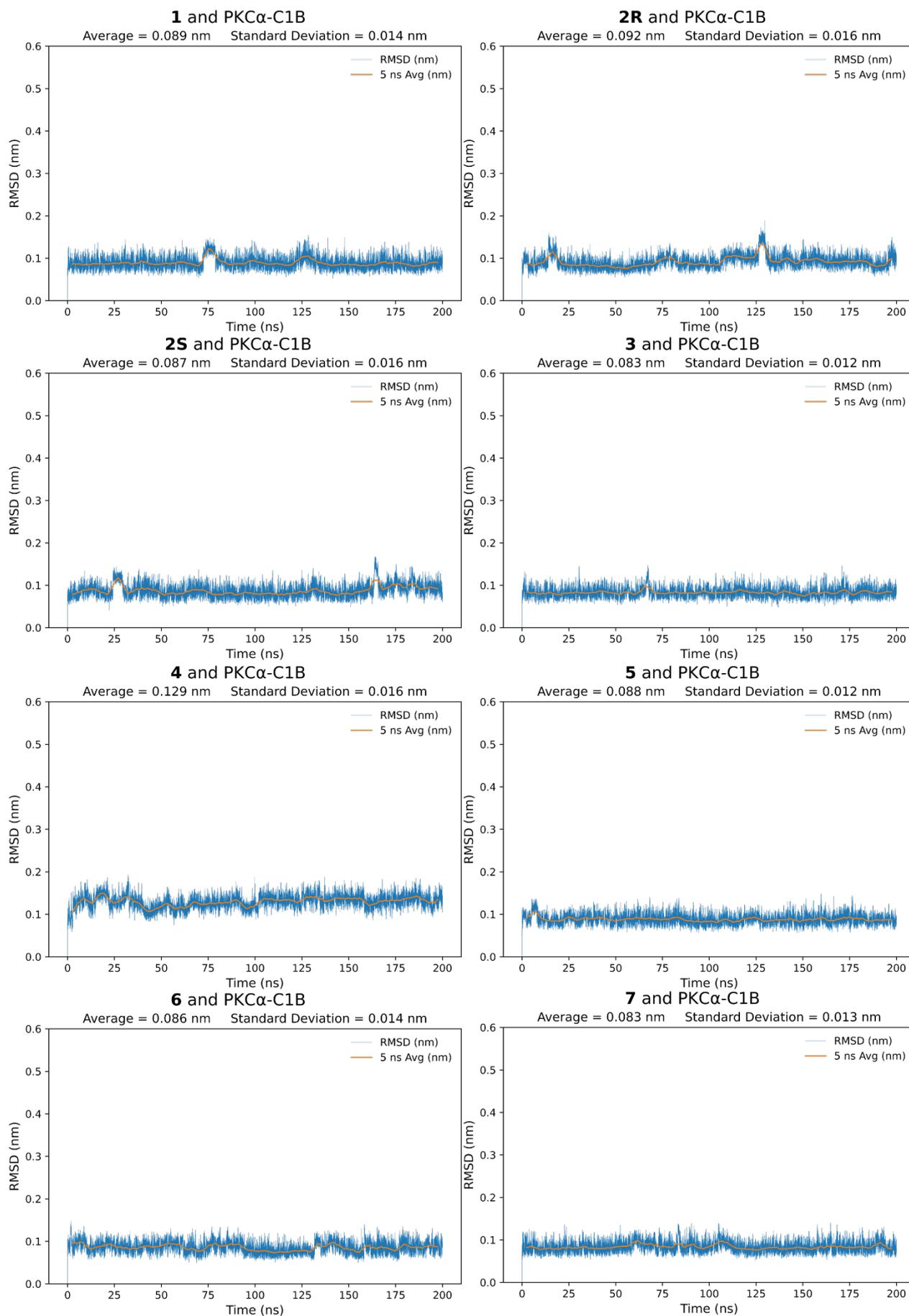
**Figure S4.** Binding enthalpies of PKC $\alpha$ -C1B with four different ligands, calculated using different settings. Each ligand was calculated twice using the same simulation trajectory but varying the parameters (aqueous vs membrane parameters) for the binding enthalpy calculation.

The membrane thickness and the location of the membrane centre are essential parameters in the binding enthalpy calculation. Since these parameters can vary in different MD replicates, they were independently calculated for each replicate and then incorporated into the input file of *gmx\_MMPBSA* using a Python script. The thickness of the membrane was determined by measuring the distance between the two phosphorus planes of the upper (extracellular) and lower (intracellular) leaflets. The centre of the membrane was defined as the average height of these two phosphorus planes. The phosphorus plane was computed by averaging the Z coordinates of their respective phosphorus atoms. The protein embedding depth was also calculated. The protein embedding depth was defined as the distance between the most deeply inserted atom of the protein and the phosphorus plane of the lower leaflet. After the replicate simulation of 5 ns, 500 frames from the trajectory were analyzed to calculate the geometrical features described above. Each frame was individually examined, and the final values were obtained by averaging the results of each frame. The *MDAnalysis*<sup>33</sup> module was employed for analysis of membrane thickness and insertion depth.

**Supplementary figures for simulations of simplexin and its analogues with PKC $\alpha$ -C1B in water**

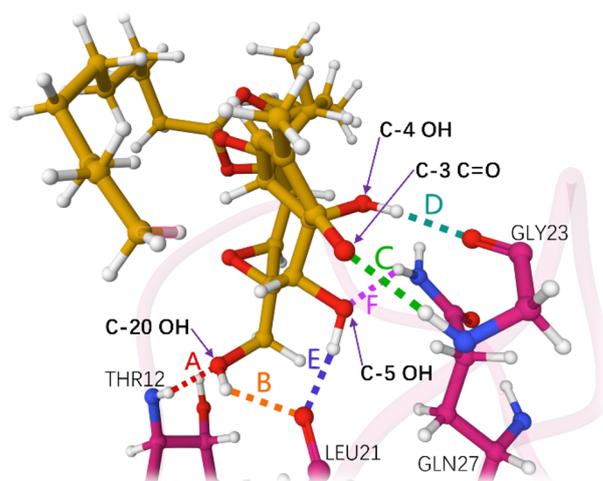


**Figure S5.** RMSD of heavy atoms of simplexin and its analogues complexed with PKC $\alpha$ -C1B in water.



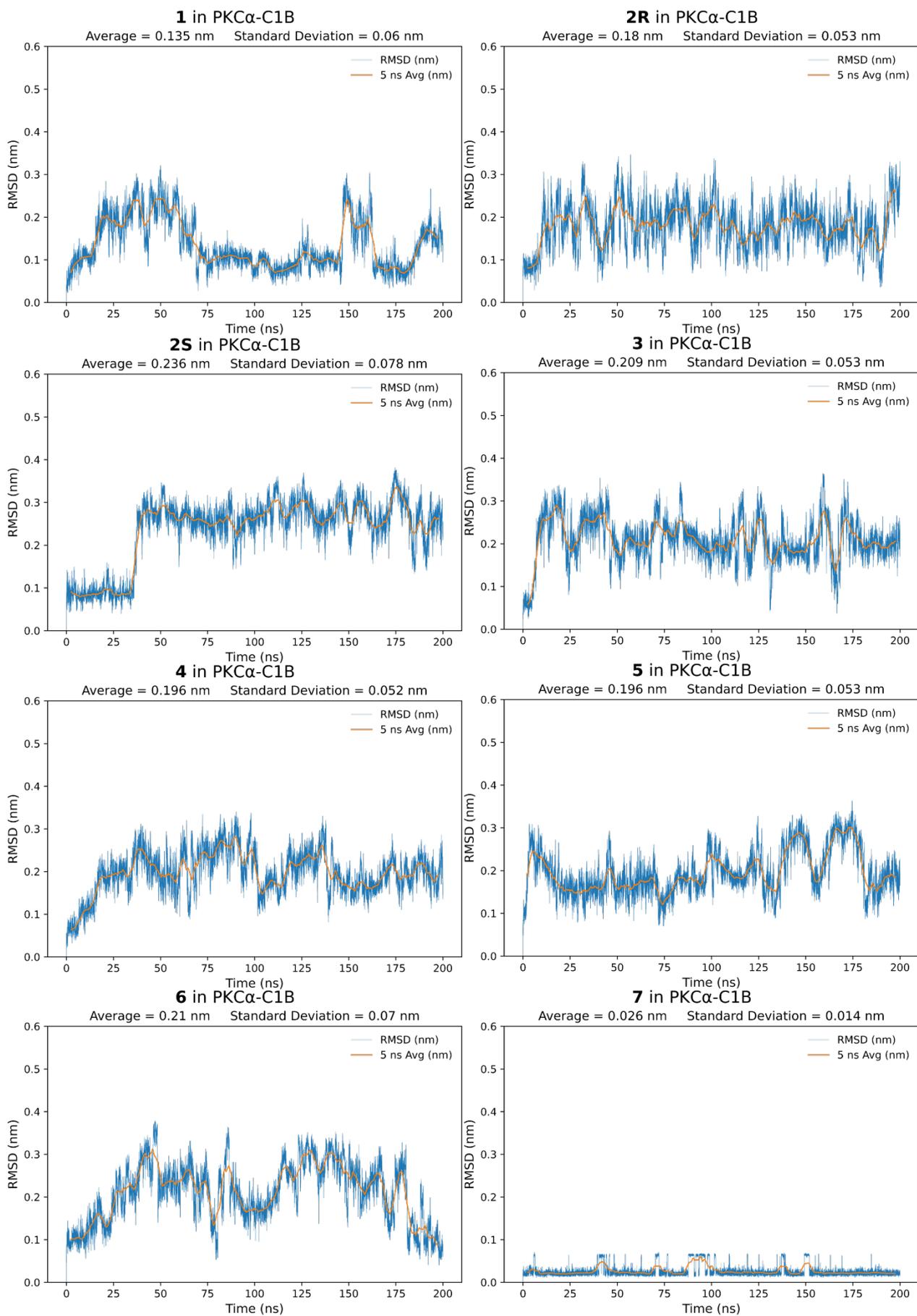
**Figure S6.** RMSD of backbone of PKC $\alpha$ -C1B complexed with simplexin and its analogues in water.

**Supplementary figures for simulations of simplexin and its analogues with PKC $\alpha$ -C1B in membrane**

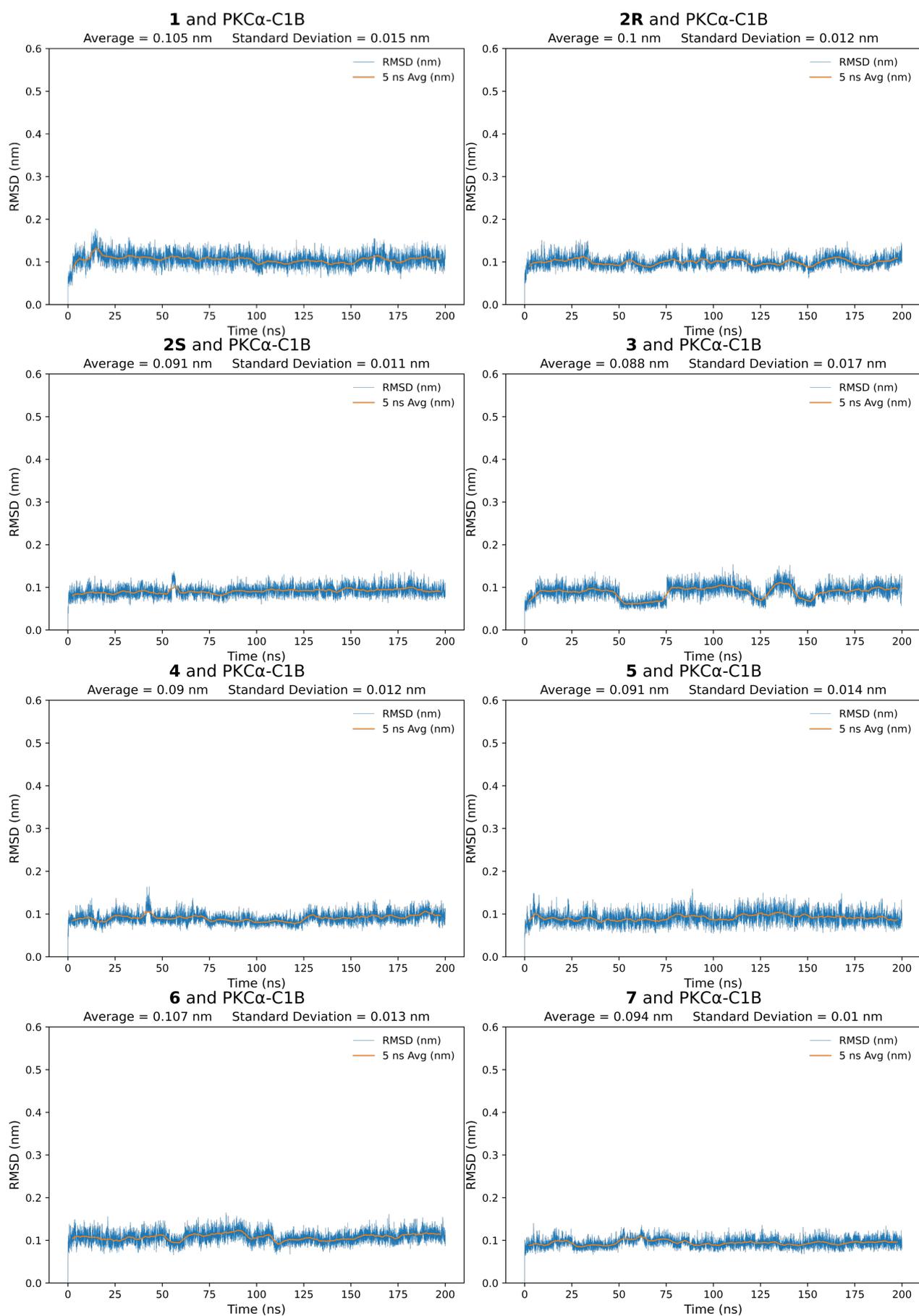


	A	B	C	D	E	F
Aqueous	89.6%	90.7%	88.7%	92.1%	96.1%	14.4%
Membrane	91.9%	91.9%	93.5%	98.5%	96.7%	32.9%

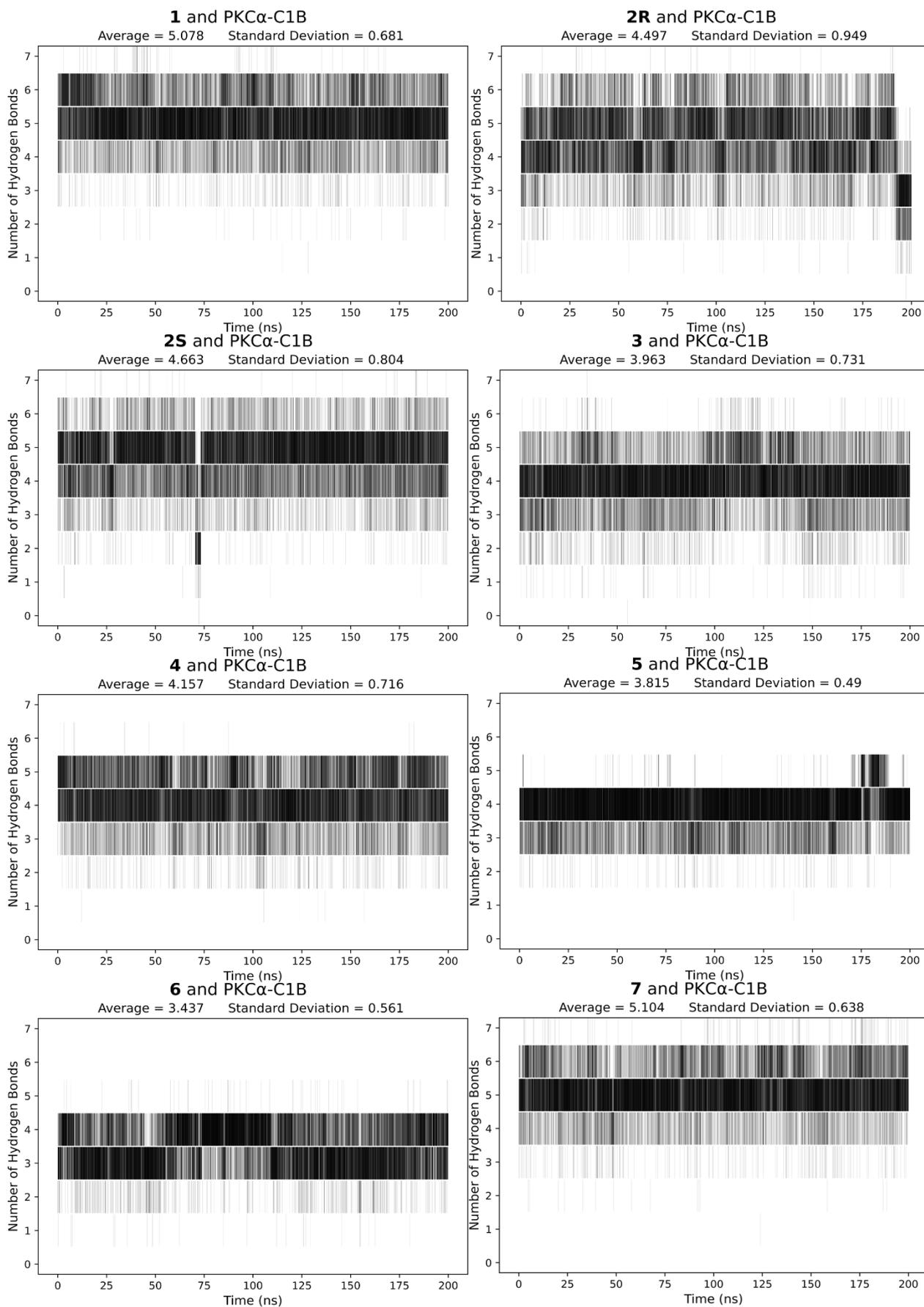
**Figure S7.** Comparison of the hydrogen bond occupancies of simplexin complexed to PKC $\alpha$ -C1B in aqueous phase vs membrane.



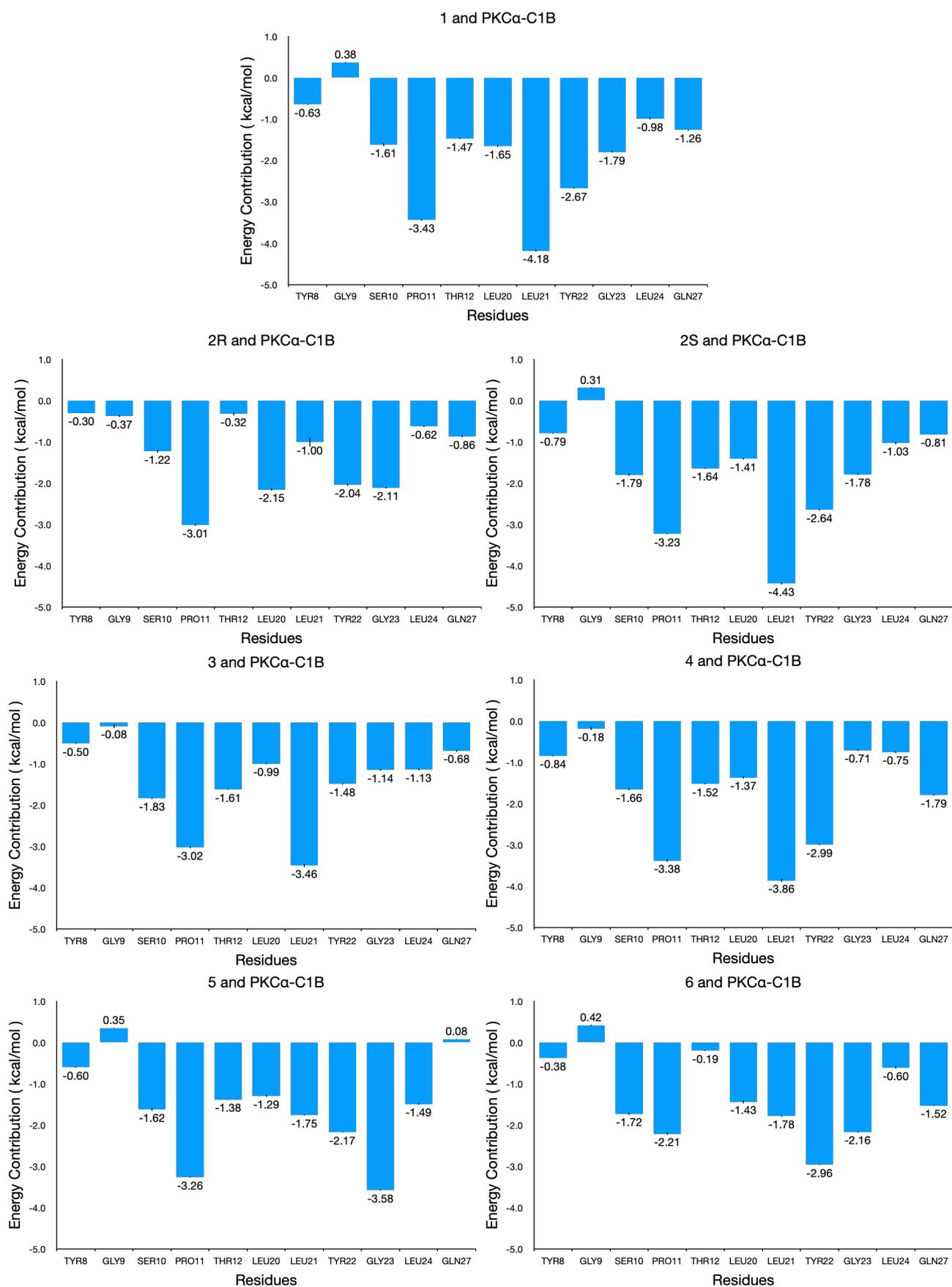
**Figure S8.** RMSD of heavy atoms of simplexin and its analogues complexed with PKC $\alpha$ -C1B in membrane.



**Figure S9.** RMSD of backbone of PKC $\alpha$ -C1B complexed with simplexin and its analogues in membrane.

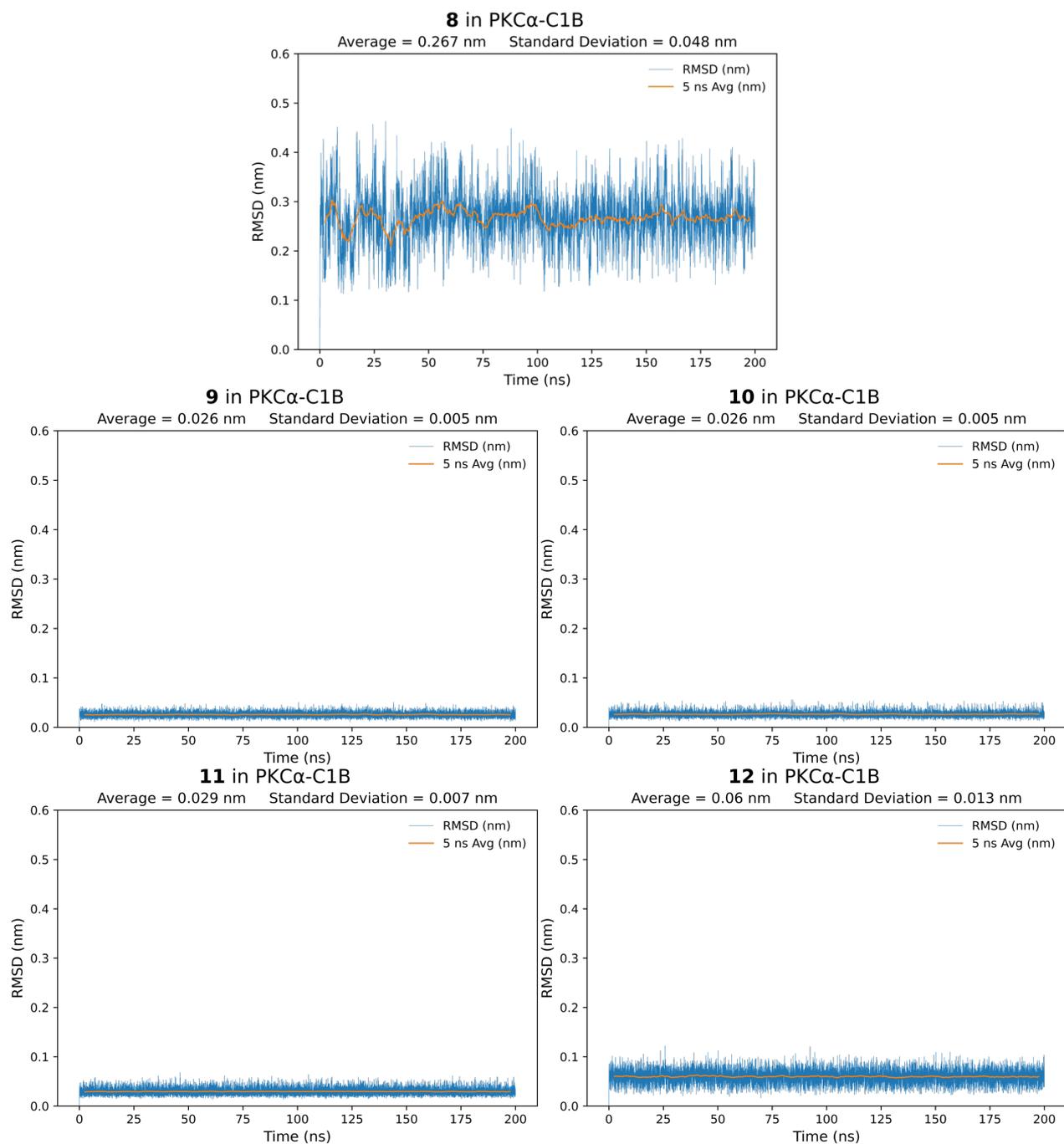


**Figure S10.** Number of hydrogen bonds between PKC $\alpha$ -C1B and simplexin and its analogues in membrane.

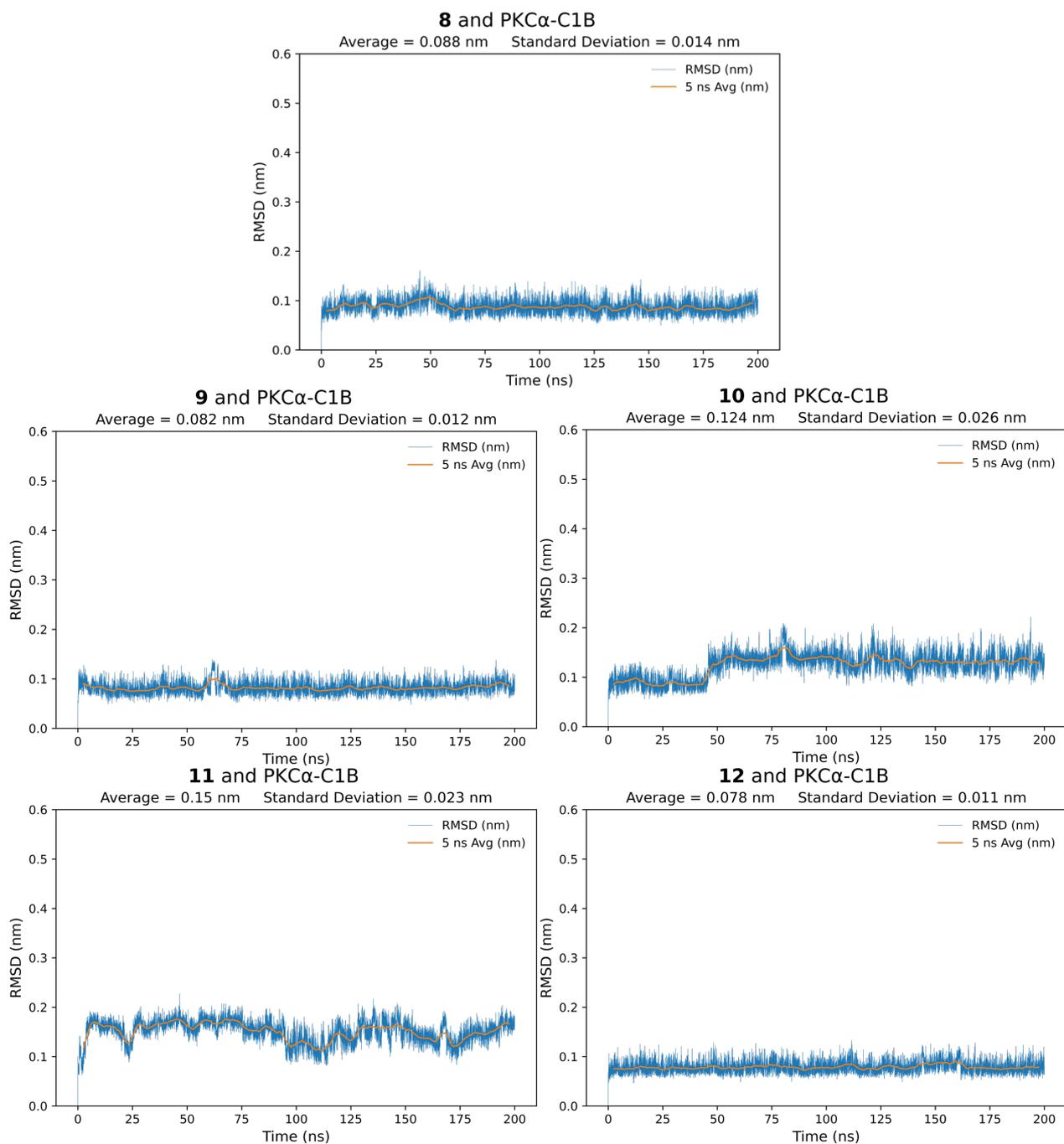


**Figure S11.** Per-residue decomposition analyses of the binding enthalpies of simplexin and its analogues complexed with PKC $\alpha$ -C1B in membrane.

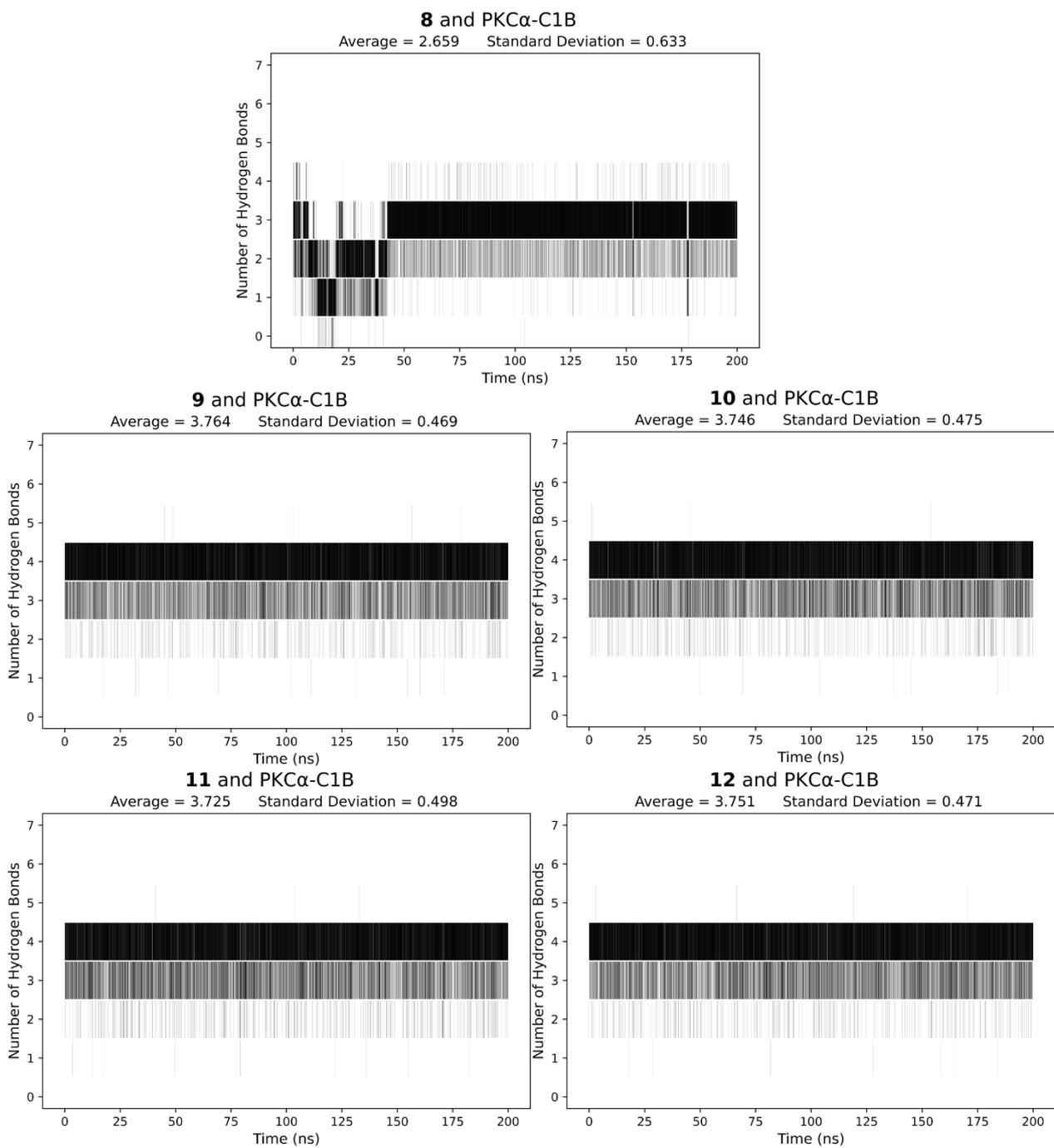
**Supplementary figures for simulations of DAG and phorbol esters with PKC $\alpha$ -C1B in water**



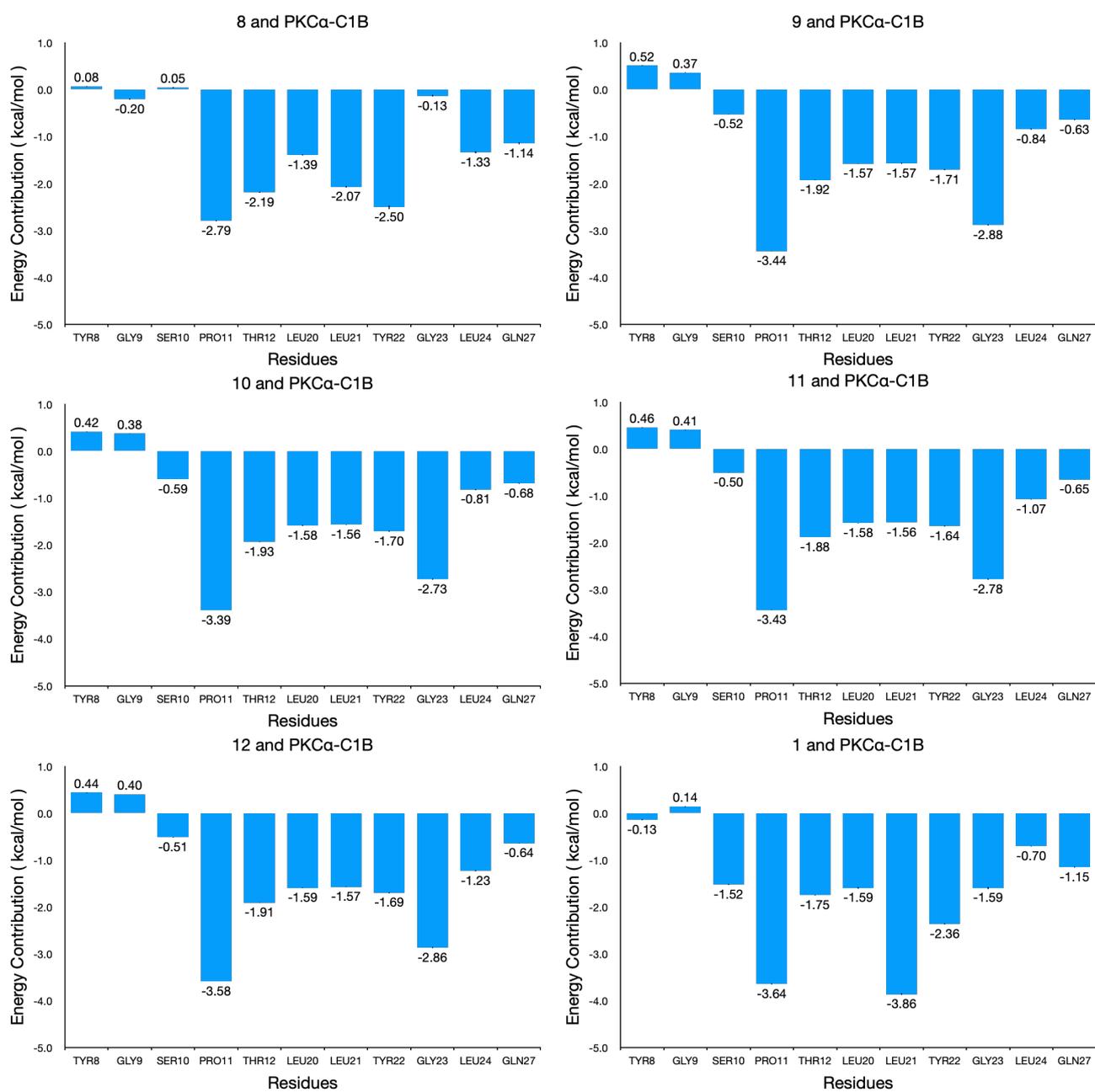
**Figure S12.** RMSD of heavy atoms of DAG and phorbol esters complexed with PKC $\alpha$ -C1B in water.



**Figure S13.** RMSD of backbone of PKC $\alpha$ -C1B complexed with DAG and phorbol esters in water.

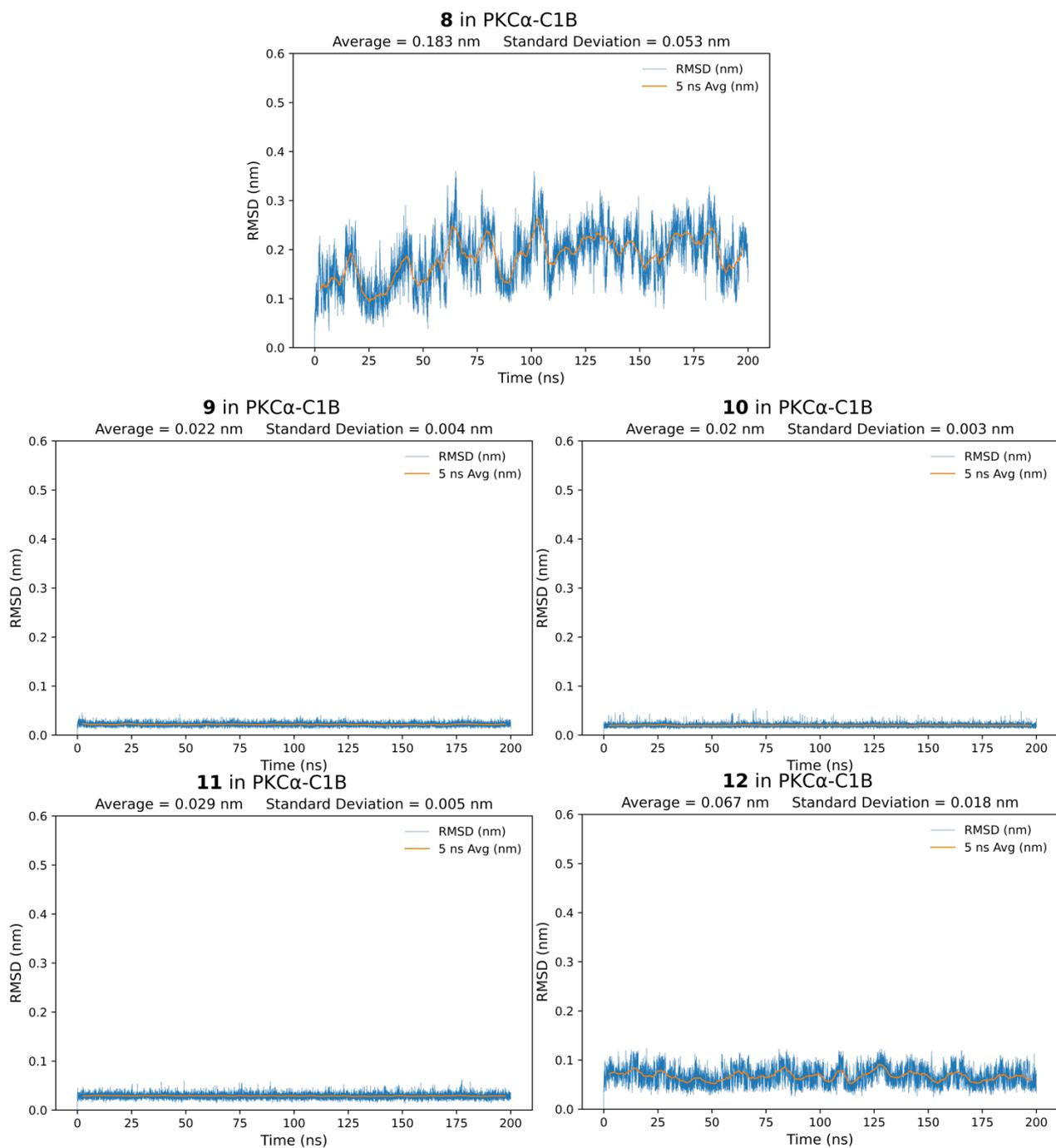


**Figure S14.** Number of hydrogen bonds between PKC $\alpha$ -C1B and ligands **9–12** in water.

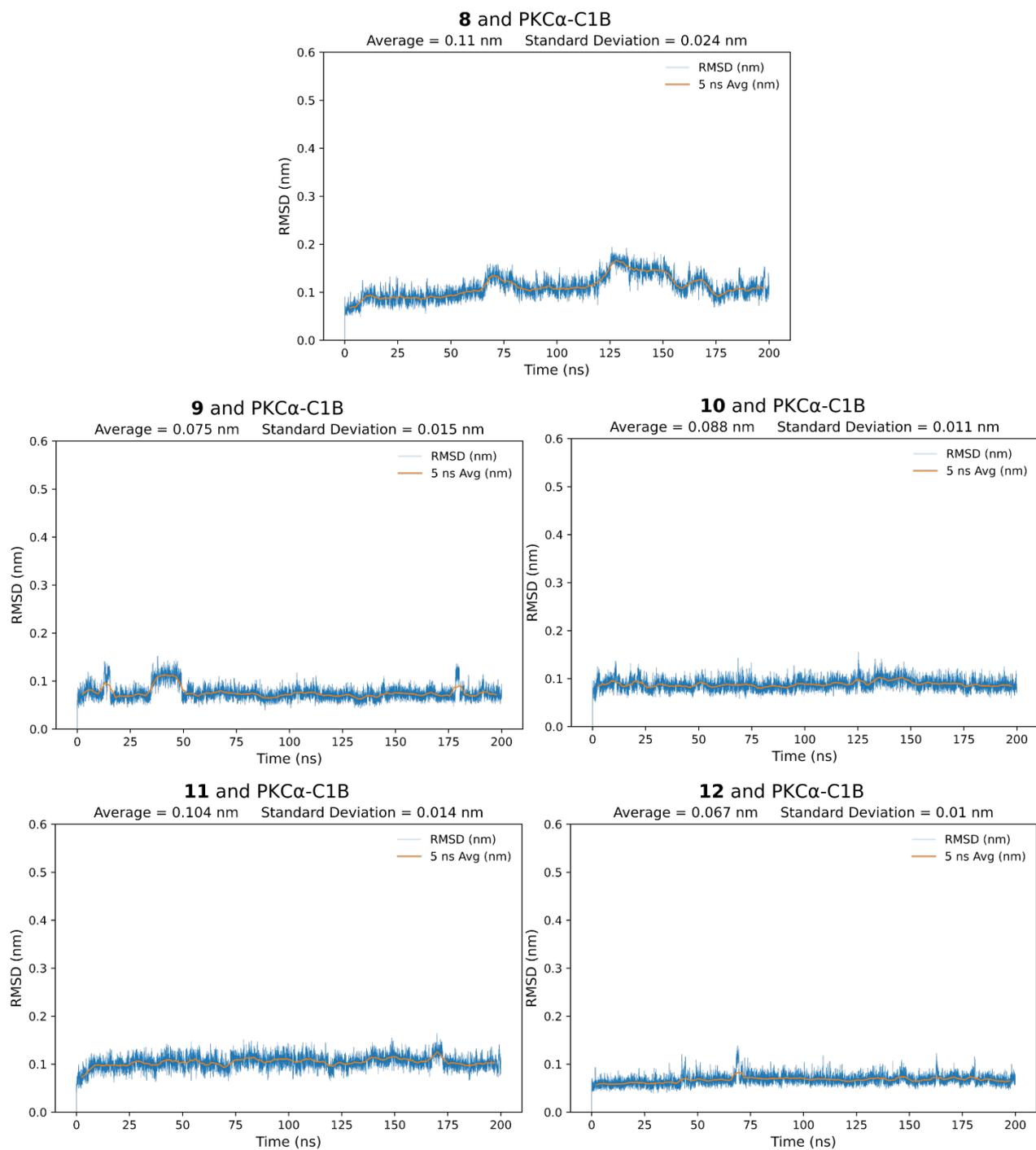


**Figure S15.** Per-residue decomposition analyses of the binding enthalpies of ligands **9–12** complexed with PKC $\alpha$ -C1B in aqueous phase. The enthalpy decomposition of simplexin with PKC $\alpha$ -C1B is also included here for the convenience of comparison.

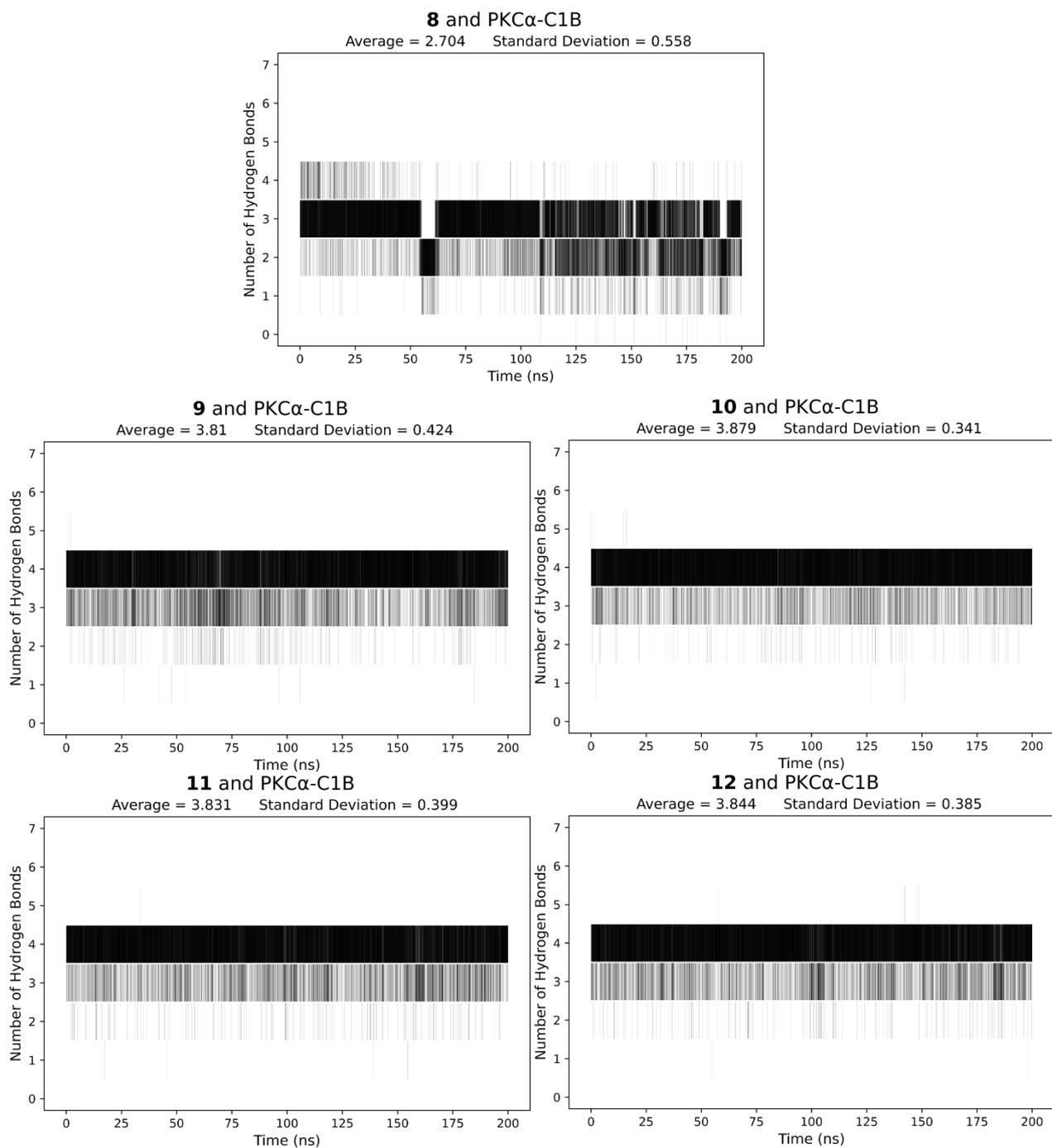
**Supplementary figures for simulations of DAG and phorbol esters with PKC $\alpha$ -C1B in membrane**



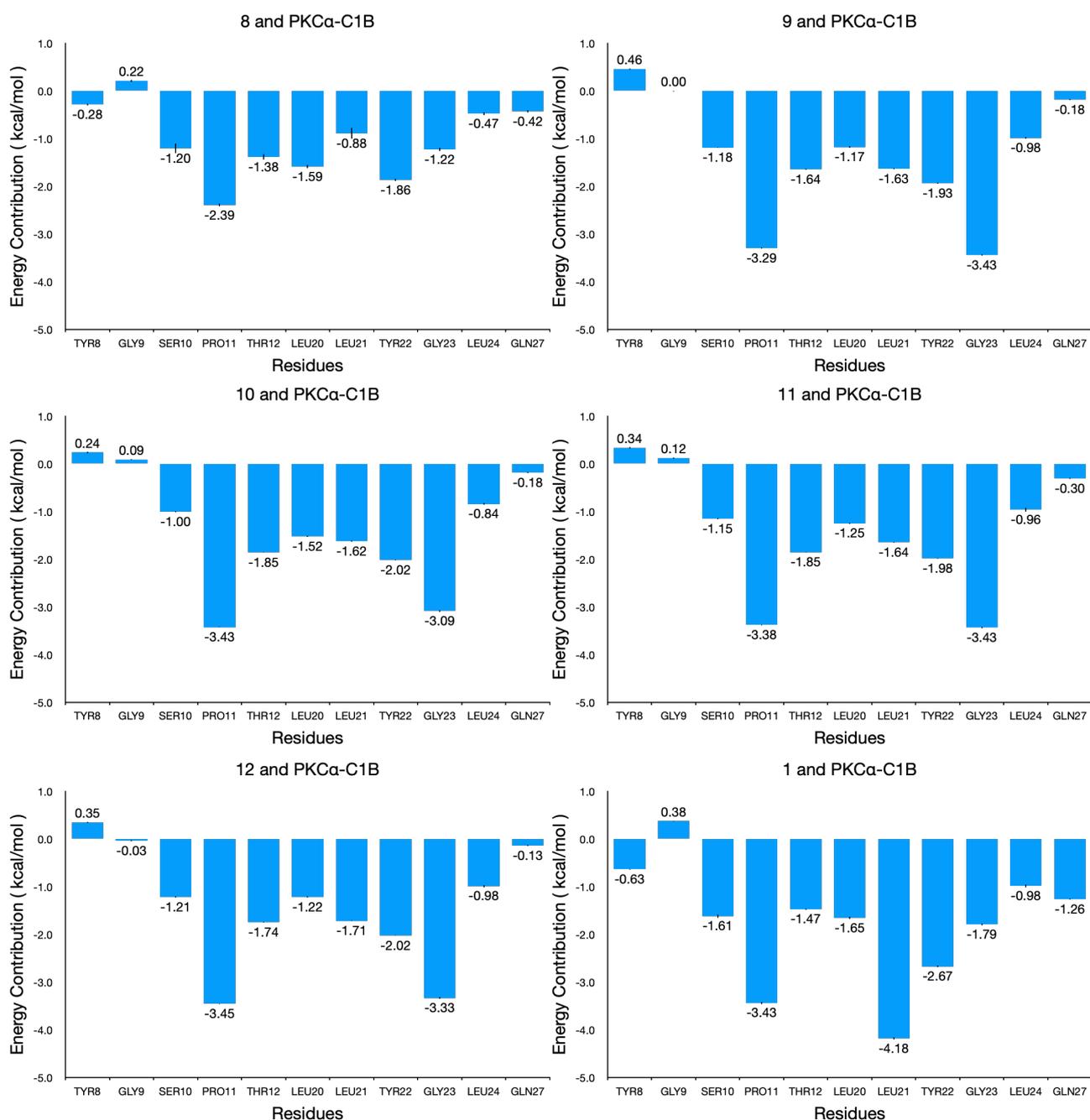
**Figure S16.** RMSD of heavy atoms of DAG and phorbol esters complexed with PKC $\alpha$ -C1B in membrane.



**Figure S17.** RMSD of backbone of PK $\alpha$ -C1B complexed with DAG and phorbol esters in membrane.



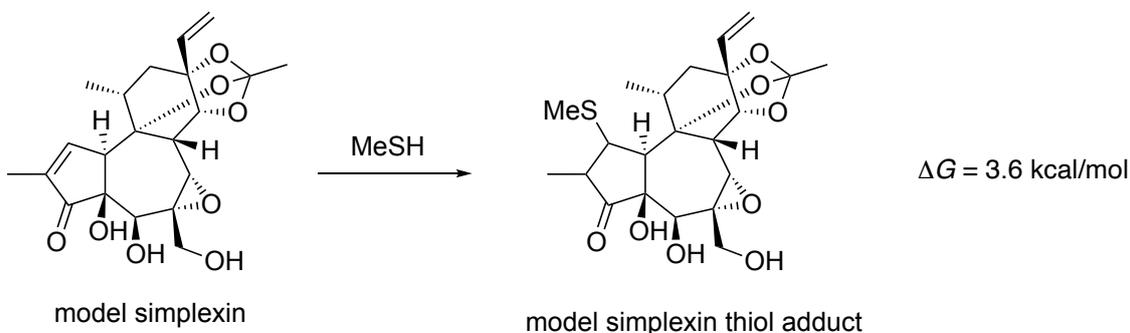
**Figure S18.** Number of hydrogen bonds between PKC $\alpha$ -C1B and ligands **9–12** in membrane.



**Figure S19.** Per-residue decomposition analyses of the binding enthalpies of ligands **9–12** complexed with PKC $\alpha$ -C1B in aqueous phase. The enthalpy decomposition of simplexin with PKC $\alpha$ -C1B is also included here for the convenience of comparison.

## Quantum Mechanical Calculations

As described in ref. 60 of the manuscript, we undertook quantum mechanical calculations with DFT to investigate the thermodynamics of thiol addition to simplexin. Our calculations, using a truncated model of simplexin (**Fig. S20**), predicted a  $\Delta G$  value of  $\geq 3.6$  kcal/mol for the conjugate addition of MeSH to the Michael acceptor moiety.



**Figure S20.** Addition of MeSH to a model simplexin, calculated with  $\omega$ B97X-D/6-311+G(d,p)//M06-2X/6-31+G(d) in CPCM water.

Conformational sampling was performed on the model simplexin and its various diastereomeric thiol adducts using the macrocycle conformational sampling algorithm of MacroModel.<sup>34</sup> In the searches, the OPLS3e force field was used,<sup>35</sup> and the water solvent was simulated with a constant-dielectric solvent model. For each species, the low-energy conformers lying within 3 kcal/mol of the global minimum were then optimized using density functional theory at the M06-2X/6-31+G(d) level of theory<sup>36</sup> in *Gaussian 16*.<sup>12</sup> The CPCM implicit solvent model<sup>37</sup> of water was used. Harmonic vibrational frequency calculations at this level confirmed that each stationary point corresponded to an energy minimum. Truhlar's quasiharmonic approximation<sup>38</sup> was employed to treat low frequency modes ( $\leq 100 \text{ cm}^{-1}$ ) in the calculation of entropies. Single-point energy calculations were performed with  $\omega$ B97X-D/6-311+G(d,p)<sup>39</sup> in CPCM water. Gibbs free energies were computed by adding the quasiharmonically corrected M06-2X thermochemical quantities to the  $\omega$ B97X-D single-point potential energies and are reported at a standard state of 298.15 K and  $1 \text{ mol L}^{-1}$ .<sup>40</sup>

Below are reported the optimized coordinates for the most stable conformer of each species, along with the following energies (in Hartree):

$E_{\text{M06-2X}}$	M06-2X electronic potential energy		
$H_{\text{M06-2X}}$	M06-2X enthalpy at 298.15 K		
$G_{\text{M06-2X}}$	M06-2X Gibbs free energy at 298.15 K and $1 \text{ mol L}^{-1}$		
$E_{\text{SP-}\omega\text{B97X-D}}$	$\omega$ B97X-D single-point electronic potential energy		

### MeSH

S	0.048360	-0.666557	0.000000
H	-1.282490	-0.836073	0.000000
C	0.048360	1.156567	0.000000
H	-0.437791	1.547455	0.894067
H	1.094163	1.466680	0.000000
H	-0.437791	1.547455	-0.894067

0 imaginary frequencies

$$E_{\text{M06-2X}} = -438.633414$$

$$H_{M06-2X} = -438.582229$$

$$G_{M06-2X} = -438.608524$$

$$E_{SP-\omega B97X-D} = -438.710141$$

**Model simplexin**

C	-1.495519	-2.168953	0.968422
C	-2.732483	-2.482989	0.541533
C	-3.253518	-1.336408	-0.216115
O	-4.383691	-1.183957	-0.655076
C	-2.131618	-0.290529	-0.379894
O	-1.724198	-0.316634	-1.736866
C	-2.647496	1.103562	-0.012695
C	0.628728	0.792099	-0.571216
C	0.417192	-0.600006	0.087540
O	1.168043	-0.678598	1.325456
C	-1.033144	-0.788425	0.571512
C	0.911226	-1.666071	-0.941266
C	2.131069	-1.138061	-1.740760
C	1.227338	-3.035628	-0.330456
C	-3.549705	-3.704275	0.808951
C	-2.161664	3.592607	0.033193
O	-1.236183	4.568232	0.462314
H	-0.906511	-2.792361	1.635054
H	-2.362258	0.233371	-2.227221
H	-3.166492	1.043976	0.956888
H	0.363739	0.699101	-1.627537
H	-1.080778	-0.182044	1.483137
H	0.083248	-1.789819	-1.646997
H	2.761565	-1.976884	-2.053588
H	1.619952	-3.691453	-1.114318
H	1.980204	-2.956417	0.457142
H	0.342145	-3.526639	0.077832
H	-3.824001	-4.196446	-0.129881
H	-3.000071	-4.415514	1.429344
H	-4.481083	-3.437945	1.319553
H	-3.080551	3.631570	0.635448
H	-2.421899	3.822924	-1.001077
H	-0.910835	4.296573	1.336808
C	-1.580708	2.183700	0.128494
C	-0.127978	2.007401	-0.062065
H	0.396676	2.937126	-0.289647
O	-3.545394	1.536140	-1.025144
H	-4.301791	0.920635	-1.022389
O	-0.705671	2.034309	1.249791
C	2.131071	1.065756	-0.464849
H	2.426596	1.984271	-0.975906
C	2.974178	-0.176729	-0.903348
O	2.457202	1.183615	0.918250
H	1.818087	-0.622371	-2.656130
C	4.247456	0.228890	-1.591450
H	4.117312	0.610787	-2.604178
C	5.459702	0.154730	-1.045555
H	5.598153	-0.225292	-0.037592
H	6.341348	0.470365	-1.594801
O	3.266235	-0.801283	0.354025
C	2.475633	-0.161835	1.324682
C	3.061471	-0.324583	2.698663

H	2.440949	0.201347	3.426477
H	3.095740	-1.386029	2.952905
H	4.072422	0.087074	2.708740

**0 imaginary frequencies**

$E_{M06-2X} = -1417.103449$

$H_{M06-2X} = -1416.614010$

$G_{M06-2X} = -1416.691969$

$E_{SP-\omega B97X-D} = -1417.591352$

**Model simplexin thiol adduct (most stable diastereomer)**

C	-3.059047	-0.712964	-1.071342
O	-3.843622	-0.584870	-1.983024
C	-1.887632	0.260442	-0.825345
O	-1.381182	0.675280	-2.074726
C	-2.411400	1.466341	-0.029104
C	0.872698	1.240995	-0.452660
C	0.615160	-0.284771	-0.274791
O	1.282699	-0.730362	0.932008
C	-0.878101	-0.583126	0.001477
C	1.203254	-1.000480	-1.535305
C	2.444345	-0.245167	-2.074523
C	1.583312	-2.474836	-1.334263
C	-1.842046	3.807080	0.795187
O	-3.143529	3.710322	1.369834
H	-2.127193	1.086040	-2.547197
H	-2.982357	1.088245	0.833194
H	0.652427	1.501343	-1.490488
H	-1.023681	-0.276232	1.039672
H	0.416989	-0.939433	-2.296375
H	3.091646	-0.941800	-2.617833
H	1.855766	-2.897581	-2.307435
H	0.786131	-3.086149	-0.916647
H	2.446111	-2.562750	-0.671624
H	-1.880809	4.362238	-0.149621
H	-1.147307	4.308267	1.477402
H	-3.481593	4.601495	1.542556
C	-1.340842	2.402217	0.539984
C	0.115245	2.238644	0.406571
H	0.681265	3.164820	0.537711
O	-3.265214	2.168201	-0.911556
H	-3.772595	2.804150	-0.377098
O	-0.557510	1.850776	1.604763
C	2.365975	1.448264	-0.188980
H	2.685096	2.474267	-0.383412
C	3.237815	0.386606	-0.928708
O	2.603891	1.143224	1.183263
H	2.167217	0.544462	-2.782461
C	4.554189	0.955363	-1.378626
H	4.492472	1.625215	-2.236562
C	5.727004	0.697251	-0.803306
H	5.797088	0.030152	0.050881
H	6.643545	1.145386	-1.174404
O	3.445232	-0.592652	0.098383
C	2.587939	-0.262792	1.164583
C	3.078573	-0.847381	2.459097
H	2.412726	-0.547583	3.270763
H	3.084728	-1.936951	2.382336

H	4.089750	-0.487487	2.658196
C	-2.943162	-1.887578	-0.133649
C	-1.406751	-2.039208	-0.088573
H	-1.100932	-2.507094	-1.031880
S	-0.843802	-3.198290	1.203883
C	-1.064316	-2.216934	2.712862
H	-0.311473	-1.428846	2.774945
H	-0.931639	-2.910172	3.546274
H	-2.069473	-1.791318	2.776246
C	-3.725856	-3.109659	-0.577691
H	-3.584304	-3.930503	0.129691
H	-4.793183	-2.880513	-0.637623
H	-3.391899	-3.439969	-1.567066
H	-3.287943	-1.548250	0.856779

**0 imaginary frequencies**

**$E_{M06-2X} = -1855.760833$**

**$H_{M06-2X} = -1855.219101$**

**$G_{M06-2X} = -1855.299371$**

**$E_{SP-\omega B97X-D} = -1856.320807$**

## References

1. M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Zidek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis and S. Velankar, *Nucleic Acids Res.*, 2022, **50**, D439-D444.
2. G. Zhang, M. G. Kazanietz, P. M. Blumberg and J. H. Hurley, *Cell*, 1995, **81**, 917-924.
3. S. Jo, T. Kim and W. Im, *PLoS One*, 2007, **2**, e800.
4. S. Jo, T. Kim, V. G. Iyer and W. Im, *J. Comput. Chem.*, 2008, **29**, 1859-1865.
5. E. L. Wu, X. Cheng, S. Jo, H. Rui, K. C. Song, E. M. Dávila-Contreras, Y. Qi, J. M. Lee, V. Monje-Galvan, R. M. Venable, J. B. Klauda and W. Im, *J. Comput. Chem.*, 2014, **35**, 1997-2004.
6. Yinfo Cloud Computing Platform, <https://cloud.yinfotek.com>, (accessed 12 Dec. , 2023).
7. O. Trott and A. J. Olson, *J. Comput. Chem.*, 2010, **31**, 455-461.
8. P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623-11627.
9. S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
10. S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456-1465.
11. R. Krishnan, J. S. Binkley, R. Seeger and J. A. Pople, *J. Chem. Phys.*, 1980, **72**, 650-654.
12. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian 16, Revision C.01, Gaussian, Inc., Wallingford CT, 2016.
13. T. I. Igumenova, *Biochemistry*, 2015, **54**, 4953-4968.
14. F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297-3305.
15. G. Scalmani and M. J. Frisch, *J. Chem. Phys.*, 2010, **132**, 114110.
16. D. A. Case, H. M. Aktulga, K. Belfon, D. S. Cerutti, G. A. Cisneros, V. W. D. Cruzeiro, N. Forouzes, T. J. Giese, A. W. Gotz, H. Gohlke, S. Izadi, K. Kasavajhala, M. C. Kaymak, E. King, T. Kurtzman, T. S. Lee, P. Li, J. Liu, T. Luchko, R. Luo, M. Manathunga, M. R. Machado, H. M. Nguyen, K. A. O'Hearn, A. V. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, A. Rishch, S. Schott-Verdugo, A. Shajan, J. Swails, J. Wang, H. Wei, X. Wu, Y. Wu, S. Zhang, S. Zhao, Q. Zhu, T. E. Cheatham, 3rd, D. R. Roe, A. Roitberg, C. Simmerling, D. M. York, M. C. Nagan and K. M. Merz, Jr., *J. Chem. Inf. Model.*, 2023, **63**, 6183-6191.
17. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1**, 19-25.
18. A. W. Sousa da Silva and W. F. Vranken, *BMC Res Notes*, 2012, **5**, 367.
19. C. J. Dickson, R. C. Walker and I. R. Gould, *J. Chem. Theory Comput.*, 2022, **18**, 1726-1736.
20. S. M. Ryckbosch, P. A. Wender and V. S. Pande, *Nat. Commun.*, 2017, **8**, 6.
21. K. Huang and A. E. Garcia, *J. Chem. Phys.*, 2014, **141**, 105101.
22. I. S. Joung and T. E. Cheatham, 3rd, *J. Phys. Chem. B*, 2008, **112**, 9020-9041.
23. J. Åqvist, *J. Phys. Chem.*, 1990, **94**, 8021-8024.

24. J. Pan, X. Cheng, L. Monticelli, F. A. Heberle, N. Kucerka, D. P. Tieleman and J. Katsaras, *Soft Matter*, 2014, **10**, 3716-3725.
25. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926-935.
26. User contributions, [https://www.gromacs.org/user\\_contributions.html](https://www.gromacs.org/user_contributions.html), (accessed 12 Dec., 2023).
27. B. Hess, H. Bekker, H. J. Berendsen and J. G. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463-1472.
28. T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089-10092.
29. M. A. Lomize, I. D. Pogozheva, H. Joo, H. I. Mosberg and A. L. Lomize, *Nucleic Acids Res.*, 2012, **40**, D370-376.
30. M. S. Valdés-Tresanco, M. E. Valdés-Tresanco, P. A. Valiente and E. Moreno, *J. Chem. Theory Comput.*, 2021, **17**, 6281-6291.
31. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157-1174.
32. Q. Lu and R. Luo, *J. Chem. Phys.*, 2003, **119**, 11035-11047.
33. N. Michaud-Agrawal, E. J. Denning, T. B. Woolf and O. Beckstein, *J. Comput. Chem.*, 2011, **32**, 2319-2327.
34. *MacroModel*, Schrödinger LLC, New York, NY, 2020.
35. K. Roos, C. Wu, W. Damm, M. Reboul, J. M. Stevenson, C. Lu, M. K. Dahlgren, S. Mondal, W. Chen, L. Wang, R. Abel, R. Friesner and E. D. Harder, *J. Chem. Theory Comput.*, 2019, **15**, 1863.
36. Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* 2008, **120**, 215.
37. (a) V. Barone and M. Cossi, *J. Phys. Chem. A*, 1998, **102**, 1995; (b) M. Cossi, N. Rega, G. Scalmani and V. Barone, *J. Comput. Chem.*, 2003, **24**, 669.
38. Y. Zhao and D. G. Truhlar, *Phys. Chem. Chem. Phys.*, 2008, **10**, 2813.
39. J. D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615.
40. J. Y. C. Ting, R. B. Roseli and E. H. Krenske, *Org. Biomol. Chem.*, 2020, **18**, 1426.