Supporting information

"Membrane-mediated regulation of the intrinsically disordered CD3ε cytoplasmic tail of the TCR"

Cesar A. López*, Anurag Sethi†, Byron Goldstein*, Bridget S. Wilson‡, S. Gnanakaran*∓.*

*Department of Theoretical Biology and Biophysics, Los Alamos National Laboratory, USA. †Department of Molecular Biophysics and Biochemistry, Yale University, USA. ‡Department of Pathology and Cancer Center, University of New Mexico Health Sciences Center. ∓New Mexico Consortium, 100 Entrada Rd., Los Alamos, NM 87544, USA

SI Materials and Methods

System set-up

We built different membrane models in order to study the preferential solvation and co-localization of the CD3ε chain. First, we set up an independent fluid domain by mixing palmitoyloleoylphosphatidylglycerol (POPG), palmitoyloleoylphosphatidylcholine (POPC), palmitoyloleoylphosphatidylserine (POPS), palmitoyloleoylphosphatidylethanolamine (POPE) and phosphatidylinositol (PI). The lipid mixture was solvated with 15000 CG polarizable water beads and equilibrated during 5 μs. At the end of the simulation, the membrane was observed in a homogenously mixed state, with no clear presence of nucleation. Afterwards the CD3ε with the transmembrane region was inserted by removing overlapping lipids.

Similarly, a multicomponent system was built to study the preferential co-localization of the protein. A lamellar structure was set up by adding both saturated (dipalmitoylphosphatidylcholine) and unsaturated (palmitoyloleoylphosphatidylglycerol, dilnoylphosphatidilcholine: DUPC) lipids, in combination with cholesterol (CHOL) in common lipid ratio for Liquid order (L_o)/Liquid disordered (L_d) domain formation (DPPC:4, CHOL:3, DUPC:2, POPG:1). The membrane was fully solvated using 20000 coarse-grain (CG) polarizable water beads and equilibrated (see results). Ganglioside GM1 was also incorporated in the membrane, accounting for 1% of the total lipid content.

Computational details

All-atom simulations were performed using a 2 fs time step to integrate Newton's equations of motion. The LINCS algorithm[1] was applied to constrain all bond lengths with a relative geometric tolerance of 10^{-4}. Non-bonded interactions were handled using a twin-range cutoff[2] scheme. Within a short-range cutoff of 0.9 nm, the interactions were evaluated every time step based on a pair list recalculated every five time steps. The intermediate-range interactions up to a long-range cutoff radius of 1.4 nm were evaluated simultaneously with each pair list update and were assumed constant in between. A PME approach[3] was used to account for electrostatic interactions with a grid spacing set to 0.15 nm. Constant temperature was maintained by weak

coupling of the solvent and solute separately to a Berendsen heat bath[4] with a relaxation time of 1.0 ps. Bilayers were coupled at 1.0 bar through a semiisotropic approach with relaxation time of 1.0 ps. Trajectories frames were stored every 20 ps. In the simulations at the CG level, we followed the simulation protocol optimized for the polarizable variant of the MARTINI model[5]. The non-bonded interactions are cut-off at a distance r_{cut} of 1.2 nm. To reduce the generation of unwanted noise, the standard shift function of GROMACS[6] is used in which both the energy and force smoothly vanish at the cutoff distance. Notice that no detectable effects have been observed by using PME or shift independently. Since the polarization of water is treated explicitly, the global dielectric constant is adjusted to ε=2.5. The LJ and Coulomb potentials are shifted from r = 0.0 and r = 0.9 nm to the cutoff distance, respectively. The time step used to integrate the equations of motion is 25 fs. Constant temperature is maintained by weak coupling of the solvent and membranes separately to a Berendsen heat bath[4] with a relaxation time of 1.0 ps. Bilayers were coupled to a semiisotropic approach with a relaxation time of 1.0 ps and tight to 1.0 bar.

CG mapping

AA configurations were converted to CG coordinates using the center of mass of the appropriate fine-grained beads[7].

$$r_i^{CG} = \frac{\sum_{j=1}^{P} r_j . m_j}{\sum_{j=1}^{P} m_j}$$

The vector r^{CG} describes the position of the CG bead, p is the number of atoms mapped to a given coarse bead, m_j is the mass of the atom j, and r_j is its coordinates.

Free energy calculations

An umbrella sampling approach was used to calculate the potential of mean force (PMF) for the detachment of the CD3ε ITAM region from the different lipid bilayers. For the PMF calculations, we used the umbrella sampling method[8] with 90 window points, spaced by 1 A, restraining the C-terminal (specifically the two last residues of the chain) of the protein with respect to the center of mass of the bilayer. The restraining potential was harmonic with a force constant of 1,000 kJ mol^{-1} nm^{-2}. 100 ns of simulation were performed for each window, covering a total of 9 μs per system. In addition, we also calculated the affinity constant of the cytoplasmic tail to pure POPG bilayers. The procedure was repeated exactly as mentioned before, however, the peptide was pulled from its total center of mass. The PMFs were reconstructed using the weighted histogram analysis method[9], with 200 bins for each profile.

Preferential partitioning of the CD3ε chain

The preferential partitioning of the CD3ε chain with the membrane components is calculated as the relative number of contacts of a lipid species with the protein, corrected by the total number

of lipids in the system:

$$P_A = \frac{\frac{C_A}{n_A}}{\frac{\sum_x C_x}{n_x}}$$

where P_A is the preferential partitioning with the membrane component A, C_A the number of contacts with component A, and n_A the number of molecules of component A. Contacts were defined with respect to the glycerol beads and beads corresponding to the protein. Two molecules were counted within contact if they were inside the first solvation shell (0.8 nm). The GROMACS analysis tool g_mindist was used to calculate the number of contacts, analyzing the frame every 1 ns. The first μs was considered as equilibration time, and the following 4 μs were used for the analysis.

SI Results

Structural conformation of the CD3ε upon binding

The conformation of the CD3ε chain has been observed to be dependent of the binding mode to different lipids. As depicted in Figure S5, the ITAM region folds into an alpha helix upon binding to the either the POPG or POPC membrane. As discussed in the main manuscript, the close interaction with the membrane interface, combined with the burying of the tyrosines, stabilizes the cytoplasmic chain, resulting in the presence of the helix. Although the protein is not fully buried in the POPC bilayer, the region in closer contact with the bilayer is seen to adopt a stable alpha helix along the simulation time. A different behavior is observed upon the interaction with the DPPC bilayer, namely a structureless configuration. Contrary to the previous interactions, the protein is unable to stick closer to the interface, and as a result, is fully solvated in the water bulk. Thus, our results provide clear atomic evidence for the facilitated folding upon binding theory, as proposed by Sigalov[10–12].

Membrane properties affected by the transmembrane region of CD3ε

In the main manuscript it is concluded that the binding mechanism is not affected by the presence of the trans-membrane region. However, the comparison of the free energies profiles between the cleaved and uncleaved protein did show a clear difference. We attribute that difference to the perturbing effect of the transmembrane helix into the structure of the bilayer. Our hypothesis is corroborated by the electrodensity profile presented in Figure S6. For the comparison, equivalent simulations were run for 10 μs with either the cleaved CD3ε chain (solid lines) or the cytoplasmic region attached to the transmembrane region (broken lines). The insertion of the helix clearly shifts the relative position of all the peaks, suggesting a decrease of the bilayer thickness. The downshift displacement of the peaks of the head groups (orange lines) and the glycerol moiety (red lines) also reveals a less structured configuration of the inter-region facing the water molecules. As a result, the stability of the protein in the interface is affected as explained in detail in the main manuscript.

Preferential lipid association to the CD3ε chain

Figure S7 shows the detailed Radial distribution function used to quantify the relative preference of the CD3ε chain to different lipids. As explained in detailed in the main manuscript, the protein has shown stronger interactions with phosphoinositols rather than POPG. In our set-up (Figure S4) the protein was colocalized in a mixed bilayer composed of POPC, POPS, POPE PI and POPG and the RDF averaged along 4 μs. As expected, the bottom line of interaction corresponds to the zwitterionic POPE and POPC lipids. Higher in the interaction, we found POPS and POPG. However, the presence of PI at the top level of interaction results unexpected, nevertheless, the strength of the interaction counts by fifty percent of the total lipid contacts, as further illustrated in Table S1. We also found that the PI is preferentially localized around the basic residues stretch (BRS) as clearly shown in Figure S7-A. These results are not only in line with different experimental findings, but suggest that indeed the BRS region is used as a specific lipid-binding motif. In conclusion, our data suggest: i) the BRS region pivots specific interactions to different lipid species, and ii) these interactions are not only restricted to charged counterparts but to molecules bearing high dipole moments.

Domain formation in a lipid mixed bilayer

We set up a simulation in order to establish the long-term lipid segregation in a plane bilayer. To that goal we used a combination of saturated lipids (DPPC), unsaturated lipids (DUPC, POPG) and cholesterol. Lipids were added in ratios commonly known to form equilibrated Lo-Ld domains (DPPC:4, CHOL:3, DUPC:2, POPG:1). The temperature of the system was set up at 288 K and coupled to a semi isotropic pressure scheme. The transition during the domain formation is detailed in Figure S8, for a total 10 μs simulation. Extra simulations were set up, adding either GM1 or the phosphoinositol PI at 4% and 10% of the lipid content respectively. In both cases, strong lipid segregation was also observed.

Localization of a CD3ϵ chain in a lipid domain

We have considered and entire CD3ϵ chain containing phosphorylated ITAM tyrosines in a lipid mix mimicking a liquid ordered-liquid disordered array (Figure S9). We do not observe any difference in the preferential localization of the entire chain, as discussed in the main manuscript.

The CD3ε chain makes specific interactions with the ganglioside GM1.

We have described in the main manuscript that GM1 is able to modify the preferential partitioning of the CD3 chain. Closer inspection reveals that indeed this mechanism is favored through interactions between the protein and the sugars present in the ganglioside. Figure S10-A shows an equilibrated configuration of the peptide surrounded by the gangliosides along our simulation trajectory. We found that in average, the protein makes closer interaction with up to 4 gangliosides in a very specific manner. The sidechains in closer interactions are depicted in Figure S10-B. Importantly, most of the residues are polar in nature and are tightly bound to the sugar rings of GM1. However, an aromatic (TYR 67) residue is also found in very close interaction, and probably favored by the flatness of the sugars.

Supporting references

[1] Hess B, Henk B, Berendsen HJC, Fraaije J (1997) LINCS: A linear constraint solver for

molecular simulations. J. Comput. Chem. 18:1463–1472.

[2] van Gunsteren WF, Berendsen HJC (1990) Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry. Angew. Chem. Int. Edit. 29:992–1023.

[3] Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An Nlog (N) method for Ewald sums in large systems. J. Chem. Phys. 98:10089–10092.

[4] Berendsen HJC, Postma JPM, van Gunsteren WF, Dinola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81:3684–3690.

[5] Yesylevskyy SO, Schafer LV, Sengupta D, Marrink SJ (2010) Polarizable water model for the coarse-grained MARTINI force field. Plos Comput. Biol. 6:e1000810.

[6] Pronk S et al. (2013) GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29:845–854.

[7] Rzepiela AJ, et al. (2010) Reconstruction of atomistic details from coarse-grained structures. J. Comput. Chem. 31:1333–1343.

[8] Torrie G, Valleau J (1977) Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. J. Comput. Phys. 23:187–199.

[9] Hub JS, De Groot BL, van der Spoel D (2010) g_wham-A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. J. Chem. Theory Comput. 6:3713–3720.

[10] Sigalov AB (2010) The SCHOOL of nature: II. Protein order, disorder and oligomericity in transmembrane signaling. Self Nonself 1:89–102.

[11] Sigalov AB (2012) Evolution of immunity: no development without risk. Immunol. Res. 52:176–181.

[12] Sigalov AB (2010) Protein intrinsic disorder and oligomericity in cell signaling. Mol. Biosyst. 6:451–461.

	POPC	POPG	POPE	POPS	PI
CDε chain	0.065	0.31	0.07	0.082	0.46

Table S1: Normalized number of contacts P_A of CD3ε chain and surrounding lipids. Errors 5 kJ mol^{-1} at most.

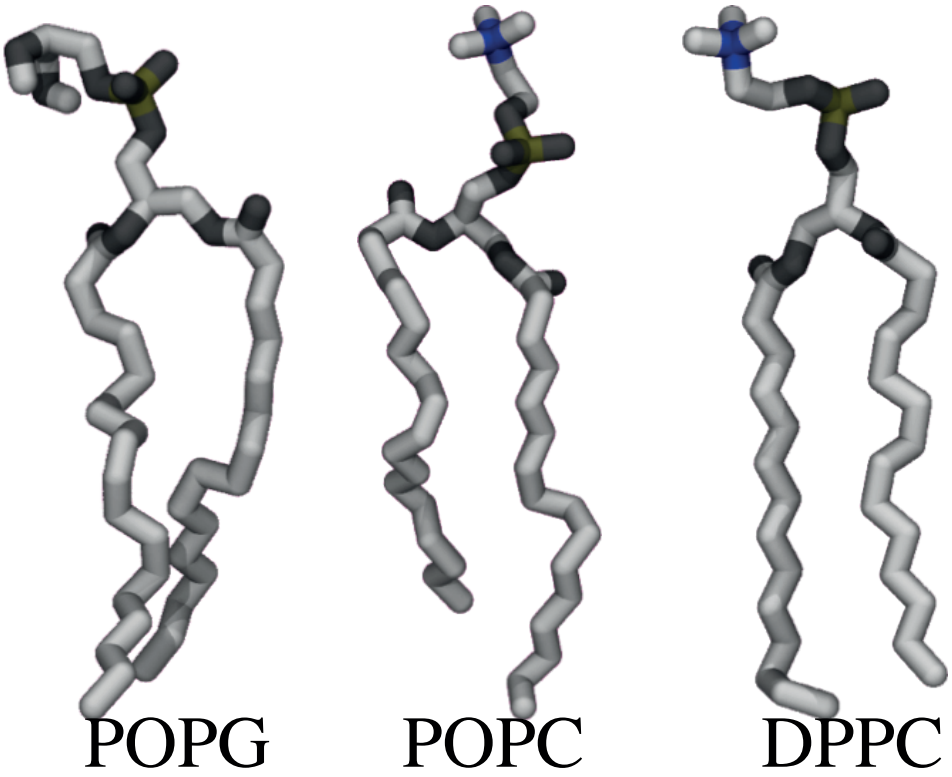

Figure S1: Atomic representations of the lipids used in all atom simulations. The negatively charged POPG and the zwitterionic POPC differ in the head group but have common aliphatic tails (one single unsaturation). DPPC on the other hand is fully saturated and commonly found in Liquid ordered membrane domains

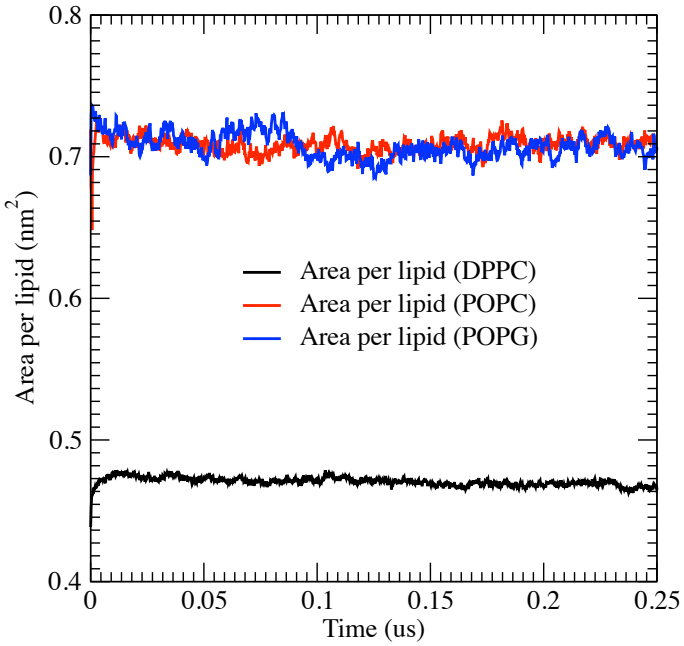

Figure S2. Area per lipid dependency. Area per lipid equilibration was easily reached within the time scale of our simulations using a united atom force field. Note that lipids where composed exclusively of a single specie.

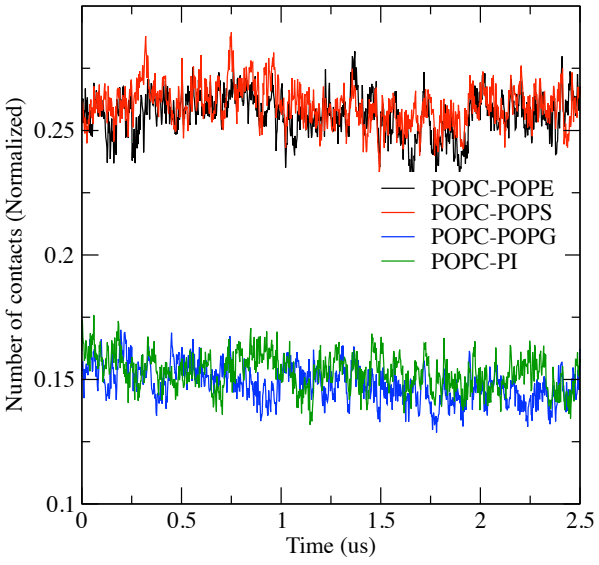

Figure S3. Normalized number of contacts for a mix of lipids in a bilayer. Simulations were performed using a coarse-grained force field, where the degree of freedom is reduced and therefore allows faster sampling with less computational effort. Normalization is done considering the total amount of lipids within the bilayer

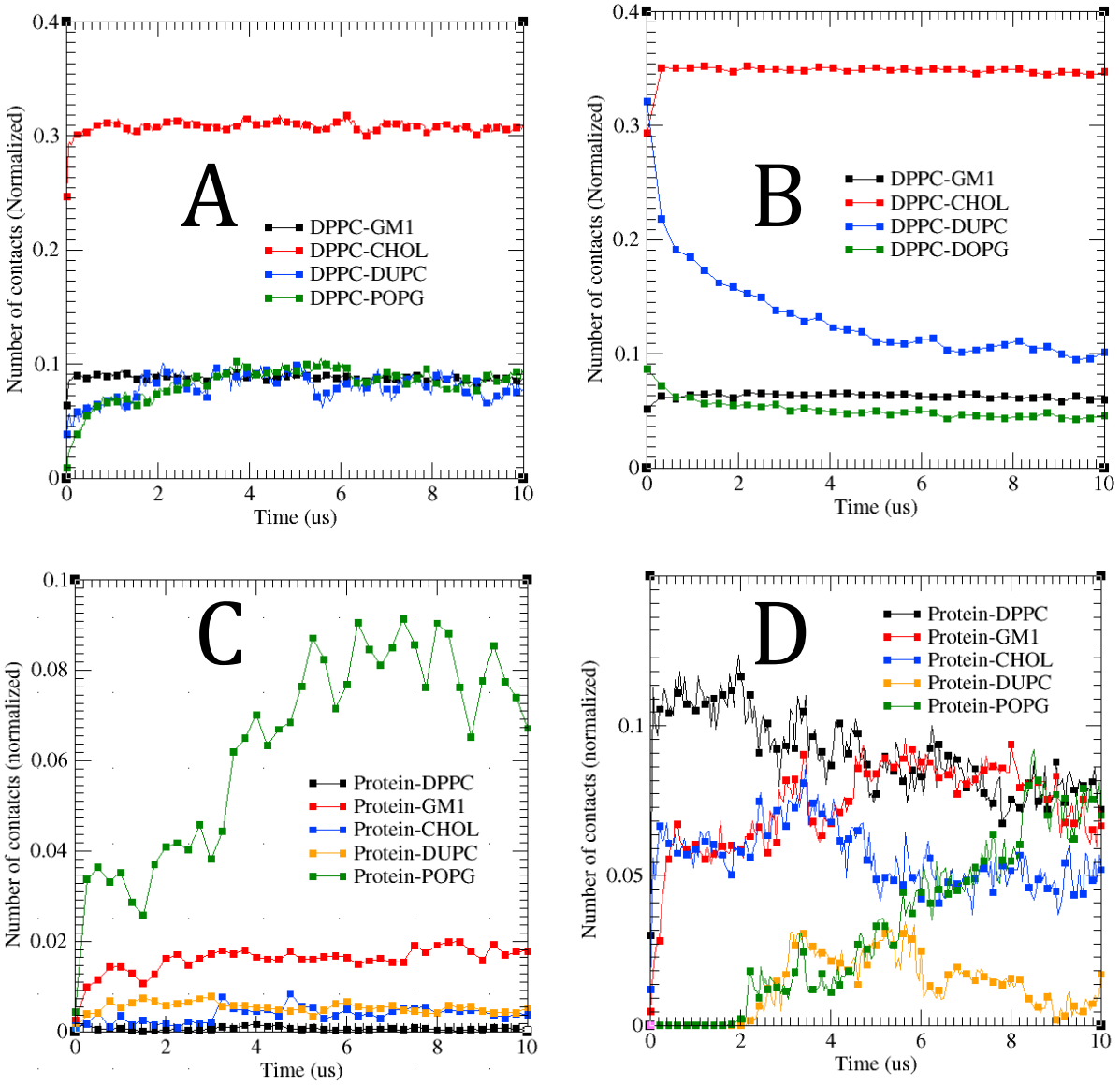

Figure S4. Normalized number of contacts in Liquid ordered-Liquid disordered membrane patches. Simulations of either pre-equilibrated (A-C) and randomized (B-D) placed lipid systems were run for 10 us. A and B plots the normalized number of contacts to the DPPC lipid. C and D highlight the contacts to the full CD3 chain (extracellular, transmembrane and cytoplasmic domains).

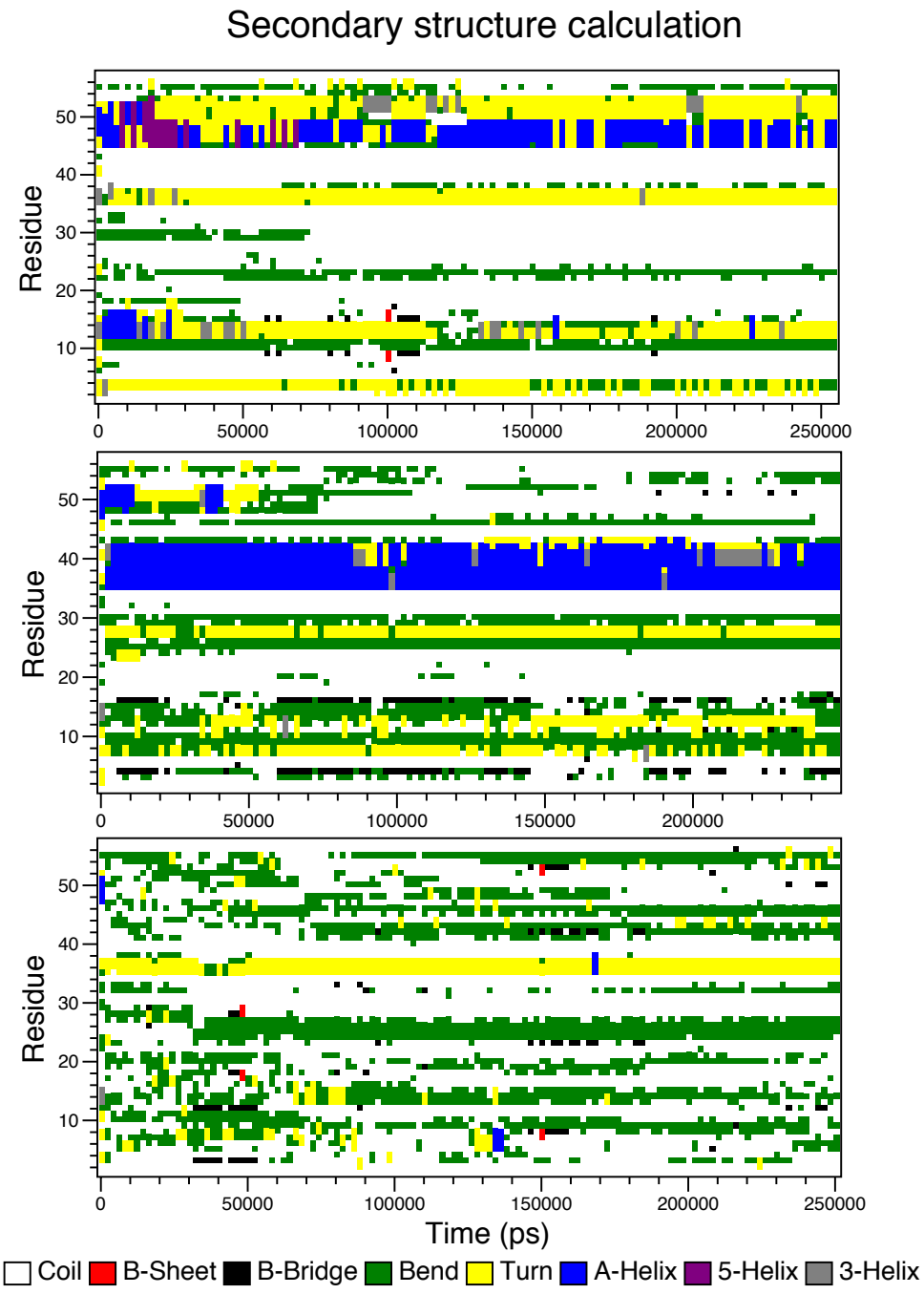

Figure S5: Averaged secondary structure of the CD3ε chain upon interacting with different lipids. The calculation was performed with the dssp program and run along the last 250 ns for each system. Alpha helices are observed when the peptide interacts with either POPG (top panel) or POPC (middle panel). However, the protein renders in a more disordered state when interacting with a pure DPPC lipid bilayer (bottom panel). Legend shows the corresponding colors for the different secondary structure patterns.

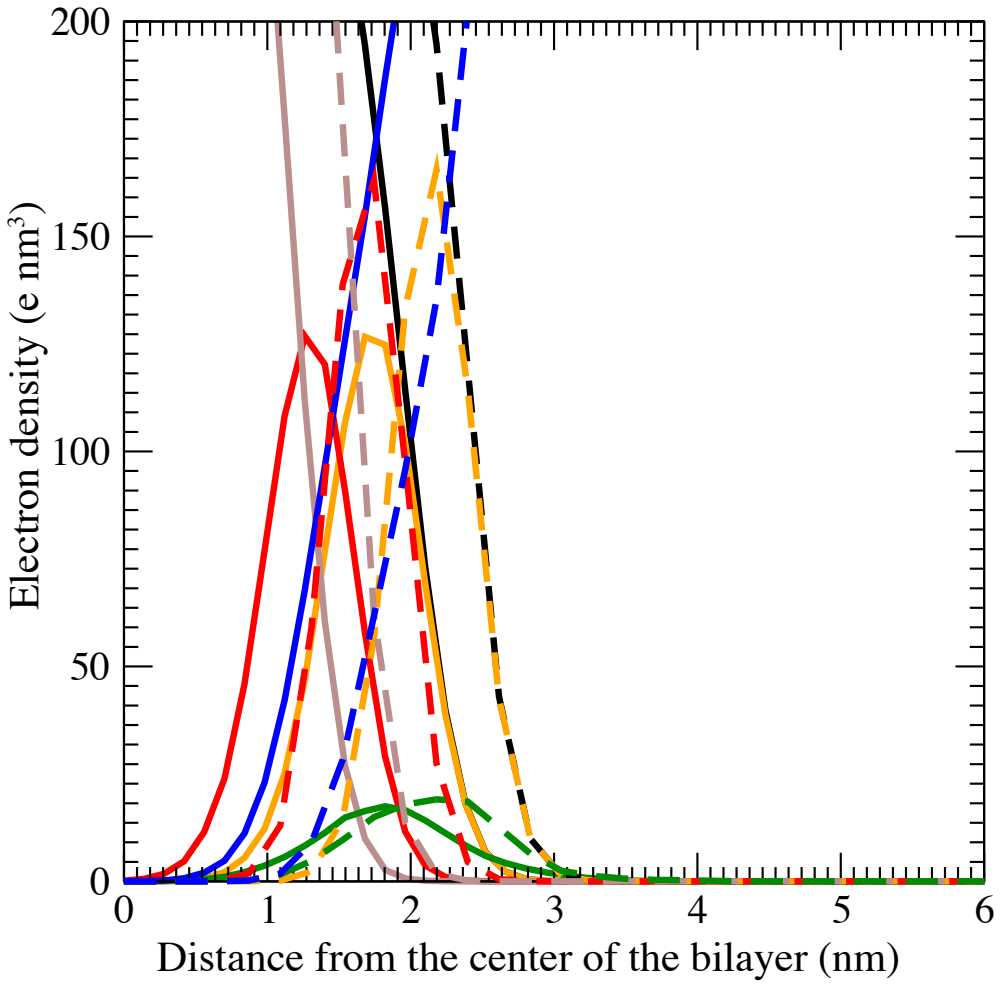

Figure S6: Comparison of the electron density profile of a pure coarse-grained POPG lipid bilayer interacting with either a cleaved cytoplasmic CD3 chain (solid lines) or a transmembrane helix attached variant (broken lines). The representative peaks are shifted by 0.5 nm in the latter case, suggesting a decrease of area per lipid. Color code: phosphate groups-orange, glycerol moiety-red, aliphatic tails-brown, water accessibility-blue. Total lipid density is represented by black lines.

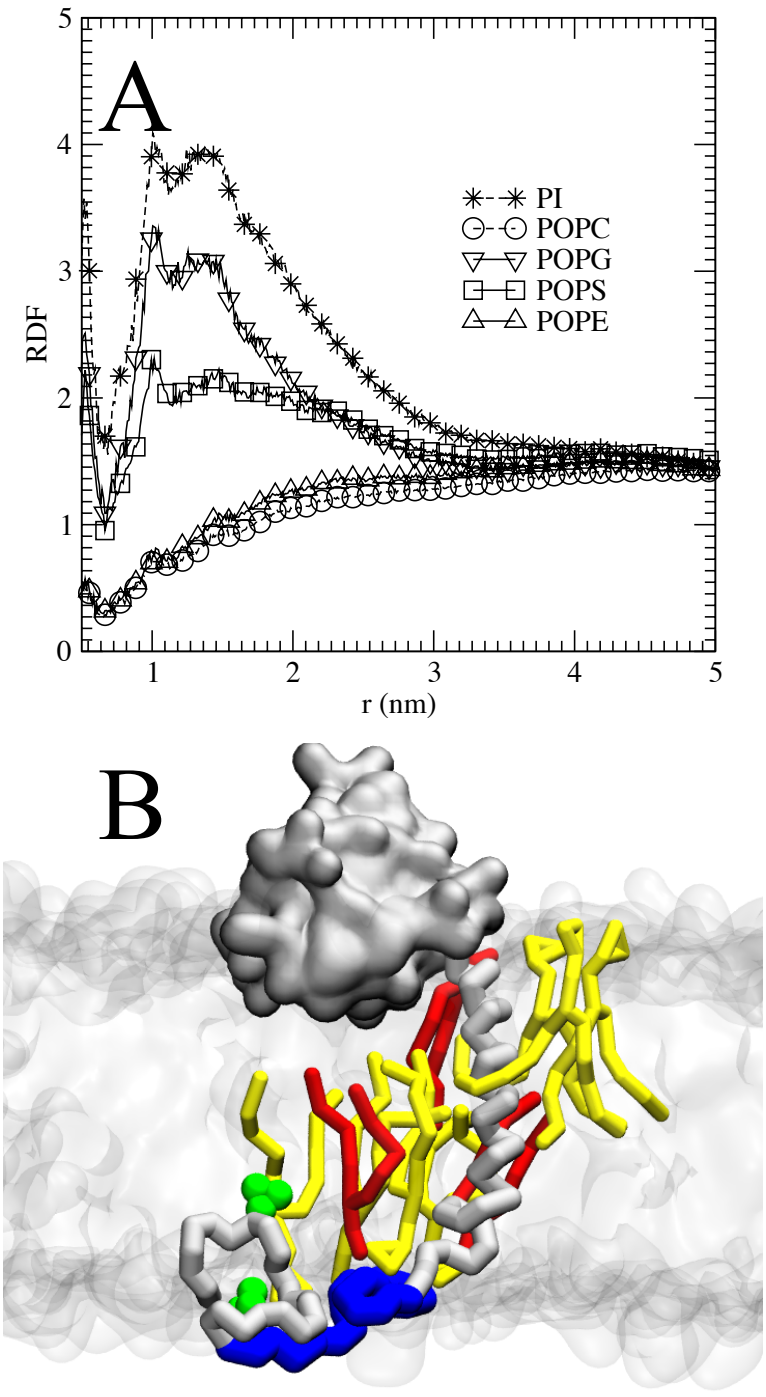

Figure S7: (A) Radial distribution function (RDF) of different lipids with respect to the CD3 chain. Calculations were performed considering the X and Y axes of the membrane plane. (B) Snapshot of the CD3 chain in interaction with PI (yellow) and POPG (red). The extracellular region of the protein is represented by a fulfilled surface. For clarity, the BRS region (blue licorice representation) and the two ITAM tyrosines (green balls) are shown.

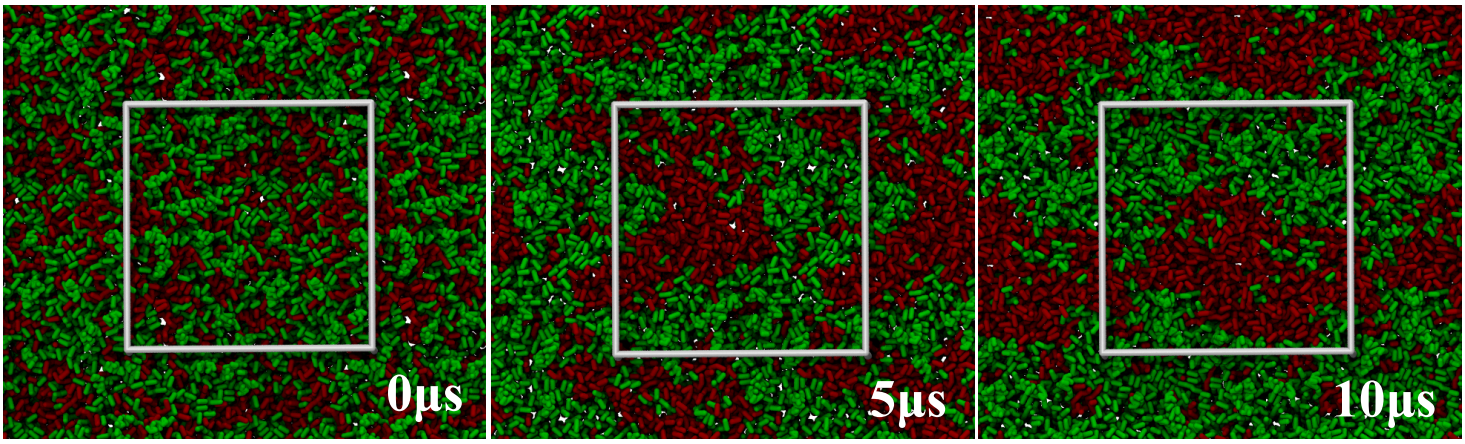

Figure S8: Liquid ordered (green)-Liquid disordered (red) domain formation of a coarse-grained quaternary lipid mixture. Four different lipids (DPPC:4, CHOL:3, DUPC:2, POPG:1) were randomly placed in a rectangular box and equilibrated for 10 μs. The temperature limit for segregation was observed at 300 K. Optimal domain formation was observed below 290 K and after 8 μs.

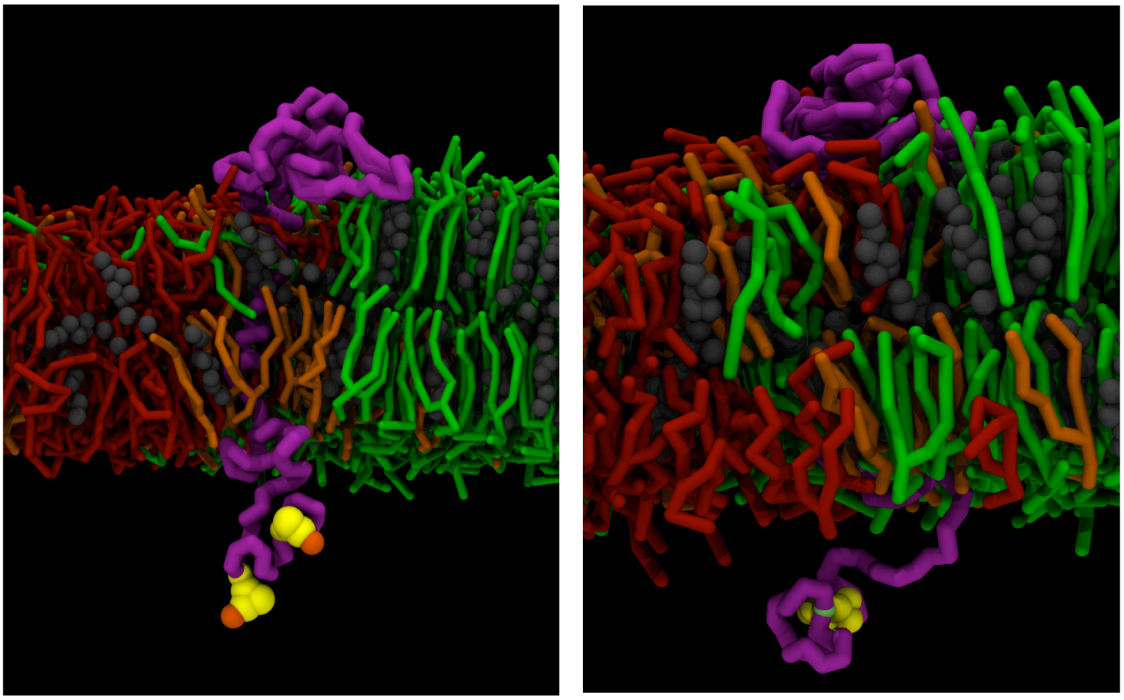

Figure S9. Snapshots of an equilibrated structure of the CD3 ε chain (Extracellular, transmembrane and cytoplasmic regions) embedded in a Liquid ordered-Liquid disordered patch. The protein (purple frame) was initially placed in the Liquid ordered domain. After 10 us, the protein was found co-localized within the Liquid ordered (green lipids) and Liquid disordered (red lipids) interface. Note that no structural difference was found when the ITAM tyrosines were phosphorylated (left panel, yellow and orange spheres). Negatively charged lipids (POPG) are depicted as orange molecules.

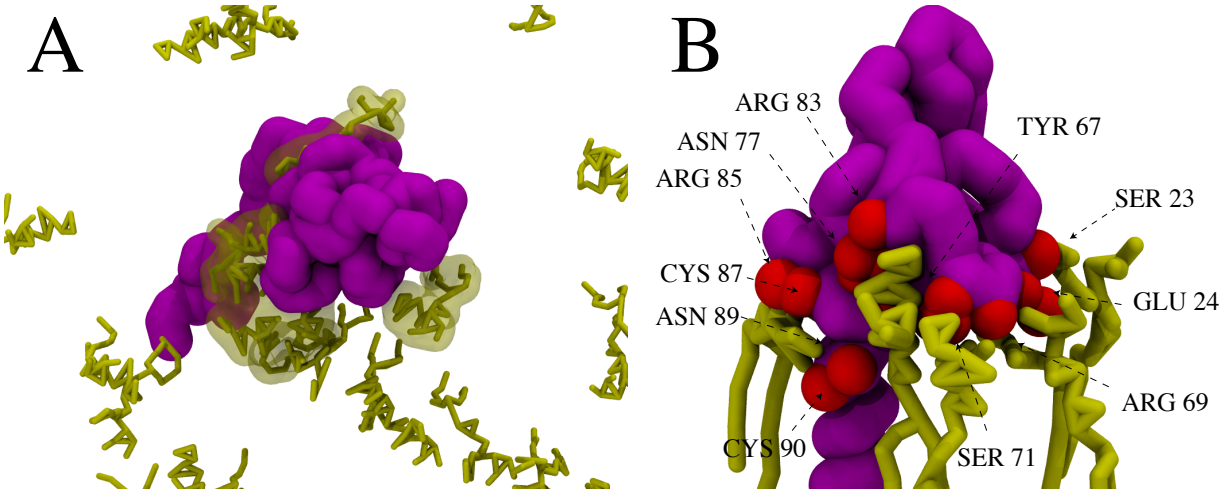

Figure S10. Interaction of the CD3ε chain with the ganglioside GM1. (A) An equilibrated configuration of the CD3 chain in close interaction with gangliosides. On average, four potential binding sites were found. (B) Specific residues in tight interaction with GM1. These tight interactions lasted throughout the entire simulation trajectory. Color code: Protein backbone: Purple, Ganglioside GM1: Yellow, Binding residues: Red balls.