The Role of a Crystallographically Unresolved Cytoplasmic Loop in Stabilizing the Bacterial Membrane Insertase YidC2

Thomas Harkey[1], Vivek Govind Kumar[1], Jeevapani Hettige[1], Seyed Hamid Tabari[1], Kalyan Immadisetty[1], and Mahmoud Moradi[1*]

[1]Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, Arkansas 72701, United States

[*]Correspondence: Email: moradi@uark.edu

ABSTRACT

YidC, a bacterial member of the YidC/Alb3/Oxa1 insertase family, mediates membrane protein assembly and insertion. Cytoplasmic loops are known to have functional significance in membrane proteins such as YidC. Employing microsecond-level molecular dynamics (MD) simulations, we show that the crystallographically unresolved C2 loop plays a crucial role in the structural dynamics of *Bacillus halodurans* YidC2. We have modeled the C2 loop and used all-atom MD simulations to investigate the structural dynamics of YidC2 in its *apo* form, both with and without the C2 loop. The C2 loop was found to stabilize the entire protein and particularly the C1 region. C2 was also found to stabilize the alpha-helical character of the C-terminal region. Interestingly, the highly polar lipid head groups of the simulated membrane were found to interact with and stabilize the C2 loop. These findings demonstrate that the crystallographically unresolved loops of membrane proteins could be important for the stabilization of the protein despite the apparent lack of structure, which could be due to the absence of the relevant lipids to stabilize them in crystallographic conditions.

Keywords: C2 loop; structural stability; microsecond-level molecular dynamics simulations; membrane proteins; lipid-protein interactions

INTRODUCTION

One third of all genes encode membrane proteins, which must be folded and inserted into the plasma membrane co-translationally[1,2]. The YidC/Oxa1/Alb3 family of membrane proteins mediates the proper folding and insertion of incoming peptides and proteins in the membrane[3,4,5,6]. Mammalian Oxa1 (found in mitochondria), plant Alb3 (found in chloroplasts), and YidC (found in bacteria) are homologous insertases[5]. Insertases exist in all domains of life and are essential for the viability of cells[7,8,9]. They are able to function either in a Sec-dependent or a Sec-independent manner[10,11,12,13,14,15,16]. In this study, we investigate the structural dynamics of YidC, which is the most well characterized member of the family.

Several studies have investigated the role of the YidC protein in different organisms. YidC plays an important role in folding of the LacY lactose permease membrane protein and is essential for the insertion of the *c* subunit of the F_0F_1-ATPase (F0c) into the plasma membrane of *Escherichia coli* (a gram-negative bacterium) in a Sec-independent manner[11,14,18]. The genomes of most gram-positive bacteria encode two YidC proteins, YidC1 and YidC2[19]. Both paralogs have functional overlap but YidC2 may have a function not shared by YidC1[20]. Multiple crystal structures of YidC are available. Kumazaki et al. have crystallized YidC2 from the gram-positive bacterium *Bacillus halodurans* and proposed a binding and insertion mechanism for single-spanning membrane proteins[5]. Here, we investigate the structural dynamics of functionally important regions of YidC2 from *Bacillus halodurans* (PDB entry: 3WO7)[5].

YidC2 consists of five transmembrane (TM) helices (TM1-5) connected by two cytoplasmic regions (C1 and C2) and two extracellular regions (E2 and E3)[5]. The C1 region consists of two helices (CH1 and CH2) connected by a short loop[21]. The insertase function

appears to be localized to the TM region[22], with TM3 being most crucial for function[23]. A positively charged arginine at position 72 (R72) along with several other hydrophilic residues form a groove that is accessible to both the cytoplasm and the lipid bilayer[5]. Figure 1 shows the position of these different domains.

Kumazaki *et al.*[5] have hypothesized that the incoming substrate first interacts with the C1 loop, and is subsequently netted into the hydrophilic groove of YidC, with negatively charged residues on the incoming substrate interacting with the positively charged R72 in the groove region[5,24]. The importance of a conserved arginine residue in the hydrophilic groove has also been reported for *E. coli* YidC[25]. High B-factor values have been reported for the C1 region, indicating that this region fluctuates greatly[5]. MD simulations also agree with the assessment that the C1 region is flexible[5]. In another study, Kumazaki *et al.* report that the C1 region is flexible in *E. coli* YidC (based on B-factor values), suggesting that the C1 region flexibility is a universal characteristic of YidC[21]. The C-terminal domain is also thought to have functional relevance and may be involved in the folding of the periplasmic regions of inserted substrates[26]. The C2 loop is not resolved in any of the YidC2 crystal structures[5]. The C2 loop of *E. coli* YidC was recently resolved in an *E. coli* YidC crystal structure[27] and suggested to play a role in the activation mechanism of YidC by covering the hydrophilic groove in its inactive state (as in the resolved structure) and by exposing the hydrophilic groove upon activation triggered by ribosome[27]. However, the C2 loop of *E. coli* YidC is considerably shorter than that of YidC2 in gram-positive bacteria, and the two may have different functional roles due to the significant difference in their length.

The importance of cyto- and periplasmic loops for the functioning of membrane proteins has been studied extensively in various proteins such as NS4B[28], Wzy[29], and LolCDE[30]. Due to

the importance of cytoplasmic loops in membrane proteins, we have modeled the crystallographically unresolved C2 loop of YidC2 in order to investigate the impact of this loop on overall protein stability as well as the stability of functionally important regions such as the C1 loop and the C-terminal domain. All-atom MD provides a reliable method for the investigation of membrane protein structural dynamics particularly in a comparative manner[31]. Nanosecond-level MD simulations, however, as are often used for such studies, have been shown to be questionable for reliably describing the functionally relevant conformational dynamics of membrane proteins[32]. Therefore, either microsecond-level[31] and/or enhanced sampling[33,34,35] simulations must be employed to investigate the conformational behavior of membrane proteins. Here we have employed microsecond-level all-atom MD simulations of membrane-embedded YidC2 that show a key role for the C2 cytoplasmic loop in the structural dynamics of YidC2.

RESULTS AND DISCUSSION

We have performed two sets of unbiased all-atom MD simulations of YidC2, each for two microseconds. The initial models are both based on the crystal structure of YidC2 (PDB entry: 3WO7)[5] in the presence of explicit membrane and water. The missing C2 loop was modeled carefully in one of the simulation sets and was not present in the other (see Methods for simulation details). Since the two systems are virtually identical besides the presence/absence of the C2 loop, we can make meaningful comparisons between the two sets of simulations. Any difference that is observed in the flexibility of the protein in each system can be attributed to the presence or absence of the C2 loop, given that the two trajectories are converged.

We have identified several regions of interest in YidC2 to characterize the conformational

dynamics of both systems. Figure 1 identifies some of these regions in the system with the modeled C2 loop. These regions include the C1 region, the C2 modeled loop, the TM helices, and the carboxyl terminal region. The C1 region (including the C1 loop and CH1/CH2 helices), carboxyl terminal region, and C2 loop all reside in the cytoplasmic area.

The presence of the C2 loop stabilizes the global protein conformation

Comparing the C_α root mean square deviation (RMSD) of the two systems indicates that the presence of the C2 loop stabilizes the protein (Figure 2A). The RMSD of YidC2 was calculated and plotted as a function of simulation time. The system without the loop is clearly less stable (RMSD=3.0±0.41 Å) than the system with the loop (RMSD=2.4±0.40 Å) (Figure 2A). Intriguingly, root mean square fluctuation (RMSF) analysis indicates that the stabilization is not localized but a large part of the protein shows a more rigid conformation in the presence of the C2 loop, evidenced by a lower value of RMSF for most residues (Figure 2B). The difference in the RMSF values, however, is not the same for all domains, indicating that some domains are more directly influenced by the presence of the C2 loop. Overall, the protein shows a relatively dynamic behavior in both cases (Figure 2A), however, our results indicate that the flexibility of the protein is considerably exaggerated when the C2 loop is not considered, as revealed more clearly using the principal component analysis (PCA).

We performed PCA to verify our claim that the C2 loop stabilizes the protein (Figure 3A). When both of the trajectories are projected onto the space of their first two principal components (PC1 and PC2), it clearly demonstrates that the system with the loop clusters tightly around a specific region as compared to the system without the loop that samples a number of

scattered regions in the configuration space. This clearly indicates that the system with the loop is more stable than the system without the loop.

The C2 loop increases the stability of the C1 loop

We observe that in the system with the C2 loop, the C1 region is significantly more stable than it is in the system without the C2 loop (Figure 2B). Since the only difference between the two systems is the presence of the C2 loop, we conclude that the presence of the C2 loop stabilizes the C1 region of YidC2. RMSD analysis of the C1 region confirms that the difference in C1 fluctuations is due to the presence of the C2 loop. We calculated the overall and internal RMSD of the C1 loop (Supplemental Figure S1). The former reflects the internal conformational changes plus the translational and rotational motions, while the latter only reflects the internal conformational changes. The overall RMSD for the C1 region without the C2 loop (3.5±0.90 Å) is greater than that with the C2 loop (2.9±0.61 Å) (Supplemental Figure S1A), confirming that the presence of the C2 loop stabilizes the C1 region of YidC. However, there is no significant difference in the behavior of the internal RMSD of the two systems (1.6±0.33Å vs. 1.7±0.30 Å) (Supplemental Figure S1B). A comparison of the internal and overall RMSDs of the C1 region indicates that while the internal conformation of the C1 region stays close to that captured in the crystal structure of YidC2, its orientation does not stay the same in the two cases. PCA analysis illustrates this observation more clearly. The most significant protein collective motion (represented by PC1) in each system is associated with a distinct motion in the C1 region (Figure 3B), which clearly illustrates a difference in behavior of the C1 region between the two systems. The C1 region moves upward towards the C2 loop and becomes stable when the C2 loop is

present. However, in the absence of the C2 loop, the C1 region can move downward away from the membrane and fluctuate strongly.

Salt bridge interactions play a key role in the stabilization of the C1 region by the C2 loop

The difference in the flexibility of YidC2 with and without the C2 loop is particularly pronounced in the C1 region as discussed above. This is due, to a great extent, to the electrostatic interactions between the C1 region and C2 loop. The negatively charged side chain of D205 in the C2 loop can specifically interact with K104 and K109 of the C1 loop (Figure 4A), although the interactions are stronger and more frequent between D205 and K109, as reflected in the salt bridge distance time series (Figure 4B). This is expected considering that D205 is closer to K109 than K104. These salt bridges cannot form in the system without the C2 loop. Therefore, we propose that these salt bridges play a key role in stabilizing the C1 region in the presence of the C2 loop and that in the system without the C2 loop, the C1 region is anchored less strongly to the rest of the protein and is able to fluctuate in the cytoplasm more freely.

We also identified an intradomain salt bridge in the CH1 region of the C1 loop between R93 and E97. The E97-R93 salt bridge distance was significantly higher in the system without the loop (5.6±1.7 Å) (Supplemental Figure S2A,B) than the system with the loop (4.3±0.64 Å) (Supplemental Figure S2C,D). We hypothesize that this salt bridge is destabilized in the system without the loop due to the fluctuating behavior of the C1 region in the absence of the C2 loop. These findings clearly indicate a relationship between the functionally important C1 region and the neighboring C2 loop in the cytoplasmic side of YidC2. Our findings are in correlation with those of Geng *et al.*[26], where they studied the C2 mediated YidC-ribosome binding in *E. coli* and

determined that the C2 region of YidC was involved in ribosome binding, but not the C1 region. They proposed that the C1 region may be involved in downstream activity but not ribosome binding. They also found that the positively charged C-terminus of YidC does play a significant role in the ribosome binding.

Carboxyl-terminal domain stability and conformation

The C-terminal domain of YidC has been proposed to play a significant role in ribosome binding[26]. Due to its potential functional importance, we probed the conformational dynamics of the C-terminal region of YidC2. The available crystal structure of YidC2[5] has a modified C-terminal domain with several missing and mutated residues (see Methods). However, we observed a significant difference in its secondary structure in the absence and presence of the C2 loop, an observation that could indicate a potential allosteric interaction between the C-terminal domain and the C2 loop. Over the course of the simulation without the C2 loop, the C-terminal region appears to unravel into a random coil (Figure 5A) as evidenced by the secondary structure analysis, where it completely loses its α-helical character within the first 0.3 μs of simulation (Figure 5B). On the other hand, the C-terminal region exists as an α-helix for the majority of the simulation in the system with the loop (Figure 5A,B). This indicates that the C2 loop stabilizes the C-terminal α-helix.

We also analyzed the overall and internal RMSD of the C-terminal domain to characterize its flexibility. The presence of the C2 loop corresponds to a more stable C-terminal domain (overall RMSD = 4.7±1.6 Å) as compared to the model without the C2 loop (overall RMSD = 7.4±1.9 Å) (Supplemental Figure S3A). The internal RMSD for the systems with and without the loop are

2.0±0.81 Å and 3.6±0.62 Å, respectively (Supplemental Figure S3B), demonstrating that there is a difference in internal conformation of the carboxyl terminus regions between the two systems, as also evidenced by our secondary structure analysis. These observations confirm the role of the C2 loop in promoting the helical structure of the C-terminal domain.

C2 loop allosterically impacts the behavior of other functionally important regions of YidC2

Thus far, we showed that the C2 loop directly interacts with the C1 region and changes its behavior, while the C-terminal region is also influenced by the presence or absence of the C2 loop. The C-terminal region, however, does not interact directly with the C2 loop. Instead, the C2 loop is affecting the behavior of the C-terminal region indirectly through the C1 region as we will discuss in more detail below.

To systematically investigate the allosteric interactions of C2 loop with different protein domains, we employed dynamic network analysis[36], which characterizes the linear correlation between different residue pairs. Figure 6 shows the correlation coefficient of each residue pair calculated from the trajectory with the C2 loop, subtracted from the same quantity calculated from the trajectory without the loop, and reported as its absolute value. The reported quantity for each residue pair quantifies the magnitude of the difference in the correlation behavior of the two residues caused by the C2 loop. We observe that the introduction or removal of the C2 loop leads to significant changes in the correlations of different domains of YidC2. Specifically, differences in inter-domain correlations between TM1/C1 region and TM3/TM4 region as well as intra-domain correlations of the TM1 helix are quite significant (Fig. 6). However, the specific residues that exhibit the most significant C2-dependent behavior in their correlations with other

residues are located in the carboxyl tail that interact strongly with various residues in the C1 region and TM4 helix (Supplemental Fig. S4).

The influence of C2 loop on the C-terminal tail is clear from the analyses above and since there is no direct interaction between the two domains, it is likely that the C1 region mediates the change in the conformation and interactions of the C-terminal domain. For instance, a salt bridge can form between E266 (C-terminal region) and K81 (C1 region) in the system without the C2 loop, which is completely absent in the system with the C2 loop (Figure 7). In the absence of the C2 loop, the C1 region can more freely move and interact with the C-terminal tail. The attraction between the E266 and K81 residues can pull the C-terminal close to the C1-TM1 region. The interaction between the C1 region and the C-terminal domain is likely to be the cause of the disruption of its helical structure, accounting for the difference in C-terminal flexibility between the two systems. E266 in the C-terminal tail is able to move closer to K81 due to the fact that the C1 region, TM4, and TM5 helices of the protein are not interacting/connected to the missing C2 loop. Overall, these findings indicate that the presence of the C2 loop influences the behavior of the functionally important regions of YidC2.

Interactions with the membrane contributes to the stability of the C2 loop

The absence of the C2 loop not only results in some local conformational changes in YidC2, as discussed above, but also induces global changes. Interestingly, the protein is placed in the membrane somewhat differently in the absence and presence of the C2 loop. Figure 8A illustrates how the tilt angle of the protein with respect to the membrane normal is distributed differently when the C2 loop is present or absent. The difference in tilt angle is most likely due

to the C2 loop membrane interactions. Our simulations provide evidence for such interactions, particularly between the side chains of the C2 loop and the lipid head groups. Specifically, residue D207 of the C2 loop forms one or more hydrogen bonds with the lipid head groups throughout the simulation (Figure 8B). Our interaction energy analysis indicates that the C2-membrane interactions, which are predominantly electrostatic, are steady throughout the simulation (Figure 8C). The interactions between the POPE head groups and the C2 loop provides a possible mechanism for the stabilization of the C2 loop.

Phospholipids are specific participants in determining membrane protein organization[37]. We have recently shown that a slight change in the polarity of the head groups of membrane's constituent lipids could substantially change the structural dynamics of a transmembrane protein[31]. Specifically, the PE head groups were shown to play a key role in the structural dynamics of a bacterial ATP-binding cassette transporter[31]. The importance of the membrane composition has been shown extensively for various membrane proteins and other membrane-related phenomena[38,39,40,41]. It is thus reasonable to assume that the absence of the phospholipids in the crystallographic conditions could result in deviations of the resolved protein structure from its native conformation. This is also consistent with previous MD simulations of *E. coli* YidC, where the structure of protein in a POPE containing lipid bilayer showed a more compact conformation compared with the crystallographic structure[25].

Earlier we mentioned that the C2 loop was not resolved in the crystal structure of YidC2, where the crystals were grown in a lipidic cubic phase, using monoolein lipids[5], which are quite different from phospholipids. The absence of the phospholipids during the crystallization process of YidC2 could explain the fact that the C2 loop was not resolved. We propose that the absence of a physiologically relevant membrane in the crystallization process could cause the C2 loop to

appear disordered. We note that the recently resolved *E. coli* YidC crystal structure contains a C2 loop, which is much shorter than that of the YidC2 C2 loop in gram-positive bacteria. The C1 and C2 loops in the resolved structure of *E. coli* YidC do not seem to be able to interact as in our YidC2 model. This observation is consistent with the fact that the two YidC proteins have significant functional differences.

We conclude that the second cytoplasmic loop of YidC2 could have a functional role perhaps by stabilizing the protein structure not only through its direct interactions with the C1 region and transmembrane helices TM4 and TM5 but also through its indirect effect on other transmembrane helices (particularly TM1) and the C-terminal region. The C2 loop of YidC2 is also significant due to its proximity to the periphery of the membrane on the cytoplasmic side and its strong interactions with the lipid head groups, a feature which is potentially absent in gram-negative YidC proteins. Due to its interactions with the membrane, the C2 loop was also found to change the tilt angle of the protein within the membrane. The C2 loop forms a salt bridge network with the functionally important C1 region and reduces its flexibility. The presence of the C2 loop also appears to reduce the flexibility of the carboxyl terminal region of the protein by increasing its helical propensity. Further research is needed to elucidate the importance of the C2 loop in the sec-independent insertion mechanism of small single-spanning membrane proteins such as the pf3 coat protein, whose interactions with the C1 region has been proposed to be crucial[5]. In the context of molecular dynamics simulations, our study suggests that modeling crystallographically unresolved loops may be necessary for accurate description of membrane protein dynamics, particularly if the missing loops are likely to interact with the periphery of the membrane.

METHODS

We have used brute-force all-atom MD simulations to characterize the conformational transitions of bacterial YidC2 in a modeled membrane environment. We built two YidC2 systems; one with the C2 loop and another without the C2 loop. The crystal structure of bacterial YidC2 from *Bacillus halodurans* (PDB entry: 3WO7) [5] with the missing C2 loop was initially processed using the Molecular Operating Environment (MOE)[42] software to remove the crystal waters and assign the appropriate protonation states for the titratable residues at the physiological pH (7.4) using protonate3D facility of MOE. For the system with the C2 loop, a Monte Carlo algorithm was used to model the missing C2 loop using the program Modeller[43] to iteratively minimize the energy of the system. The initial loop generated by Modeller was substantially different from the converged loop. The overall shape of the loop was converged within about 1000 iterations, but we continued the Monte Carlo iterations for another 9,000 steps (10,000 in total) to ensure a reliable convergence. CHARMM-GUI[44,45,46] was then used to build the simulation systems. The protein was placed in 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) lipids, solvated in a box of TIP3P waters, and 0.15 M NaCl (in addition to the counterions used to neutralize the protein) using CHARMM-GUI[44,45,46]. The system with the loop consisted of one protein, 90/88 POPE lipids in the upper/lower leaflet, 11,797 TIP3P waters, and 36/38 sodium/chloride ions with a total of 61,994 atoms in a box size of ~98×96×100 $Å^3$. The system without the loop consisted of 90/96 POPE lipids in the upper/lower leaflet, 11,827 TIP3P waters, and 35/38 sodium/chloride ions with a total of 62,254 atoms in a box size of ~ 92×96×101 $Å^3$.

Both systems were simulated with NAMD 2.10-12[47] and the CHARMM36 all-atom additive force field48[37,48]. Initially each system was energy-minimized for 10,000 steps using the conjugate gradient algorithm[49]. Then, we relaxed the systems by releasing the restraints in a stepwise manner (for a total of ~1 ns) using a procedure that is explained elsewhere[44]. Production runs were performed for 2 μs for each system (total 4 μs of simulation data). The initial relaxation was performed in an NVT ensemble while all production runs were performed in an NPT ensemble. Simulations were carried out using a 2-fs time step at 310 K using a Langevin integrator with a damping coefficient of $\gamma = 0.5$ ps^{-1}. The pressure was maintained at 1 atm using the Nosé−Hoover Langevin piston method[49,50]. The smoothed cutoff distance for non-bonded interactions was set to 10−12 Å and long-range electrostatic interactions were computed with the particle mesh Ewald (PME) method[51]. The trajectories were collected every 0.5 ns, resulting in 4,000 data points for each system for statistical analysis. Supercomputers Razor and Bridges at the University of Arkansas and Pittsburgh Supercomputing Center, respectively, were used to run the simulations.

The TM helices and other subdomains were defined as follows: TM1 (63-83), TM2 (134-155), TM3 (175-190), TM4 (219-233), TM5 (233-258), C1 region (84 to 133), C2 loop (195 to 216), and modified C-terminal region (256 to 272). The last 13 C-terminal residues (268-280) were deleted from the wild-type sequence and replaced by a tag[5]. Residues 268 to 272 in the YidC2 model (modified C-terminal domain) used in our study belong to the tag.

The RMSD trajectory tool of VMD[52] was used to calculate the RMSD and C_α atoms were considered for these calculations. For overall RMSD, the protein was aligned against the crystal structure and RMSD was calculated for the region of interest with respect to its initial configuration. For internal RMSD, we aligned the region of interest against its own initial

configuration and calculated RMSD with respect to this configuration. For calculating the average RMSD, the entire trajectory was considered and error bars represent the standard deviation in the data. RMSF of individual residues was calculated using the C_α atoms by aligning the trajectory against the crystal structure. The VMD timeline plugin[52] was used to identify salt bridges and the cutoff distance used was 3.5 Å. The salt bridge plugin of VMD[52] was used to calculate the distance between the two salt bridge residues over the course of the simulation, which is the distance between the oxygen atom of the participating acidic residue and the nitrogen atom of the basic residue. PRODY software[53] was used to carry out the PCA analysis. Only C_α atoms were used for PCA calculations. The VMD HBond plugin[52] was used for hydrogen bond analysis; the cut-off distance and angles used were 3.5Å and 30° respectively. Dynamic network analysis was carried out using the dynamic network analysis tool in VMD[36] and the program Carma[54]. In brief, the correlation of a residue pair i and j is defined as:

$$C_{ij} = \frac{\langle \Delta\vec{r}_i(t).\Delta\vec{r}_j(t)\rangle}{\sqrt{\langle \Delta\vec{r}_i(t)^2\rangle\langle \Delta\vec{r}_j(t)^2\rangle}},$$

where $\Delta\vec{r}_i(t) = \vec{r}_i(t) - \langle\vec{r}_i(t)\rangle$, $\vec{r}_i(t)$ is the position of C_α atom of residue i at time t, and $\langle\cdot\rangle$ is an average over all t. C_{ij} quantifies the linear correlation of the motion of C_α atoms of residues i and j. Values 1.0 and -1.0 indicate strongest positive and negative correlations possible and 0.0 indicates the complete lack of any linear correlations. If C_{ij} and C'_{ij} are measured under two different simulation conditions (e.g., YidC2 with and without the C2 loop), $|C_{ij} - C'_{ij}|$ (the absolute value of the difference in the correlations between the two conditions) quantifies the absolute change due to the change in the condition (e.g., introduction/removal of the C2 loop in YidC2).

ACKNOWLEDGMENTS

We thank the University of Arkansas, Fayetteville and the Arkansas Biosciences Institute for funding. This research was also supported by the Arkansas High Performance Computing Center, which was funded through multiple NSF grants and the Arkansas Economic Development Commission. This work also used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. This work used Bridges at the Pittsburgh Supercomputing Center through XSEDE allocation MCB150129. We also thank Colin Heyes, James Walker, Mack Ivey, Adithya Polasa, and Mitchell Benton for helpful discussions.

AUTHOR CONTRIBUTIONS

T.H., J.H., and M.M conceived the project and designed the simulations. T.H. conducted the simulations. T.H., J.H., S.H.T., V.G.K. and K.I. analyzed the data. T.H., V.G.K., K.I., and M.M. prepared the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

AVAILABILITY OF DATA

The molecular dynamics trajectories and the analyses generated will be shared upon request to corresponding author.

REFERENCES

1. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J. Mol. Biol. *305*, 567-580.

2. Rapoport, T.A. (2007). Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. Nature *450*, 663-669.

3. Jiang, F.L., Yi, L., Moore, M., Chen, M.Y., Rohl, T., van Wijk, K.J., de Gier, J.W.L., Henry, R., and Dalbey, R.E. (2002). Chloroplast YidC homolog Albino3 can functionally complement the bacterial YidC depletion strain and promote membrane insertion of both bacterial and chloroplast thylakoid proteins. J. Biol. Chem. *277*, 19281-19288.

4. Kedrov, A., Sustarsic, M., de Keyzer, J., Caumanns, J.J., Wu, Z.C., and Driessen, A.J.M. (2013). Elucidating the Native Architecture of the YidC: Ribosome Complex. J. Mol. Biol. *425*, 4112-4124.

5. Kumazaki, K., Chiba, S., Takemoto, M., Furukawa, A., Nishiyama, K., Sugano, Y., Mori, T., Dohmae, N., Hirata, K., Nakada-Nakura, Y.*, et al.* (2014a). Structural basis of Sec-independent membrane protein insertion by YidC. Nature *509*(7501), 516-520.

6. Samuelson, J.C., Chen, M.Y., Jiang, F.L., Moller, I., Wiedmann, M., Kuhn, A., Phillips,

G.J., and Dalbey, R.E. (2000). YidC mediates membrane protein insertion in bacteria. Nature *406*, 637-641.

7. Borowska, M.T., Dominik, P.K., Anghel, S.A., Kossiakoff, A.A., and Keenan, R.J. (2015). A YidC-like Protein in the Archaeal Plasma Membrane. Structure *23*, 1715-1724.

8. Dalbey, R.E., and Kuhn, A. (2015). Membrane Insertases Are Present in All Three Domains of Life. Structure *23*, 1559-1560.

9. Kuhn, A., and Kiefer, D. (2017). Membrane protein insertase YidC in bacteria and archaea. Mol. Microbiol. *103*, 590-594.

10. van Bloois, E., ten Hagen-Jongman, C.M., and Luirink, J. (2007). Flexibility in targeting and insertion during bacterial membrane protein biogenesis. Biochem. Biophys. Res. Commun. *362*, 727-733.

11. van Bloois, E., Nagamori, S., Koningstein, G., Ullers, R.S., Preuss, M., Oudega, B., Harms, N., Kaback, H.R., Herrmann, J.M., and Luirink, J. (2005). The Sec-independent function of Escherichia coli YidC is evolutionary-conserved and essential. J. Biol. Chem. *280*, 12996-13003.

12. Dalbey, R.E., and Kuhn, A. (2004). YidC family members are involved in the membrane insertion, lateral integration, folding, and assembly of membrane proteins. J. Cell Biol.

166, 769-774.

13. Facey, S.J., and Kuhn, A. (2004). Membrane integration of E-coli model membrane proteins. BBA-Mol. Cell Res. *1694*, 55-66.

14. Lewis, N.E., and Brady, L.J. (2015). Breaking the bacterial protein targeting and translocation model: oral organisms as a case in point. Mol. Oral Microbiol. *30*, 186-197.

15. Scotti, P.A., Urbanus, M.L., Brunner, J., de Gier, J.W.L., von Heijne, G., van der Does, C., Driessen, A.J.M., Oudega, B., and Luirink, J. (2000). YidC, the Escherichia coli homologue of mitochondrial Oxa1p, is a component of the Sec translocase. EMBO J. *19*, 542-549.

16. Urbanus, M.L., Scotti, P.A., Froderberg, L., Saaf, A., de Gier, J.W.L., Brunner, J., Samuelson, J.C., Dalbey, R.E., Oudega, B., and Luirink, J. (2001). Sec-dependent membrane protein insertion: sequential interaction of nascent FtsQ with SecY and YidC. EMBO Rep. *2*, 524-529.

17. Nagamori, S., Smirnova, I.N., and Kaback, H.R. (2004). Role of YidC in folding of polytopic membrane proteins. J. Cell Biol. *165*, 53-62.

18. Serdiuk, T., Balasubramaniam, D., Sugihara, J., Mari, S.A., Kaback, H.R., and Muller, D.J. (2016). YidC assists the stepwise and stochastic folding of membrane proteins. Nat.

Chem. Biol. *12*, 911-917.

19. Funes, S., Hasona, A., Bauerschmitt, H., Grubbauer, C., Kauff, F., Collins, R., Crowley, P.J., Palmer, S.R., Brady, L.J., and Herrmann, J.M. (2009). Independent gene duplications of the YidC/Oxa/Alb3 family enabled a specialized cotranslational function. Proc. Natl. Acad. Sci. U.S.A. *106*, 6656-6661.

20. Dalbey, R.E., Kuhn, A., Zhu, L., and Kiefer, D. (2014). The membrane insertase YidC. BBA-Mol. Cell Res. *1843*, 1489-1496.

21. Kumazaki, K., Kishimoto, T., Furukawa, A., Mori, H., Tanaka, Y., Dohmae, N., Ishitani, R., Tsukazaki, T., and Nureki, O. (2014b). Crystal structure of Escherichia coli YidC, a membrane protein chaperone and insertase. Sci. Rep. *4*.

22. Oliver, D.C., and Paetzel, M. (2008). Crystal structure of the major periplasmic domain of the bacterial membrane protein assembly facilitator YidC. J. Biol. Chem. *283*, 5208-5216.

23. Yuan, J., Phillips, G.J., and Dalbey, R.E. (2007). Isolation of cold-sensitive yidC mutants provides insights into the substrate profile of the YidC insertase and the importance of transmembrane 3 in YidC function. J. Bacteriol. *189*, 8961-8972.

24. Kol, S., Nouwen, N., and Driessen, A.J.M. (2008). Mechanisms of YidC-mediated

Insertion and Assembly of Multimeric Membrane Protein Complexes. J. Biol. Chem. *283*, 31269-31273.

25. Chen, Y.Y., Capponi, S., Zhu, L., Gellenbeck, P., Freites, J.A., White, S.H., and Dalbey, R.E. (2017). YidC Insertase of Escherichia coli: Water Accessibility and Membrane Shaping. Structure *25*, 1403-1414.

26. Geng, Y.P., Kedrov, A., Caumanns, J.J., Crevenna, A.H., Lamb, D.C., Beckmann, R., and Driessen, A.J.M. (2015). Role of the Cytosolic Loop C2 and the C Terminus of YidC in Ribosome Binding and Insertion Activity. J. Biol. Chem. *290*, 17250-17261.

27. Tanaka, Y., Izumioka, A., Abdul Hamid, A., Fujii, A., Haruyama, T., Furukawa, A., and Tsukazaki, T. (2018) 2.8-Å crystal structure of Escherichia coli YidC revealing all core regions, including flexible C2 loop. Biochem. Biophys. Res. Commun. *505*(1), 141-145.

28. Zou, J., Lee, L.T., Wang, Q.Y., Xie, X.P., Lu, S.Y., Yau, Y.H., Yuan, Z.M., Shochat, S.G., Kang, C.B., Lescar, J.*, et al.* (2015). Mapping the Interactions between the NS4B and NS3 Proteins of Dengue Virus. J. Virol. *89*, 3471-3483.

29. Islam, S.T., Huszczynski, S.M., Nugent, T., Gold, A.C., and Lam, J.S. (2013). Conserved-residue mutations in Wzy affect O-antigen polymerization and Wzz-mediated chain-length regulation in Pseudomonas aeruginosa PAO1. Sci. Rep. *3*.

30. Yasuda, M., Iguchi-Yokoyama, A., Matsuyama, S.I., Tokuda, H., and Narita, S.I. (2009). Membrane Topology and Functional Importance of the Periplasmic Region of ABC Transporter LolCDE. Biosci. Biotechnol. Biochem. *73*, 2310-2316.

31. Immadisetty K., Hettige, J., and Moradi, M. (2019). Lipid-Dependent Alternating Access Mechanism of a Bacterial Multidrug ABC Exporter. ACS Cent. Sci. *5* (1), 43–56.

32. Immadisetty K., Hettige, J., and Moradi, M. (2017). What Can and Cannot Be Learned from Molecular Dynamics Simulations of Bacterial Proton-Coupled Oligopeptide Transporter GkPOT? J. Phys. Chem. B *121*, 3644-3656.

33. Moradi, M., and Tajkhorshid, E. (2013). Mechanistic picture for conformational transition of a membrane transporter at atomic resolution. Proc. Natl. Acad. Sci. U.S.A. *110*, 18916-18921.

34. Moradi, M., Enkavi, G., and Tajkhorshid, E. (2015). Atomic-level characterization of transport cycle thermodynamics in the glycerol-3-phosphate:phosphate antiporter. Nat. Commun. *6*, 8393.

35. Singharoy, A., Chipot, C., Moradi, M., and Schulten, K. (2017). Chemomechanical coupling in hexameric protein–protein interfaces harness energy within V–type ATPases. J. Am. Chem. Soc. *139*, 293-310.

36. Sethi, A., Eargle, J., Black, A.A., and Luthey-Schulten, Z. (2009). Dynamical networks in tRNA: protein complexes. Proc. Natl. Acad. Sci. U.S.A. *106*, 6620-6625.

37. Klauda, J.B., Venable, R.M., Freites, J.A., O'Connor, J.W., Tobias, D.J., Mondragon-Ramirez, C., Vorobyov, I., MacKerell, A.D., and Pastor, R.W. (2010). Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. J. Phys. Chem. B *114*, 7830-7843.

38. Kasson, P.M., and Pande, P.S. (2007) Control of Membrane Fusion Mechanism by Lipid Composition: Predictions from Ensemble Molecular Dynamics. PLoS Comp. Biol., 3, e220.

39. Balusek, C., and Gumbart, J.C. (2016). Role of the Native Outer-Membrane Environment on the Transporter BtuB. Biophys. J 111, 1409-1417.

40. Tse, C.H., Comer, J., Wang, Yi, and Chipot, C. (2018) Link between Membrane Composition and Permeability to Drugs. J. Chem. Theory Comput., 14, 2895–2909.

41. Muller, M., Jiang, T., Sun, C., Lihan, M., Pant, S., Mahinthichaichan, P., Trifan, A., Tajkhorshid, E. (2019) Characterization of Lipid–Protein Interactions and Lipid-Mediated Modulation of Membrane Protein Function through Molecular Simulation. Chem. Rev. Article ASAP. DOI: 10.1021/acs.chemrev.8b00608

42. Molecular Operating Environment (MOE), 2013.08 (2019). Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7.

43. Eswar, N., Eramian, D., Webb, B., Shen, M.-Y., and Sali, A. (2008). Protein structure modeling with MODELLER. Methods Mol. Biol. *426*:145-159.

44. Jo, S., Kim, T., and Im, W. (2007). Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. PloS One *2*(9).

45. Lee, J., Cheng, X., Swails, J.M., Yeom, M.S., Eastman, P.K., Lemkul, J.A., Wei, S., Buckner, J., Jeong, J.C., Qi, Y.F.*, et al.* (2016). CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. J. Chem. Theory Comput. *12*, 405-413.

46. Wu, E.L., Cheng, X., Jo, S., Rui, H., Song, K.C., Davila-Contreras, E.M., Qi, Y.F., Lee, J.M., Monje-Galvan, V., Venable, R.M.*, et al.* (2014). CHARMM-GUI Membrane Builder Toward Realistic Biological Membrane Simulations. J. Comput. Chem. *35*, 1997-2004.

47. Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L., and Schulten, K. (2005). Scalable molecular dynamics with NAMD. J. Comput. Chem. *26*, 1781-1802.

48. Best, R.B., Zhu, X., Shim, J., Lopes, P.E.M., Mittal, J., Feig, M., and MacKerell, A.D. (2012). Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone phi, psi and Side-Chain chi(1) and chi(2) Dihedral Angles. J. Chem. Theory Comput. *8*, 3257-3273.

49. Reid, J.K. (1971). Large Sparse Sets of Linear Equations. Academic Press, London and New York, ix + 283 pp.

50. Martyna, G.J., Tobias, D.J., and Klein, M.L. (1994). Constant-pressure molecular-dynamics algorithms. J. Chem. Phys. *101*, 4177-4189.

51. Darden, T., York, D., and Pedersen, L. (1993). PARTICLE MESH EWALD - AN N.LOG(N) METHOD FOR EWALD SUMS IN LARGE SYSTEMS. J. Chem. Phys. *98*, 10089-10092.

52. Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. J. Mol. Graph. Model. *14*, 33-38.

53. Bakan, A., Meireles, L.M., and Bahar, I. (2011). ProDy: Protein Dynamics Inferred from Theory and Experiments. Bioinformatics *27*, 1575-1577.

54. Glykos, N.M. (2006). Software news and updates - Carma: A molecular dynamics analysis program. J. Comput. Chem. *27*, 1765-1768.

FIGURE LEGENDS

Figure 1. The bacterial membrane protein insertase YidC2. The cartoon representation of membrane embedded YidC2 model based on its crystal structure (**PDB ID: 3WO7**)(Kumazaki et al., 2014a) is shown. The five TM helices (blue) of YidC2, the hydrophilic groove, the side chain of residue R72 (red), the C1 region including CH1 and CH2 helices connected with the C1 loop (yellow), the modeled C2 loop (green), and the carboxyl terminal helix (orange) are shown. The modeled cytoplasmic loop C2 (residues 195-216) was not resolved in the crystal structure.

Figure 2. Protein stability assessment for YidC2 with and without the C2 loop. (a) Root mean square deviation (RMSD) time series and (b) root mean square fluctuation (RMSF) estimations for YidC2 with (blue) and without (red) the C2 loop obtained from 2-μs simulations. RMSD analysis indicates that the absence of the C2 loop destabilizes the protein and RMSF analysis indicates that the presence of the modeled C2 loop (residues 195 to 216) significantly stabilize*s* the entire protein, particularly in the C1 loop region (residues 84-133).

Figure 3. Principal component analysis demonstrates that the system with the C2 loop is more stable than the system without the loop. (a) Projection of the simulation trajectory of YidC2 with (blue) and without the C2 loop onto their first two principal components (PC1, PC2). The YidC2 model with the loop is quite stable only locally fluctuating around its average structure; however, the YidC2 model without the loop jumps between multiple conformations indicating a significant conformational flexibility. (b) First principal component (PC1) eigenvector shown for the C1 region, that shows substantially different behavior in the presence

(blue) and absence (red) of the C2 loop. The width of the arrows represents the magnitude of fluctuation.

Figure 4: A salt bridge interaction network stabilizes the C1 loop in the presence of the C2 loop. (a) D205 (red) of C2 loop (green) can potentially form a salt bridge with K109 and/or K104 (blue) of the C1 region (yellow). (b) Time series of the D205-K109/104 donor-acceptor salt bridge distances.

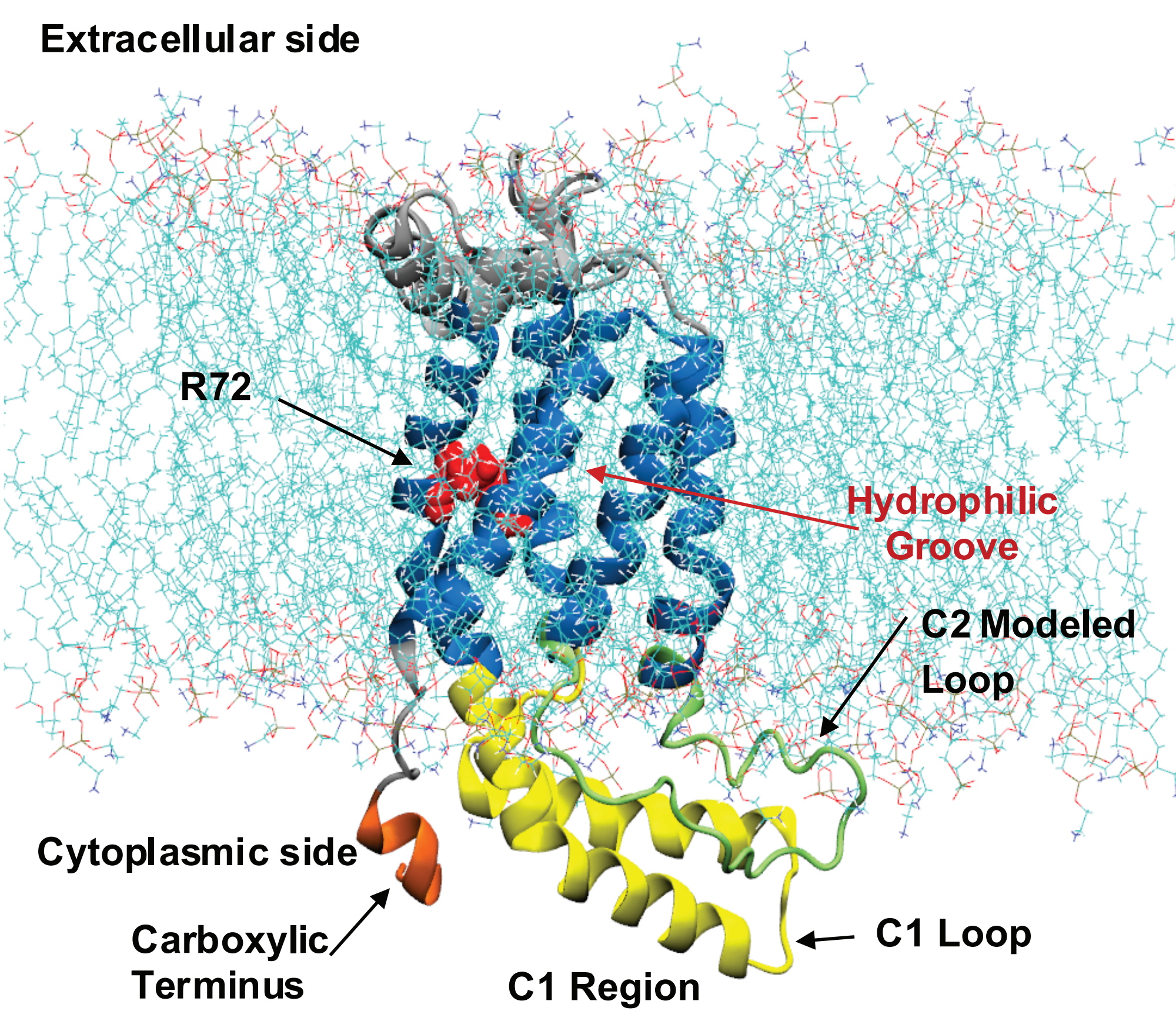
Figure 5. Carboxyl terminal secondary structure. (a) Visual representation of the carboxyl terminal domain of the systems with the loop (blue) and without the loop (red) upon equilibration. α-helical character of the residues 265 to 272 of the modified carboxyl terminal domain with (b) and without (c) the C2 loop is shown as a function of the simulation time. Yellow color indicates an alpha helical secondary structure. The carboxyl terminal helix unravels into a coil in the system without the loop, whereas it retains the secondary structure for the entire simulation in the system with the loop.

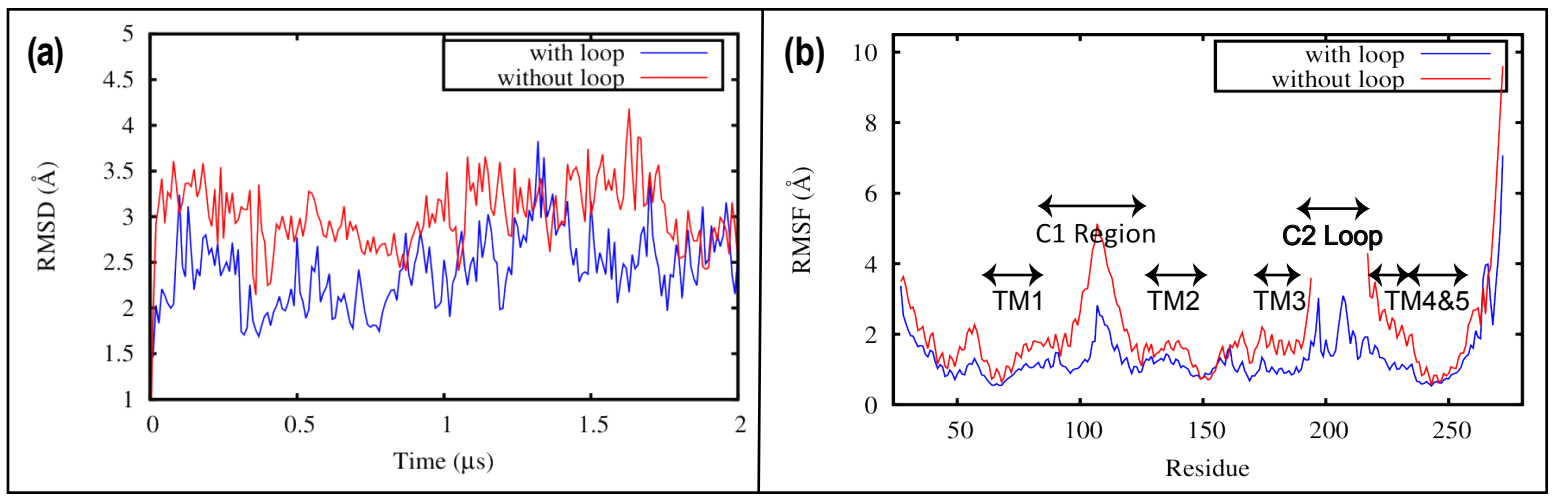
Figure 6. Dynamic network analysis of the systems with and without the C2 loop. (a) The heat map of the absolute value of the residue pair correlation difference between the systems with and without the C2 loop. The x and y axes represent the residue numbers for each residue pair. The blue color indicates no significant change in residue pair correlations due to the presence/absence of the C2 loop. The red color indicates a significant change in the residue pair correlations due to the presence/absence of the C2 loop. (b) The regions with the most significant

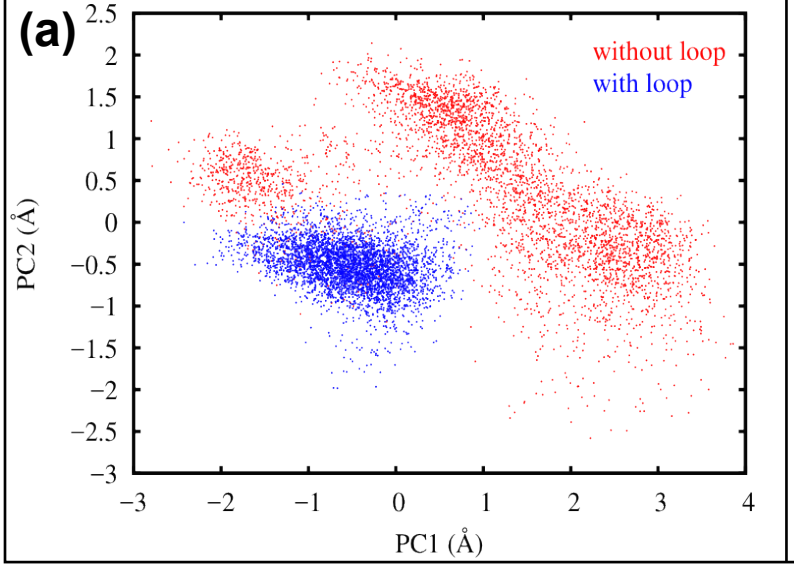
change in their correlation behavior including TM1, TM3, and TM4 (red) in the transmembrane domain and the C-terminal tail and the C1 region (blue) in the cytoplasmic side are highlighted.

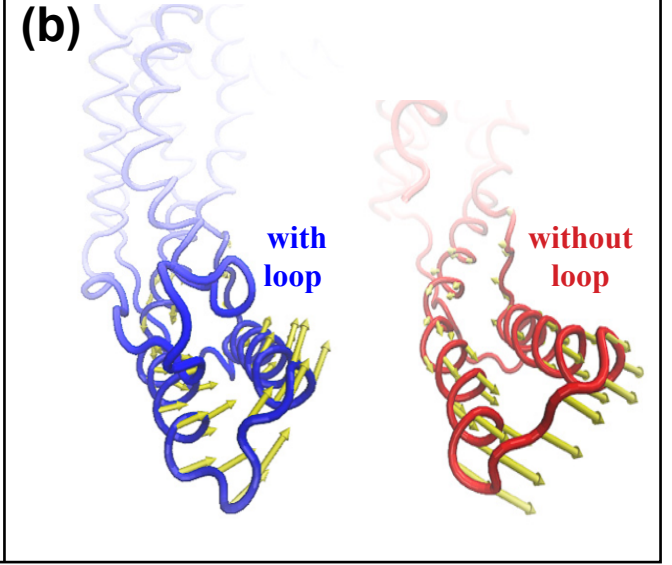
Figure 7: **An interdomain salt bridge forms between the C-terminal tail and the C1 region in the absence of the C2 loop.** (a) Donor-acceptor salt bridge distance between C-terminal residue E266 and C1 residue K81 as a function of simulation time for the systems without (red) and with (blue) the C2 loop. (b) A visual representation of the K81-E266 interaction in the system with (left) and without the C2 loop (right).

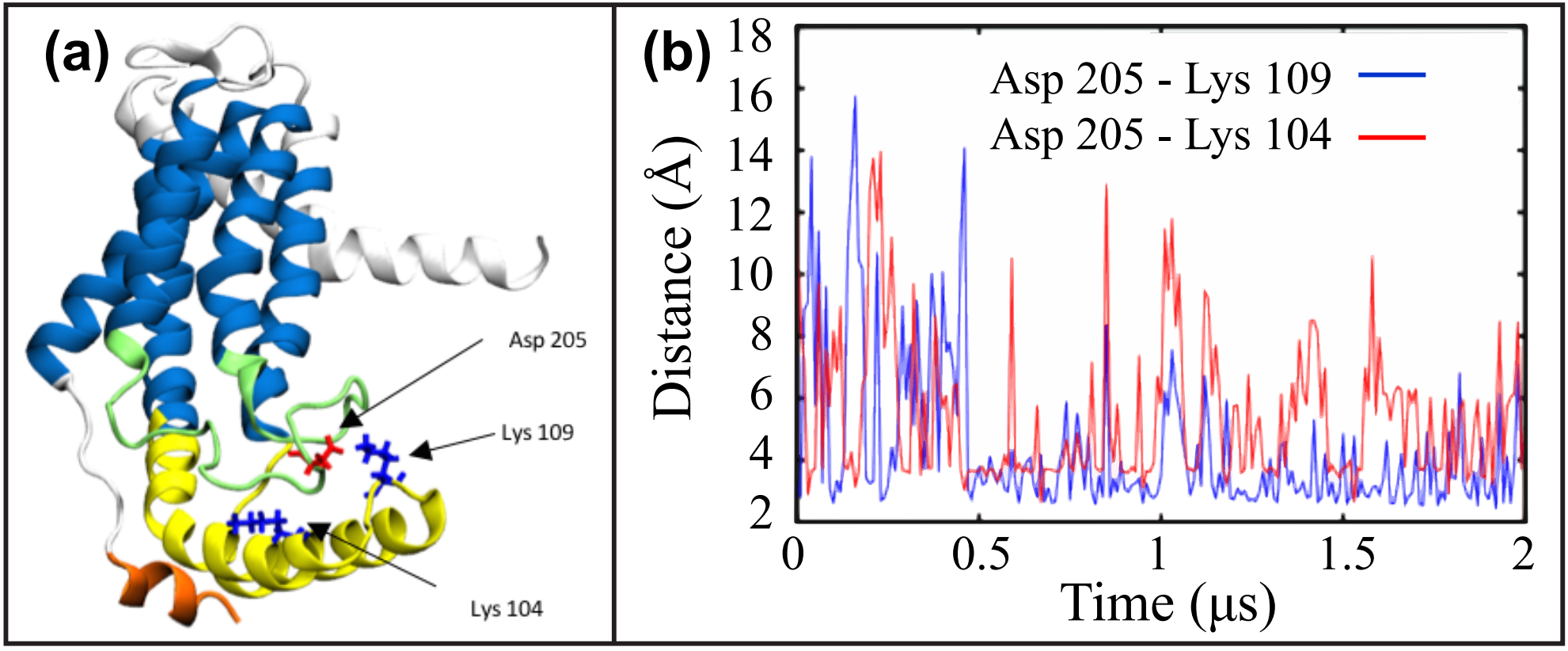
Figure 8. Influence of the C2 loop on membrane interactions. (a) The distribution of the tilt angle of YidC2 with respect to the membrane measured as the angle between the third principal axis (roll axis) of the protein and the membrane normal. In the system with the loop, the protein was tilted over 29° for 58% of the simulation. In the system without the loop, the protein was tilted over 29° for 19% of the simulation. (b) Interactions between D207 of the C2 loop and the POPE lipid head groups measured by counting the number of hydrogen bonds between the two as a function of simulation time. (c) Electrostatic (red) and van der Waals interaction energies (blue) of C2 loop and lipids.

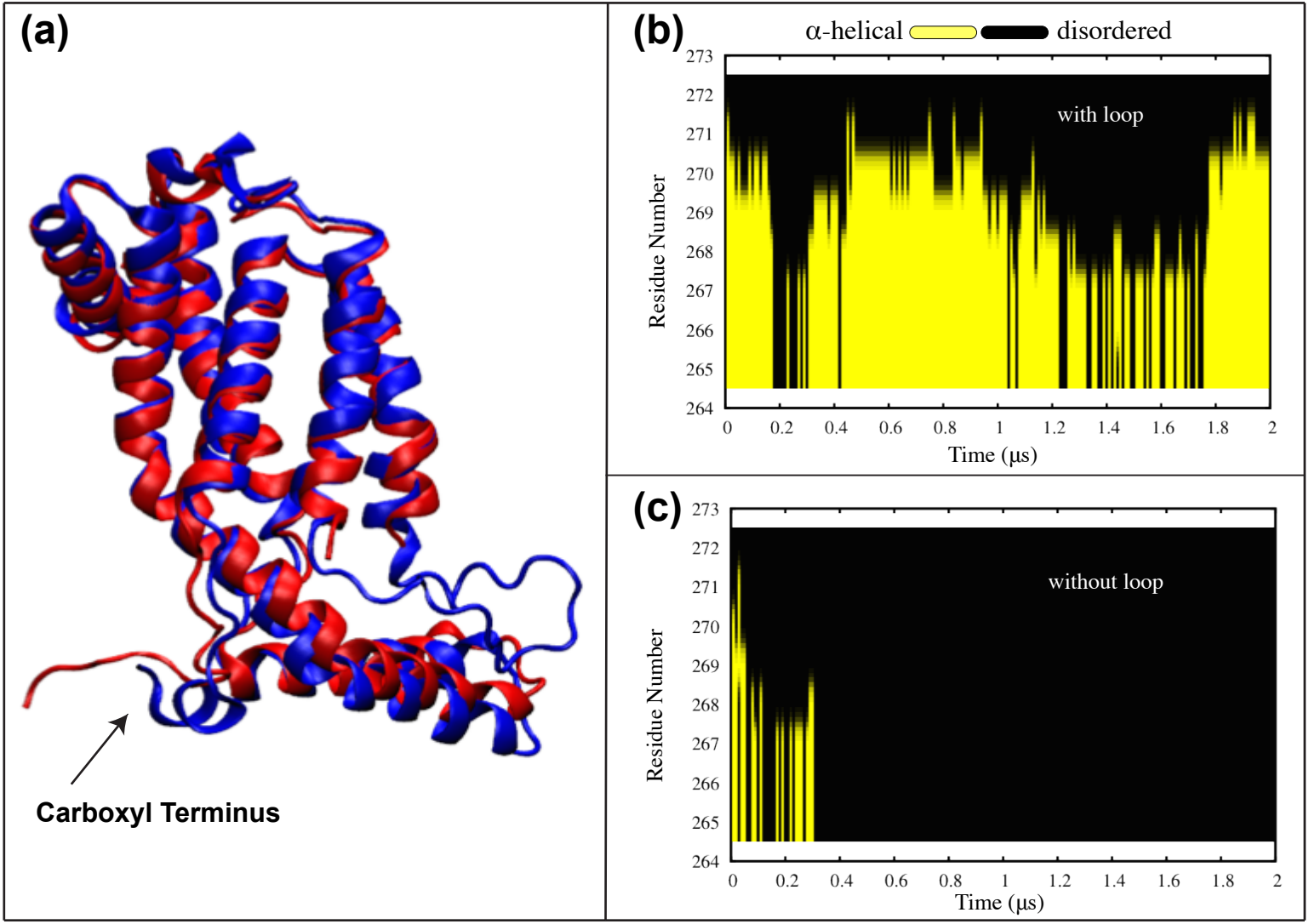

Extracellular side
R72
Hydrophilic Groove
C2 Modeled Loop
Cytoplasmic side
Carboxylic Terminus
C1 Region
C1 Loop

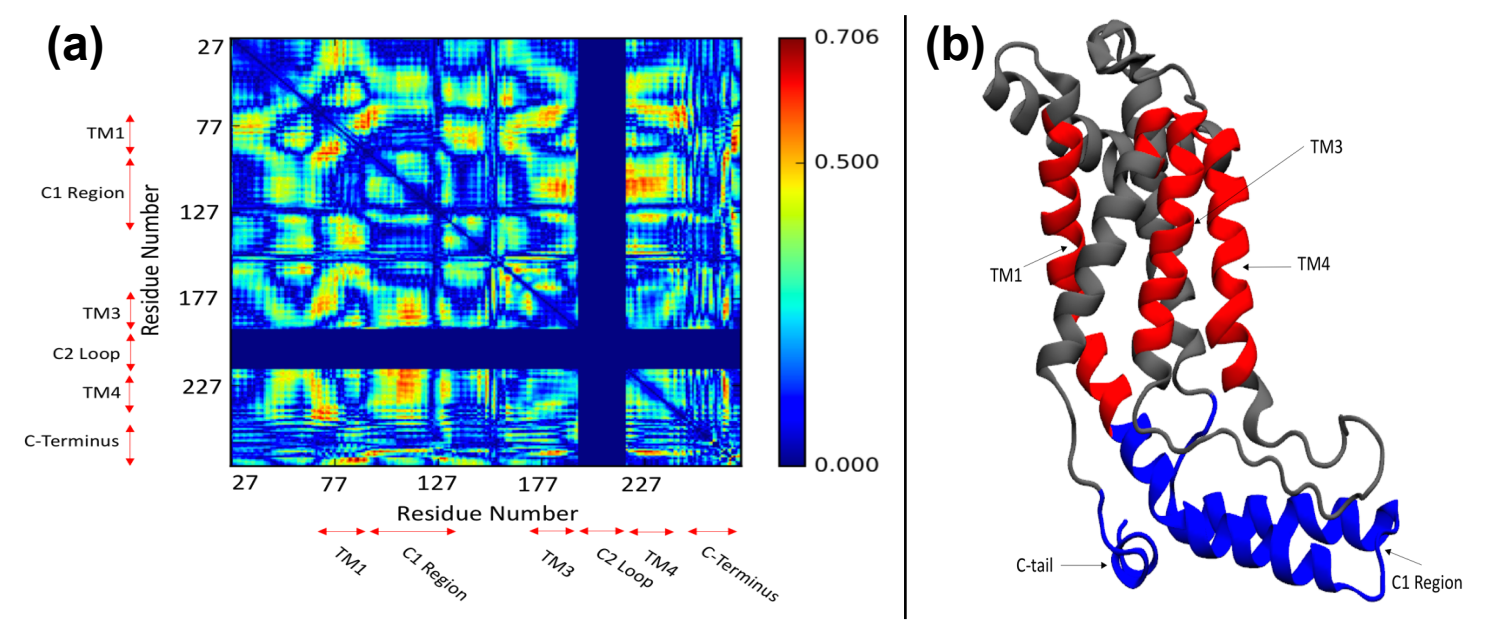

(a)
with loop
without loop
RMSD (Å)
Time (μs)
0
0.5
1
1.5
2
1
1.5
2
2.5
3
3.5
4
4.5
5
(b)
with loop
without loop
RMSF (Å)
Residue
C1 Region
C2 Loop
TM1
TM2
TM3
TM4&5
50
100
150
200
250
0
2
4
6
8
10

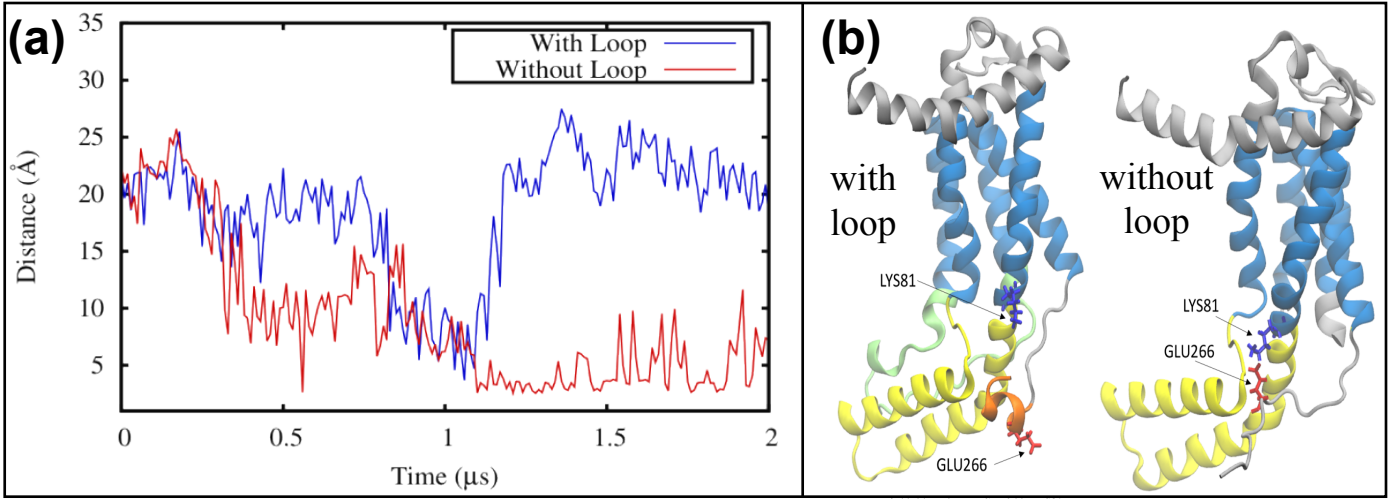

(a)
without loop
with loop
PC2 (Å)
PC1 (Å)
2.5
2
1.5
1
0.5
0
−0.5
−1
−1.5
−2
−2.5
−3
−3
−2
−1
0
1
2
3
4

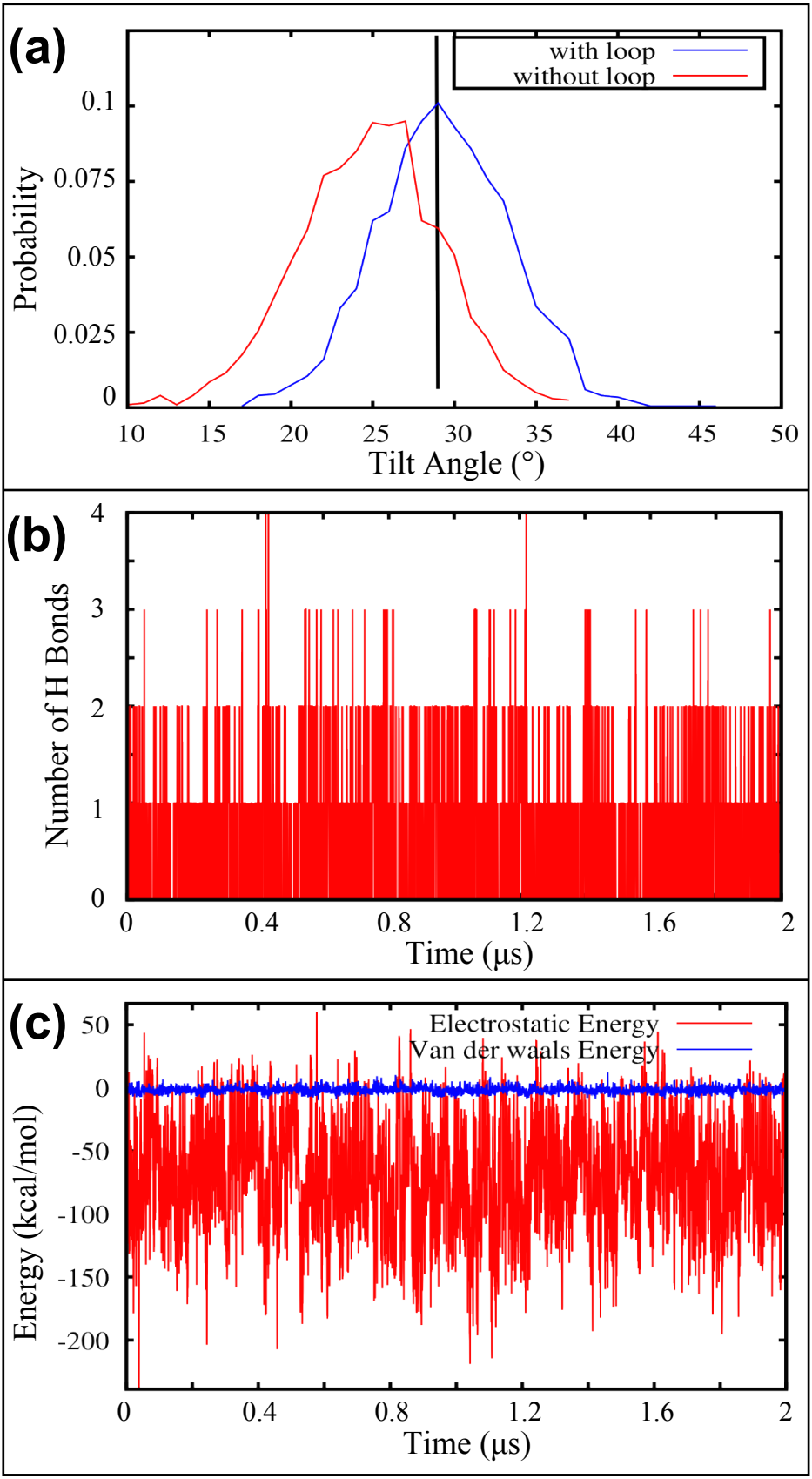

(b)
with loop
without loop


(a)
Asp 205
Lys 109
Lys 104
(b)
Distance (Å)
Time (μs)
Asp 205 - Lys 109
Asp 205 - Lys 104
18
16
14
12
10
8
6
4
2
0
0.5
1
1.5
2


(a)
Carboxyl Terminus
(b)
α-helical
disordered
with loop
Residue Number
273
272
271
270
269
268
267
266
265
264
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
Time (μs)
(c)
without loop
Residue Number
Time (μs)


(a)
27
77
127
177
227
Residue Number
TM1
C1 Region
TM3
C2 Loop
TM4
C-Terminus
0.706
0.500
0.000
(b)
TM3
TM1
TM4
C-tail
C1 Region


(a)
With Loop
Without Loop
Distance (Å)
Time (μs)
35
30
25
20
15
10
5
0
0.5
1
1.5
2
(b)
with loop
without loop
LYS81
GLU266
LYS81
GLU266


(a)
with loop
without loop
Probability
0.1
0.075
0.05
0.025
0
10
15
20
25
30
35
40
45
50
Tilt Angle (°)
(b)
Number of H Bonds
4
3
2
1
0
0
0.4
0.8
1.2
1.6
2
Time (μs)
(c)
Electrostatic Energy
Van der waals Energy
Energy (kcal/mol)
50
0
-50
-100
-150
-200
0
0.4
0.8
1.2
1.6
2
Time (μs)