Supporting Information for

Atomistic Characterization of Gramicidin Channel Formation

Delin Sun¹, Stewart He¹, W. F. Drew Bennett¹, Camille L. Bilodeau¹, Olaf S. Andersen², Felice C. Lightstone¹ and Helgi I. Ingólfsson¹

¹Biosciences and Biotechnology Division, Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA 94550.
²Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY 10065.
Simulation Methods

Replica-Exchange Umbrella Sampling Simulation of gA Dimerization. To run the REUS simulations, a pre-defined reaction coordinate or collective variable (CV) is needed. Intuitively, the movement of the two bilayer-embedded gA monomers can be described with two CVs, i.e., the gA-gA center-of-mass (COM) distance’s projections on the plane parallel to the bilayer surface (XY direction) and on the axis perpendicular to the bilayer surface (Z direction). However, the use of two separate CVs is computationally demanding in the REUS simulations, and therefore, we used a three-dimensional gA-gA COM distance (d_{gA-gA}) as the CV to run the REUS simulations.

We first used steered MD simulations to generate 52 configurations (windows), with increasing separation of the d_{gA-gA} between the two gA monomers (range: 1.35-3.9 nm), such that the distance between two neighboring windows is 0.05 nm. The COM of the gA monomer was defined using all C\alpha atoms. Then, during the REUS simulations, a harmonic biasing potential was imposed to each umbrella window to restrain the CV to fluctuate around a target CV. The functional form of the harmonic biasing potential is:

\[
V(\xi, i) = \frac{1}{2} k (\xi - \xi_i)^2, \tag{S1}
\]

where \(\xi\) is the target value of the CV, \(\xi_i\) is the instantaneous value of the CV, \(k\) is the umbrella spring constant and \(i\) is the index number of the umbrella window. The total potential energy of umbrella window \(i\) is thus determined as the sum of the force field based potential energy and the imposed harmonic biasing potential. Two neighboring umbrella windows \(i, j\) can attempt to exchange their configurations during the REUS simulations, with the replica exchange probability following the Metropolis criterion:

\[
P(i \leftrightarrow j) = \begin{cases} 
1, & \text{for } \Delta \leq 0, \\
\exp(-\Delta), & \text{for } \Delta > 0, 
\end{cases} \tag{S2}
\]

with

\[
\Delta = \beta(V(\xi, i) + V(\xi, j) - V^{ex}(\xi, i) - V^{ex}(\xi, j)), \tag{S3}
\]
where $\beta$ is the Boltzmann factor, $V^{ex}$ is the biasing potential energy calculated after the configurations of the two neighboring windows are swapped.

The REUS simulations were run in parallel on 52 NVIDIA Volta V100 GPUs with GROMACS (version 2019)\(^1\) patched with PLUMED2.5.\(^2\) For each umbrella sampling window, we ran 1 $\mu$s MD simulations. The exchange of neighboring windows was attempted every 100 ps. The umbrella spring constant was 5500 kJ/mol/nm\(^2\) and the 1d-WHAM program from A. Grossfield (http://membrane.urmc.rochester.edu/content/wham/)\(^3\) was used to derive the one-dimensional PMF profiles. It is noted that, with the $d_{gA-gA}$ as CV, each point in the derived PMF profile would contain a Jacobian factor\(^4\) whose value is in the range between $-2k_B T \ln(d_{gA-gA})$ and $-k_B T \ln(d_{gA-gA})$. The exact value for the Jacobian factor, however, is difficult to derive because of the anisotropic movement of gA in the directions parallel to and normal to the bilayer surface. Nevertheless, the Jacobian factors are expected to be equal at the same $d_{gA-gA}$ in different lipid bilayers, and therefore, the inclusion of Jacobian factor in the PMF profiles does not affect the gA dimerization free energy difference in different lipid bilayers. From the derived one-dimensional PMF profiles and the REUS simulation trajectories, it is possible to further derive two-dimensional PMF maps\(^5\) with the two reaction coordinates: (1) the $d_{gA-gA}$ and (2) the Z direction component of $d_{gA-gA}$ (or $d_Z$). First, we grouped all simulation snapshots in the REUS simulation trajectories based on the $d_{gA-gA}$ values; second, for each set of simulation snapshots with the $d_{gA-gA}$ value in the same histogram with bin width of 0.0135 nm, we constructed the histogram for $d_Z$ to estimate the conditional probability $P(d_Z | d_{gA-gA})$. In addition, the distribution probability for $d_{gA-gA}$, $P(d_{gA-gA})$ can be directly obtained from the WHAM derived one-dimensional PMF profile. Finally, $P(d_Z , d_{gA-gA})=P(d_Z | d_{gA-gA})*P(d_{gA-gA})$.

**Atomistic Unbiased MD Simulations of gA Dimerization.** We generated the initial configuration of 6HB gA channel embedded in the lipid bilayers using CHARMM-GUI.\(^6\) The starting configurations for two trans-bilayer gA monomers incorporated in DC\(_{18:1}\)PC, DC\(_{20:1}\)PC or DC\(_{22:1}\)PC bilayers were then
generated with steered MD simulations to force the two monomers in an equilibrated gA channel to slowly dissociate. Three additional starting configurations were generated by rotating one gA monomer by 180°, as illustrated in Fig. 1 in the main text. After short equilibrations, we ran 200 independent 1 µs long MD simulations for each of the 6 systems.

The CHARMM36 force field with the cross-term energy correction map (CMAP) was used to model the L- and D-amino acids in gA.7 The CHARMM compatible force field parameters for the formyl and ethanolamine groups were from previous work.8 Lipids were modeled using the CHARMM36 lipid force field.9 Each simulation system contained 200 lipids (100 lipids in each leaflet) and two gA monomers. The different systems were fully hydrated, containing 12,000-14,000 TIP3P water molecules and 0.15 M concentration of KCl. All simulations were done in the semi-isotropic ensemble at 310.15 K and 1 bar using the GPU version of GROMACS (version 2019).1 The Nosé-Hoover thermostat10,11 and the Parrinello-Rahman barostat12 were used to maintain the temperature and pressure. The LINCS algorithm13 was used to constrain water geometry and covalent bonds involving a hydrogen atom. Lennard-Jones interactions were switched off smoothly at 1-1.2 ns and the Particle Mesh Ewald method14 was used to treat long-range electrostatics with a real space cutoff distance of 1.2 nm. Long-range dispersion corrections to the energy and pressure were not applied. Snapshots of each simulation were saved every 100 ps.

Support Vector Machine Model Development. The 1.2 ms unbiased MD simulations generated a total of 12 million snapshots, which make it nontrivial to identify all gA monomer ↔ dimer transition events in the 1200 simulation trajectories. We developed an SVM machine learning model to assist the analysis of the simulation data. We first used the GROMACS trjconv utility to extract the atom coordinates for the two gA monomers in the 1200 simulations. We then used the k-means clustering approach implemented in PyEMMA2.5 package15 to assign the 12 million snapshots into 100 microstates using an intermolecular Cα-Cα atom contact feature. The Cα-Cα atom contact is defined to be 1 if the intermolecular Cα-Cα atom
distance is $\leq 1$ nm, otherwise 0. Only seven C$\alpha$ atoms of each gA monomer at the dimerization interface were considered. From the 100 groups of configurations, we were able to identify the 2HB, 4HB and 6HB dimers. Furthermore, we added the gA-gA hydrogen-bond donor-acceptor atom contact feature to train the SVM model. These include the donor-acceptor atom contact for the 1Val$^1$-2Ala$^3$, 1Ala$^3$-2Ala$^3$ and 1Ala$^3$-2Val$^1$ pairs the 1Val$^1$-2Ala$^3$ and 1Ala$^3$-2Val$^1$ pairs and the 1Val$^1$-2Val$^1$ pair. The donor-acceptor atom contact is defined to be 1 if the intermolecular donor-acceptor atom distance is $\leq 0.35$ nm, otherwise 0. We selected 10,000 snapshots for each of the five states (i.e., monomer, transition, 2HB, 4HB and 6HB states) with their corresponding features to train the SVM model.

The SVM model was implemented in the scikit-learn package. When tuning hyperparameters for the model, we used five-fold cross validation. We also explored a random forest model, using the same parameters. Both models performed well, and we chose the SVM model using a radial basis function (RBF) kernel. The classification problem proved to be a simple one, with the SVM model reaching 0.99 f1-score.

**Umbrella Sampling Simulations of K$^+$ Permeating the gA Channels.** As the convergence of sampling a single-atom K$^+$ permeating the cation-selective gA channel was expected to be much easier to achieve, we here used the plain umbrella sampling simulation method to derive the potential of mean force (PMF). The 6HB and 4HB dimer structures used for the umbrella sampling simulations were from the unbiased MD simulations. The CV was the center-of-mass distance between the ion and the 14 gA backbone C$\alpha$ atoms located at the dimer interface (7 C$\alpha$ atoms from each monomer). To obtain the PMF, two K$^+$ were selected for each system, one near the top of the channel and one near the bottom of the channel. Steered MD simulations were run to generate 31 umbrella sampling windows with the CV spanning from -1.5 to 0 nm in the negative direction and from 0 to 1.5 nm in the positive direction. The harmonic umbrella spring constant was 1000 kJ/mol/nm$^2$. The PMF was obtained with the 1d-WHAM program from Grossfield. To further improve sampling, the final PMF was symmetrized as the average.
of the PMF in the positive and negative directions such that \( \bar{w}(+r) = \bar{w}(-r) = \frac{w_{pos}(r) + w_{neg}(r)}{2} \), where \( w_{pos} \) and \( w_{neg} \) refer to the PMF arising from WHAM in the positive and negative directions respectively, \( \bar{w} \) is the averaged PMF, and the CV, \( r \) ranges from 0 to 1.5 nm. Both the 6HB and 4HB dimers umbrella sampling simulations were run in a DC_{18:1}PC bilayer and both dimers remained stable during the simulations. The PMF profiles are shown in Fig. 2 (c). The peak barriers, relative to bulk water, for K\(^+\) permeating the two channels are similar, 8.2 kcal/mol in the 4HB channel vs. 8.5 kcal/mol in the 6HB channel, suggesting that the 4HB dimer can conduct ions despite the disordereding of water molecules at the 4HB channel’s dimer interface.

**Rotational Root-Mean Squared Deviation Calculations.** To characterize the relative rotation movements of the two gA monomers, we calculated the rotational root-mean squared displacement (RRMSD) as follows: (1) For an unbiased MD simulation trajectory, we superimposed all C\(\alpha\) atoms of one gA monomer (gA1) to a pre-selected reference 6HB dimer structure; this produced a new simulation trajectory with superimposed gA1; (2) For the new simulation trajectory, we translated the second monomer gA2 to overlap with the gA2 in the pre-selected reference 6HB dimer; (3) we calculated the root-mean squared displacement (RMSD) for the gA2, and the derived RMSD value is the relatively rotational RMSD for gA1 and gA2.
**Figure S1.** Convergence analysis for the PMF of gA dimerization in (a) DC$_{18:1}$PC, (b) DC$_{20:1}$PC, and (c) DC$_{22:1}$PC bilayers. The total amount of REUS simulation time is 1000 ns for each umbrella window. The PMF profiles start to converge after 400 ns simulations and the convergence is indicated by the emergence of the V-shaped basins at gA-gA distance of ~1.5 nm.
Figure S2. Two-dimensional PMF maps for gA dimerization in (a) DC$_{18:1}$PC, (b) DC$_{20:1}$PC, and (c) DC$_{22:1}$PC bilayers. The two reaction coordinates are the three-dimensional COM distance between the gA monomers ($d_{gA-gA}$) and the Z-direction component of the COM distance ($d_z$).
Figure S3. Root-mean squared displacement (RMSD) for the simulated 6HB dimer (red dots) and 4HB dimer (blue dots) using the 1MAG and 1JNO PDB structures as reference structures. Only the backbone Cα atoms are used for the RMSD analysis. The 6HB structure is closer to the 1JNO experimental structure than to the 1MAG structure, as was also seen in§.
Figure S4, pages 10 - 34 below show each of the 1200 five-state simulation trajectories (a to f), combined panels with all simulations (g), and Fig. S5 shows the averages state information with time.

(a) **DC$_{18:1}$PC face-to-face 1-25**

![Graph showing simulation trajectories](image)

(a) **DC$_{18:1}$PC face-to-face 26-50**

![Graph showing simulation trajectories](image)
(a) DC$_{18:1}$PC face-to-face 51-75

(a) DC$_{18:1}$PC face-to-face 76-100
(a) DC_{18:1} PC face-to-face 101-125

Duration

Simulation time (ns)

Macrostate

(a) DC_{18:1} PC face-to-face 126-150

Duration

Simulation time (ns)
(b) \( \text{DC}_{18:1} \text{PC face-to-back 101-125} \)

Simulation time (ns)

---

(101) (102) (103) (104) (105)

(106) (107) (108) (109) (110)

(111) (112) (113) (114) (115)

(116) (117) (118) (119) (120)

(121) (122) (123) (124) (125)

---

(b) \( \text{DC}_{18:1} \text{PC face-to-back 125-150} \)

Simulation time (ns)

---

(126) (127) (128) (129) (130)

(131) (132) (133) (134) (135)

(136) (137) (138) (139) (140)

(141) (142) (143) (144) (145)

(146) (147) (148) (149) (150)

---
(c) DC\textsubscript{20:1}PC face-to-face 1-25

![Graph: DC\textsubscript{20:1}PC face-to-face 1-25](image)

(c) DC\textsubscript{20:1}PC face-to-face 26-50

![Graph: DC\textsubscript{20:1}PC face-to-face 26-50](image)
(c) DC_{20:1}PC face-to-face 51-75

(c) DC_{20:1}PC face-to-face 76-100
(c) DC\textsubscript{20:1}PC face-to-face 101-125

(c) DC\textsubscript{20:1}PC face-to-face 126-150
(c) DC$_{20:1}$PC face-to-face 151-175

(c) DC$_{20:1}$PC face-to-face 176-200
(d) DC_{20:1} PC face-to-back 1-25

(d) DC_{20:1} PC face-to-back 26-50
(d) DC$_{20:1}$PC face-to-back 51-75

(d) DC$_{20:1}$PC face-to-back 76-100
(d) DC$_{20:1}$ PC face-to-back 101-125

(d) DC$_{20:1}$ PC face-to-back 126-150
(d) DC$_{20:1}$ PC face-to-back 151-175

(d) DC$_{20:1}$ PC face-to-back 176-200
(e) DC$_{22:1}$PC face-to-face 1-25

(e) DC$_{22:1}$PC face-to-face 26-50
(e) DC$_{22:1}$PC face-to-face $51-75$

(e) DC$_{22:1}$PC face-to-face $76-100$
(e) DC_{22;1}PC face-to-face 101-125

(e) DC_{22;1}PC face-to-face 126-150
(f) DC$_{22:1}$PC face-to-back 1-25

| Macrostate |  |  |  |  |
|------------|---|---|---|---|
| (1)        | (2) | (3) | (4) | (5) |
| (6)        | (7) | (8) | (9) | (10) |
| (11)       | (12) | (13) | (14) | (15) |
| (16)       | (17) | (18) | (19) | (20) |
| (21)       | (22) | (23) | (24) | (25) |

Simulation time (ns)

0 500 0 500 0 500 0 500 1000

(f) DC$_{22:1}$PC face-to-back 26-50

| Macrostate |  |  |  |  |
|------------|---|---|---|---|
| (26)       | (27) | (28) | (29) | (30) |
| (31)       | (32) | (33) | (34) | (35) |
| (36)       | (37) | (38) | (39) | (40) |
| (41)       | (42) | (43) | (44) | (45) |
| (46)       | (47) | (48) | (49) | (50) |

Simulation time (ns)

0 500 0 500 0 500 0 500 1000
(f) DC22:1 PC face-to-back 151-175

(f) DC22:1 PC face-to-back 176-200
Figure S4. The 1200 five-state simulation trajectories after classification with the SVM model. There are 200 trajectories for each of the three bilayer thicknesses DC$_{18:1}$PC (a, b), DC$_{20:1}$PC (c, d), and DC$_{22:1}$PC (e, f) starting in either a face-to-face (FF) or face-to-back (FB) orientation (see Fig. 1d in main text). The five different states: monomer (M), monomer → initial dimer transition (T), 2HB, 4HB and 6HB are labeled 1, 2, 3, 4, and 5, respectively. (g) As an overview all 200 simulations for each condition are shown overlapped where each simulation has a random color and small random shift on the state (y) axes.
Figure S5. Fraction of snapshots for different states (monomer, transition, 2HB, 4HB and 6HB) as a function of simulation time. (a) DC\textsubscript{18:1}PC face-to-face; (b) DC\textsubscript{18:1}PC face-to-back; (c) DC\textsubscript{20:1}PC face-to-face; (d) DC\textsubscript{20:1}PC face-to-back; (e) DC\textsubscript{22:1}PC face-to-face; (f) DC\textsubscript{22:1}PC face-to-back.
Figure S6. Time evolution of the tryptophan residues’ $\chi_1$ (chi1, red) and $\chi_2$ (chi2, blue) dihedral angles during the (a) $M \rightarrow T \rightarrow 6HB$ and (b) $M \rightarrow T \rightarrow 4HB \rightarrow 6HB$ dimerization pathways. M: monomer state (state 1); T: transition state (state 2); 4HB: gA channel with four hydrogen bonds (state 4); 6HB: gA channel with six hydrogen bonds (state 5).
(a)

Monomer state (M)

Transition state (T)

Two hydrogen bond state (2HB)

Four hydrogen bond state (4HB)

Six hydrogen bond state (6HB)

DC_{18:1} PC
(b)

Monomer state (M)

Transition state (T)

Two hydrogen bond state (2HB)

Four hydrogen bond state (4HB)

Six hydrogen bond state (6HB)

DC_{20:1} PC
Figure S7. $\chi_1$ (chi1) and $\chi_2$ (chi2) dihedral angle density maps for the four tryptophans (amino acids 9, 11, 13, and 15) in the gA subunits. The 1200 µs unbiased MD simulations in the DC$_{18:1}$PC (a), DC$_{20:1}$PC (b) and DC$_{22:1}$PC (c) bilayers were used for the analysis and separated based on state.
Figure S8. Convergence analysis for the PMF for K⁺ permeating across the 4HB (a) and 6HB (b) gA channels.
References

(1) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. SoftwareX 2015, 1-2, 19.

(2) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for An Old Bird. Computer. Phys. Commun. 2014, 185, 604.

(3) Grossfield, A. An Implementation of WHAM: the Weighted Histogram Analysis Method Version 2.0.9.

(4) Boresch, S.; Karplus, M. The Jacobian Factor in Free Energy Simulations. J. Chem. Phys. 1996, 105, 5145.

(5) Meshkin, H.; Zhu, F. Q. Thermodynamics of Protein Folding Studied by Umbrella Sampling along a Reaction Coordinate of Native Contacts. J. Chem. Theory Comput. 2017, 13, 2086.

(6) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. J. Comput. Chem. 2008, 29, 1859.

(7) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; Mackerell, A. D. J. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ(1) and χ(2) Dihedral Angles. J. Chem. Theory Comput. 2012, 8, 3257.

(8) Allen, T. W.; Andersen, O. S.; Roux, B. Structure of Gramicidin A in a Lipid Bilayer Environment Determined Using Molecular Dynamics Simulations and Solid-State NMR Data. J. Am. Chem. Soc. 2003, 125, 9868.

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

(10) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. Phys. Rev. A. 1985, 31, 1695.

(11) Nosé, S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. Mol. Phys. 1984, 52, 255.

(12) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. J. Appl. Phys. 1981, 52, 7182.

(13) Hess, B.; Bekker, H.; Berendsen, H. J. C. Lincs: A Linear Constraint Solver for Molecular Simulations. J. Comput. Chem. 1997, 18, 1463.
(14) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577.

(15) Scherer, M.; Trendelkamp-Schroer, B.; Paul, F.; Perez-Hernandez, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J. H.; Noe, F. PyEMMA 2: A software package for estimation, validation, and analysis of markov models. *J. Chem. Theory Comput.* **2015**, *11*, 5525.

(16) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825.