SUPPLEMENTARY INFORMATION:

Computational Design of Polypeptide-Based Compartments for Synthetic Cells

Jianming Mao,[†] Yongkang Xi,[‡] Armin Shayesteh Zadeh,[¶] Allen P. Liu,[‡,§,‖,⊥] and Andrew L. Ferguson[*,¶,†]

[†]*Department of Chemistry, University of Chicago, Chicago, IL, 60637*

[‡]*Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI 48109*

[¶]*Pritzker School of Molecular Engineering, University of Chicago, Chicago, IL, 60637*

[§]*Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI 48109*

[‖]*Cellular and Molecular Biology Program, University of Michigan, Ann Arbor, MI 48109*

[⊥]*Department of Biophysics, University of Michigan, Ann Arbor, MI 48109*

E-mail: andrewferguson@uchicago.edu

Contents

S1 Free Energy Calculation by Alchemical Transfer

The free energy ΔG associated with the extraction of one ELP monomer from the ELP membrane to the water region is employed as our proxy measurement of the thermodynamic stability of a ELP membrane, and by extension, the entire bilayer ELP vesicle. We estimate ΔG using the recently proposed alchemical transfer approach[1–4] in which a molecule is decoupled in the one phase and a "ghost" molecule placed in the other phase is simultaneously coupled. This circumvents the need for co-alchemical ions to maintain charge neutrality, as required in the double decoupling method (DDM)[5–7], and enables the calculation of free energy calculation in a single set of simulations to reduce computational time and resources.

S1.1 Thermodynamic Pathway

The total free energy ΔG for the extraction process of an ELP out of a bilayer membrane is given by the expression,

$$\Delta G = -k_B T \ln \frac{Z_{sol}}{Z_{mem}} = -k_B T \ln \left(\frac{\int_{\Omega_{\text{sol}}} e^{-\beta U(\mathbf{r})}\, d\mathbf{r}}{\int_{\Omega_{\text{mem}}} e^{-\beta U(\mathbf{r})}\, d\mathbf{r}} \right), \tag{S1}$$

where Z is the configuration partition function and Ω_{sol}, Ω_{mem} represent the configuration space domains of the interacting ELP in the bulk water and membrane states, respectively. The free energy ΔG can be computed by a series of transformations following a thermodynamic path connecting the initial state and the final state as illustrated in Figure S1. In the initial state of the simulation, an unrestrained ELP within a relaxed bilayer (ELP2) is chosen for alchemical free energy calculation. A duplicate of the ELP is introduced into the aqueous phase (ELP1) in a fully decoupled state with harmonic restraints applied to both termini to maintain its position and orientation in the solvent. As the system progresses toward the final state, the ELP in the water phase becomes fully coupled to the environment,

allowing it to interact freely with the solvent and ions. Simultaneously, the ELP within the membrane is decoupled from the system, and harmonic restraints are applied to its ends to maintain its position and orientation during the free energy calculation process. The free energy difference for the complete process – removal of an ELP from the membrane into the solvent – along this pathway can be written as,

$$\Delta G = \Delta G_{res}^{\#1} + \Delta G_{transfer} + \Delta G_{res}^{\#2} + \Delta G_{correction}, \tag{S2}$$

where each term corresponds to the free energy associated with the realization of each step in Figure S1. In the following sections, we describe the details and implementation for obtaining each free energy component.

S1.2 Free Energy Perturbation: Harmonic Springs (1) ($\Delta G_{res}^{\#1}$)

The first stage in our alchemical free energy calculation involves the application of harmonic restraints on both termini of ELP2 embedded in the membrane. These restraints are introduced to preserve the positional and orientational alignment of the ELP2 within the membrane during subsequent steps when the ELP2 is partially or fully decoupled from its environment. Such restraint protocols are commonly employed in protein-ligand binding free energy calculations to prevent the bound ligand from drifting out of the binding site during decoupling. The free energy change associated with introducing these restraints is given by,

$$\Delta G_{res}^{\#1} = -k_B T \ln \frac{Z_{mem}^{1^=,2^=}}{Z_{mem}^{1^=}} = -k_B T \ln \left(\frac{\int_{\Omega_{\text{mem}}} e^{-\beta\left[U(\mathbf{r})+u_{\bar{1}}^{=}+u_{\bar{2}}^{=}\right]} d\mathbf{r}}{\int_{\Omega_{\text{mem}}} e^{-\beta\left[U(\mathbf{r})+u_{\bar{1}}^{=}\right]} d\mathbf{r}}, \right), \tag{S3}$$

where the "=" superscript indicates that each terminus is restrained, and $Z_{\text{mem}}^{1^=}$ and $Z_{\text{mem}}^{1^=,2^=}$ represent the configurational partition functions before and after the restraints are applied to ELP2. We employ a *mem* subscript on these partition functions to remind us that

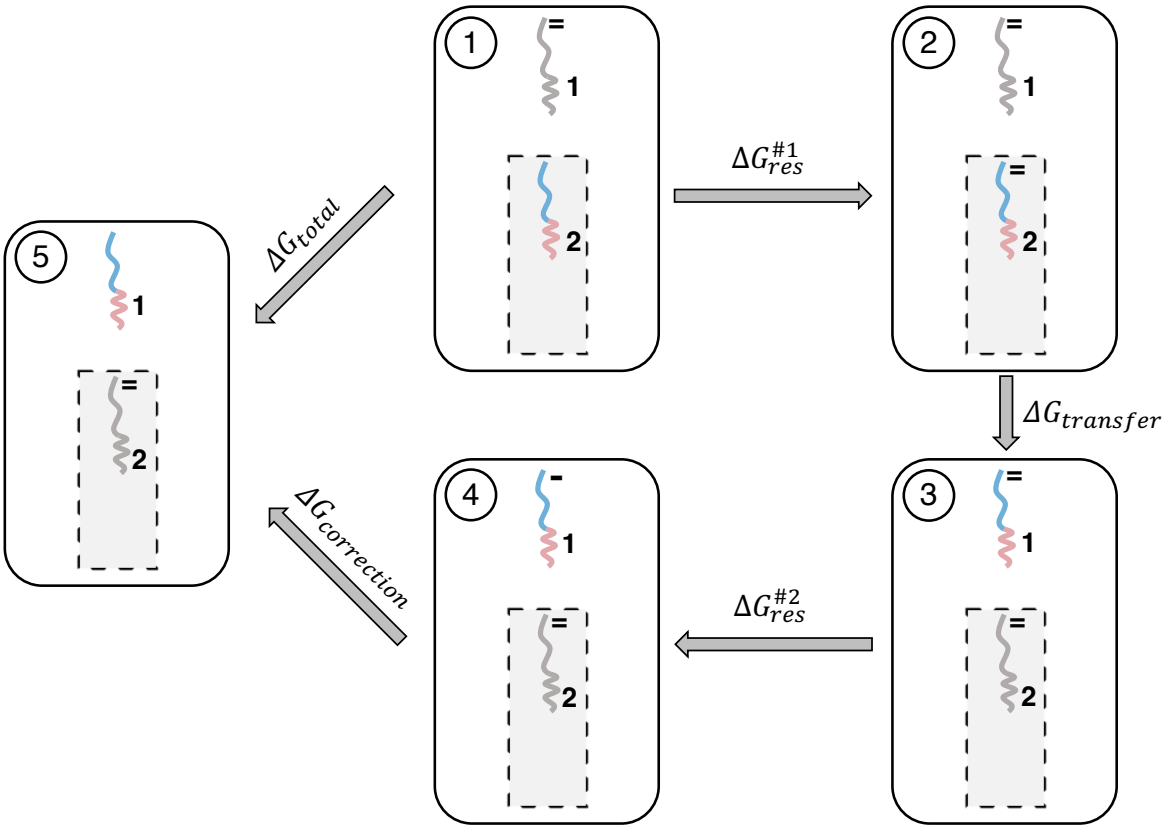

Figure S1: An illustration depicting the alchemical transfer method for computing the ΔG of the extraction of an ELP from membrane to water region, where the "–" symbol denotes that one end of the ELP is restrained by a harmonic potential while the "=" symbol denotes that both ends are restrained by harmonic potentials. In the initial state, the selected monomer (ELP2) from the membrane is unrestrained and interacting with the environment. A "ghost" version of this ELP is then duplicated and inserted to the water region with both ends restrained by harmonic potentials. In the first step, the harmonic restraints are gradually imposed on the C and N-termini of ELP2 to obtain $\Delta G_{res}^{\#1}$. Secondly, the ELP2 in membrane is gradually decoupled and the "ghost" ELP1 in bulk water is coupled simultaneously to acquire $\Delta G_{transfer}$. Finally, harmonic potential on the lower tail of ELP1 in water is gradually released to obtain the $\Delta G_{res}^{\#2}$ to account the ELP in water to relax its internal configurational degrees of freedom while prohibiting translational motion of the molecule through the single harmonic restraint. The final extraction free energy ΔG is acquired by summing all the contributions and applying an entropic correction to the prevailing concentration by performing an analytical correction for the translational restraint applied to the terminal ELP in water.

it is the membrane ELP that is coupled to the environment. The additional potential energy terms from the harmonic restraint on both ends of ELP1 and ELP2 take the form of $u_1^{=} = \frac{1}{2}K[(z_1^N - \xi_1^N)^2 + (z_1^C - \xi_1^C)^2]$ and $u_2^{=} = \frac{1}{2}K[(z_2^N - \xi_2^N)^2 + (z_2^C - \xi_2^C)^2]$. In practice, harmonic restraints were applied to both termini of ELP2 with a final force constant of

5000 kJ.mol^{-1}.nm^{-2}, maintaining the Z-component of the center-of-mass (COM) distance between each terminus and the membrane. The restraints were gradually introduced across 12 λ windows with the values λ = [0.0, 0.0001, 0.0002, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.5, 1.0], where this choice of windows was optimized by trial and improvement.

Physically, $\Delta G_{res}^{\#1}$ corresponds to the transition from State 1 to State 2 in Figure S1, wherein commencing from an initial state with a fully coupled ELP in the membrane and a fully decoupled ELP in the solvent with both termini restrained, we turn on the two termini restraints for the ELP in the membrane. This is necessary since we will be decoupling the membrane ELP in the next step and these constraints are required to stop it from collapsing or drifting within the box to improve the efficiency of free energy estimates along the thermodynamic cycle.

S1.3 Free Energy Perturbation: Alchemical Transfer ($\Delta G_{transfer}$)

In the next stage, we perform an alchemical transformation where the interactions of ELP2 within the membrane are turned off while simultaneously turning on interactions of ELP1 in the aqueous phase. This effectively models the reversible transfer of an ELP molecule with both termini harmonically restrained from the membrane to the solvent. The associated free energy change is given by,

$$\Delta G_{transfer} = -k_B T \ln \frac{Z_{sol}^{1^=,2^=}}{Z_{mem}^{1^=,2^=}} = -k_B T \ln \left(\frac{\int_{\Omega_{\text{sol}}} e^{-\beta\left[U(\mathbf{r})+u_1^=+u_2^=\right]}\, d\mathbf{r}}{\int_{\Omega_{\text{mem}}} e^{-\beta\left[U(\mathbf{r})+u_1^=+u_2^=\right]}\, d\mathbf{r}} \right), \tag{S4}$$

where the partition functions $Z_{mem}^{1^=,2^=}$ and $Z_{sol}^{1^=,2^=}$ correspond to fully restrained coupled ELP configurations in the membrane and solvent, respectively. The change in subscript from *mem* in the initial state to *sol* in the final state reminds us that the coupled ELP switches from the membrane state to the solvated state. This transfer is illustrated in the transition from State 2 to State 3 in Figure S1.

The transformation was carried out over 41 equally spaced λ windows with an interval of 0.025, using a dual-topology scheme where Lennard-Jones (LJ) and electrostatic interactions were progressively switched on/off for ELP1 and ELP2. During the transfer process, the soft-core potentials[8] for both Lennard Jones (LJ) and electrostatic interactions in the following forms were employed to avoid the well-known endpoint singularity effects in the vicinity of $\lambda = 0$ and 1,

$$U_{vdW}^{sc}(\lambda) = \begin{cases} \lambda^n 4\epsilon \left[\frac{1}{\left(\alpha_{LJ}(1-\lambda)^2 + \left(\frac{r}{\sigma}\right)^6\right)^2} - \frac{1}{\alpha_{LJ}(1-\lambda)^2 + \left(\frac{r}{\sigma}\right)^6} \right] & r < r_c \\ 0 & r \geq r_c \end{cases}, \tag{S5}$$

$$U_{Q}^{sc}(\lambda) = \lambda^n \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r \left[\alpha_Q(1-\lambda)^2 + r_{ij}^2\right]^{1/2}}, \tag{S6}$$

where ϵ and σ are the LJ energy and size parameters for the atom pair in question, r_c is the cutoff for the LJ interaction, q_i and q_j are the charges on the two atoms i and j interacting by the Coulomb force, r_{ij} is the separation between the atoms, $\epsilon_0 = 8.854 \times 10^{-12}$ F.m^{-1} is the permittivity of free space, and $\epsilon_r = 15$ is the relative permittivity as recommended for the non-polarizable Martini 2.2 water model[9,10]. The switching parameters $\alpha_{LJ} = \alpha_Q = 0.5$, $\sigma = 0.3$, and $n = 1$ were set based on prior work[8]. We note that it is convenient to employ the soft-core potential in the alchemical transfer within our simulations, but if large numbers of atoms and charges were to be perturbed, such as transferring a large molecule in an all-atom simulation, we refer the readers to the original alchemical transfer work[2] where a more complex perturbation energy is used.

S1.4 Free Energy Perturbation: Harmonic Springs (2) ($\Delta G_{res}^{\#2}$)

Following the alchemical transfer, we remove the harmonic restraint applied to the C-terminus of the coupled ELP1 to allow the molecule to undergo conformational relaxation in the aqueous phase since the equilibrium conformation in water will, in general, differ from

that in the membrane. The corresponding free energy change is,

$$\Delta G_{res}^{\#2} = -k_B T \ln \frac{Z_{sol}^{1^-,2^=}}{Z_{sol}^{1^=,2^=}} = -k_B T \ln \left(\frac{\int_{\Omega_{\text{sol}}} e^{-\beta\left[U(\mathbf{r})+u_1^- +u_2^=\right]} \, d\mathbf{r}}{\int_{\Omega_{\text{sol}}} e^{-\beta\left[U(\mathbf{r})+u_1^= +u_2^=\right]} \, d\mathbf{r}} \right), \tag{S7}$$

where the "–" superscript indicates that only the N-terminus of the ELP remains restrained, and $u_1^- = \frac{1}{2}K[(z_1^N - \xi_1^N)^2)^2]$. Similar to Section S1.2, the force constant was decreased from 5000 to 0 kJ.mol^{-1}.nm^{-2} across 12 λ windows, using a sequence of λ values of λ = [0.0, 0.0001, 0.0002, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.5, 1.0]. The process corresponding to $\Delta G_{res}^{\#2}$ is illustrated in the transition from State 3 to State 4 in Figure S1.

S1.5 Free Energy Perturbation: Entropic Correction ($\Delta G_{correction}$)

The single harmonic restraint applied to the N-terminus of the solvent ELP in State 4 restricts translational diffusion of the molecule to prevent it from potentially interacting with the membrane and improving the efficiency of the free energy calculations. Calculation of the final ΔG, however, requires that we apply an analytical correction associated with the entropy gain associated with release of this harmonic restraint. This entropic correction is given by,

$$\Delta G_{correction} = -k_B T \ln \frac{(2\pi k_B T/K)^{1/2}}{V_0}, \tag{S8}$$

where K = 5000 kJ.mol^{-1}.nm^{-2} is the force constant of the harmonic potential and V_0 = 1660 Å is the molecular volume accessible to a peptide monomer at a standard concentration of 1 mol.L^{-1} [11,12]. This transfer is illustrated in the transition from State 4 to State 5 in Figure S1.

S1.6 Overall Free Energy Calculation

Combining all components along the thermodynamic cycle, the total free energy of extracting one ELP monomer from the membrane into the aqueous phase is

$$\begin{aligned}\Delta G &= -k_B T \ln \left[\frac{Z_{\text{mem}}^{1^=,2^=}}{Z_{\text{mem}}^{1^=}} \times \frac{Z_{\text{sol}}^{1^=,2^=}}{Z_{\text{mem}}^{1^=,2^=}} \times \frac{Z_{\text{sol}}^{1^-,2^=}}{Z_{\text{sol}}^{1^=,2^=}} \times \frac{(2\pi k_B T/K)^{1/2}}{V_0} \right] \\ &= \Delta G_{\text{res}}^{\#1} + \Delta G_{\text{transfer}} + \Delta G_{\text{res}}^{\#2} + \Delta G_{\text{correction}}. \end{aligned} \tag{S9}$$

The first three terms are computed from alchemical simulations, and the final term is an analytical correction, contributing a small, but significant, value of approximately 5.3 $k_B T$ per ELP. ΔG values were normalized by the number of residues in each sequence to facilitate comparison across ELP sequences of different lengths.

S1.7 Alchemical Simulation Parameters

For all the alchemical simulations, a stochastic dynamics (sd) integrator with an inverse friction constant of 1.0 ps^{-1} was used to sample the decoupled states. Each simulation window was energy minimized using the steepest descent algorithm and then equilibrated for 40 ns followed by a production run of 100 ns at a temperature of 300 K and a pressure of 1 bar. The temperature control was handled by the Langevin thermostat [13] and the pressure control was controlled employing the same scheme in the membrane equilibration. The free energy for each step was integrated using the multistate Bennett acceptance ratio (MBAR) algorithm [14] in the alchemlyb [15] package and the uncertainty was estimated by five-fold block averaging. Compared with TI [16] and Bennett acceptance ratio (BAR) scheme [17], MBAR gives more rapidly converging estimates of $\Delta G_{res}^{\#2}$ (Figure S2) and similar performance for $\Delta G_{res}^{\#1}$ and $\Delta G_{transfer}$.

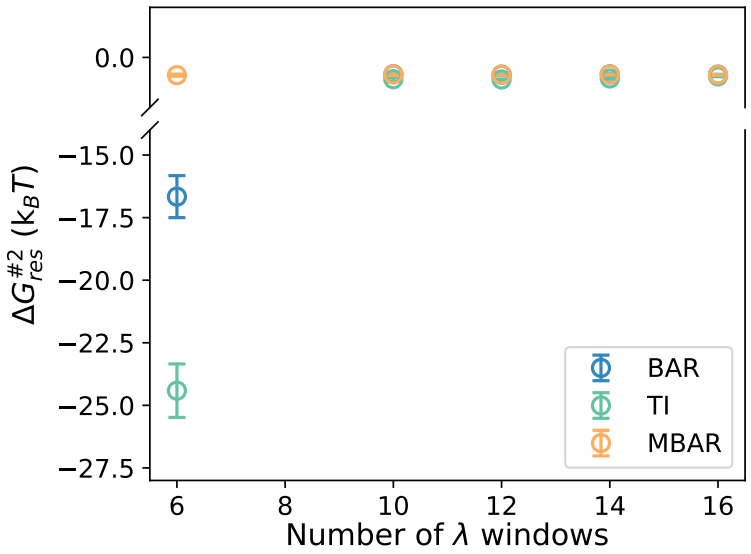

Figure S2: The convergence profiles of free energy $\Delta G^{\#2}_{res}$ estimated by TI, BAR and MBAR with respect to the number of λ windows. The MBAR method consistently yields more rapid convergence with number of λ windows. TI and BAR are able to produce similar estimates when using significantly larger numbers of λ windows.

In Figure S3, we show a cumulative free energy profile of a representative ELP T_5I_5 at the various stages illustrated in Figure S1. The extraction free energy for this ELP, measured on a per amino acid residue basis, was calculated to be $\Delta G = (1.33 \pm 0.11)\ k_BT.\text{residue}^{-1}$ where uncertainties were estimated by five-fold block averaging. We note that $\Delta G_{transfer}$ dominates the total free energy followed by $\Delta G^{\#2}_{res}$, and that $\Delta G^{\#1}_{res}$ contributes less than 3.4% to the overall free energy. The entropic correction $\Delta G_{correction}$ is estimated to be a value of 5.3 k_BT per sequence leading to a small constant offset of 0.11 k_BT or 0.12 k_BT per residue for ELPs with a total pentamer repeat of 10 or 9, respectively. We also show the overlap matrices between λ windows and convergence profiles of ΔG's by a bi-directional analysis, indicating that the total free energy is converged within the sampling schemes (Figure S4).

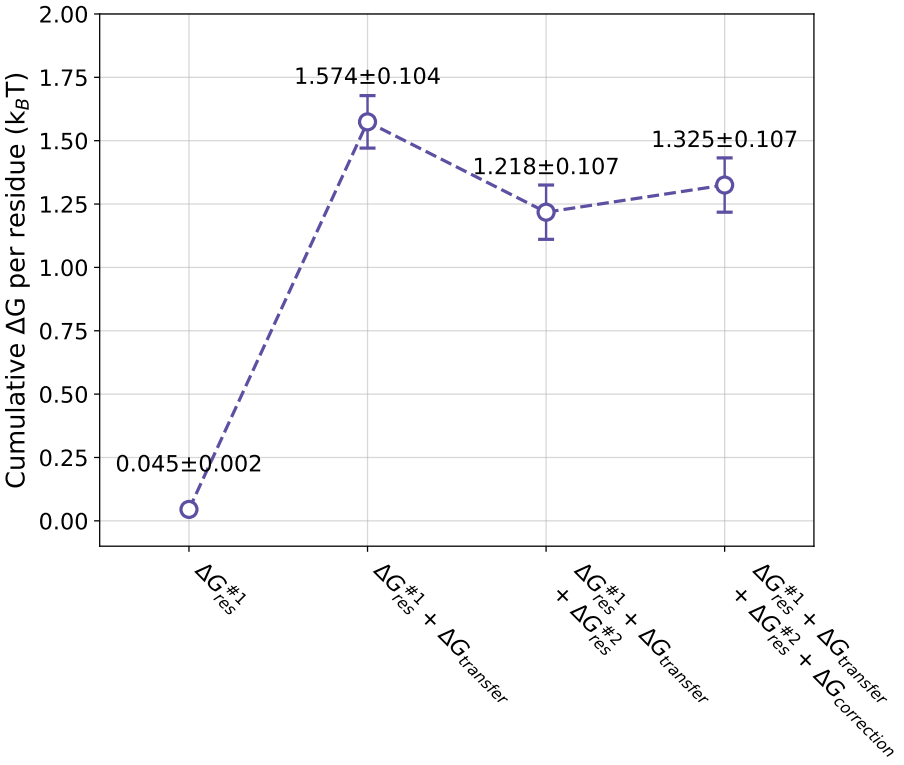

Figure S3: An illustrative cumulative free energy profile for an ELP $(VPGTG)_5(VPGIG)_5$. The uncertainties are estimated by five-fold block averaging.

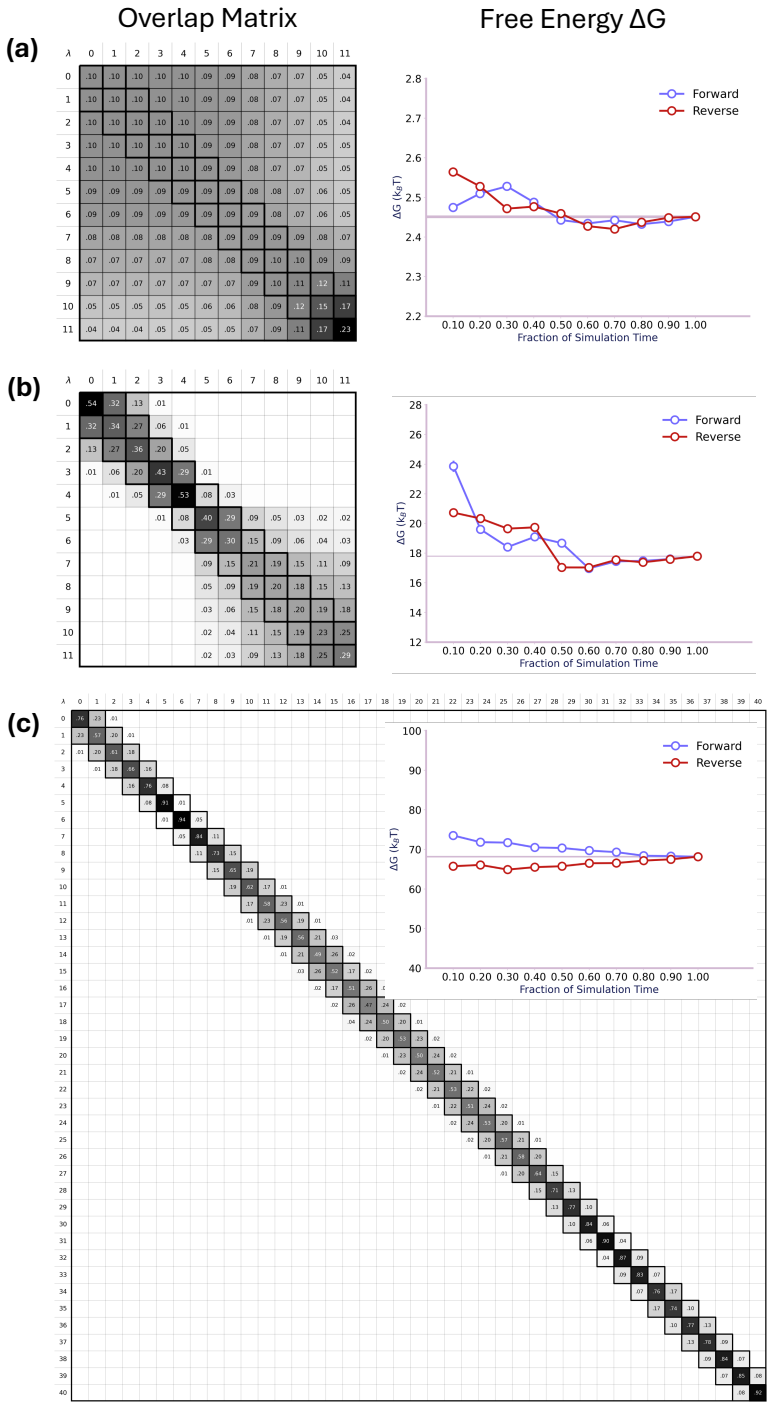

Figure S4: The convergence profiles of each component of ΔG. The overlap matrices between each λ windows and the bidirectional analysis of free energies are computed using the alchemlyb for (a) $\Delta G_{res}^{\#1}$, (b) $\Delta G_{res}^{\#2}$, and (c) $\Delta G_{transfer}$. All samples show overall reasonable overlap between adjacent states and converged free energies.

S2 Validation of the Alchemical Transfer Method

We validated the alchemical transfer method in five benchmarks where we compare against previously reported values calculated by umbrella sampling or the double decoupling method (DDM), or against known theoretical values.

In the first three validations, we employed our alchemical transfer pipeline to compute the hydration free energy calculation for three molecules – 1,2,3-trichloro-5-(2,5-dichlorophenyl)benzene (TCDP), decane, and F-uracil – by transferring the molecule from vacuum to a water phase (Figure S5a). The initial simulation systems were obtained from the FreeSolv database[18], where the small molecules were parameterized by acpype employing the General Amber Force Field (GAFF) and AM1-BCC charges[19,20]. The simulation box for alchemical transfer was expanded in the z-dimension to create a vacuum slab where a duplicated small molecule was inserted in a decoupled state. The subsequent simulation setup adopted the same protocol in Section S1. We note that application of the pipeline to relatively rigid small molecules means that we do not need to apply the dual harmonic restraining potentials to each molecule that were necessary to control and relax the chain-like ELP molecule. In this case, we apply a single harmonic restraint to conduct the transfer, maintain the restraints on the terminal states, and do not perform the entropic correction. The LJ and electrostatic interactions were simultaneously perturbed in 21 equally spaced λ windows which effectively transferred the molecule from the bulk water to the vacuum. Each sampling window was energy minimized with the steepest descent algorithm, equilibrated for 100 ps, and sampled for 2 ns in the NVT ensemble at a temperature of 298.15 K maintained by a v-rescale thermostat[21]. In all simulations, the equations of motion were integrated by a stochastic dynamics integrator[13] with a time step of 2 fs. Position restraints were applied to the COM of the small molecules in all three dimensions adopting a harmonic potential with a force constant of 1000 kJ.mol^{-1}.nm^{-2}. The hydration free energy was computed by the `gmx bar` scheme and the uncertainty was estimated by five-fold block averaging. We compated our calcu-

lated ΔG values against those reported in the FreeSolv database computed using DDM[18] (Figure S5c).

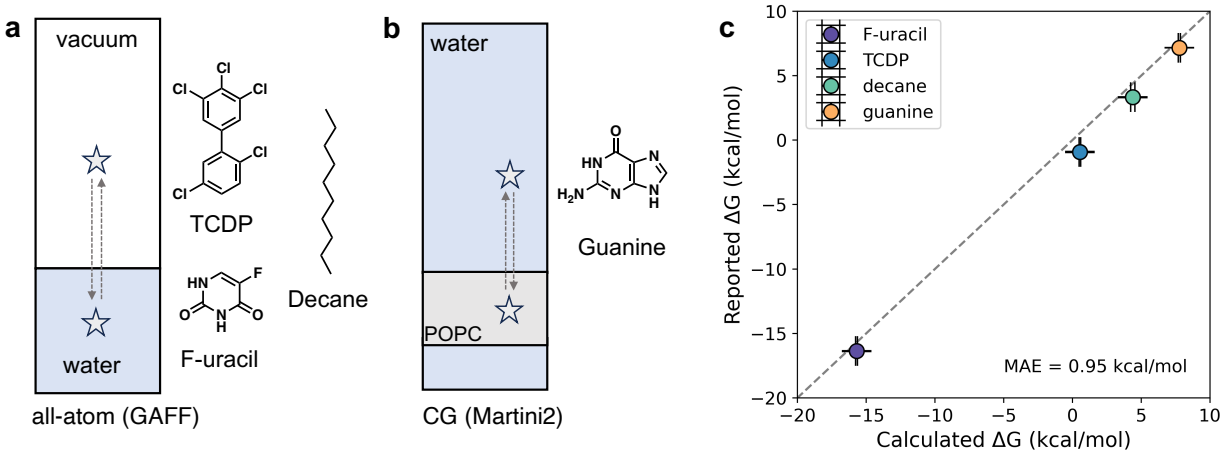

Figure S5: Validation of the alchemical transfer method for free energy calculations in four benchmark systems for which previously reported transfer free energies are available. The simulation box for computing the (a) hydration free energy of three small molecules – TCDP, decane, and F-uracil – and (b) the partition free energy of a guanine from water to the interior of a POPC membrane. The hydration free energy is obtained by transferring the molecule from vacuum to bulk water, and the partition free energy by transferring the guanine from water region to the membrane interior. (c) A parity plot showing the comparison between the ΔG computed using our alchemical transfer approach (x-axis) and against previously reported values from the FreeSolv database computed using DDM[18] or umbrella sampling[22] (y-axis). Images of chemical structures were created using ChemDraw from Revvity Signals Software, Inc. (`https://revvitysignals.com/`).

In the fourth validation, we calculated the partition free energy of a guanine molecule from water to a POPC membrane (Figure S5b). In this case, we modeled the molecule using the Martini 2.2 force field[10,23] as a test of the method for a coarse-grained system using the same force field employed for the ELP systems in this work. In the first step, a system consisting of a guanine molecule inside a POPC bilayer membrane (64 molecules each leaflet) was built by packmol[24] and solvated by the non-polarizable water beads. The system was equilibrated for 50 ns at a temperature of 303 K and a pressure of 1 bar using a v-rescale thermostat[21] and Parrinello-Rahman barostat[25]. The guanine molecule was then copied and inserted into the bulk solvent in a decoupled state. We performed the alchemical transfer simulations in 41 equally spaced λ windows. Each window was sampled for 50 ns to converge the free energy. The additional simulation parameters were identical to those specified in Section

S1. Harmonic restraints with a force constant of 400 kJ.mol^{-1}.nm^{-2} were applied to the COM of both guanine molecules along the membrane normal. The partition free energy for guanine was computed by the `gmx bar` scheme and the uncertainty was estimated by five-fold block averaging. The calculated partition free energy was compared against a reference value computed by umbrella sampling reported by Uusitalo *et al.*[22] (Figure S5c).

In Figure S5c, a parity plot of the free energy values computed by our method (x-axis) against previously reported values (y-axis) demonstrates excellent agreement for all four benchmarks, with a mean average error of 0.95 kcal.mol^{-1}. We do observe that the alchemical transfer method appears to systematically slightly overestimate the free energy, which might be attributed to the somewhat shorter sampling time in our calculation compared with those from the reported protocols. Nevertheless, the extremely good concordance of our method with previously reported results within better than 1 kcal.mol^{-1} provides strong support for the alchemical transfer approach to computing the free energy difference between two states.

In our fifth validation, we sought to confirm the applicability of the approach directly to our ELP systems. In this case we constructed a benchmark in which we decoupled one monomer from within a $(VPGHG)_2(VPGAG)_2$ ELP bilayer membrane while simultaneously coupling another ELP monomer elsewhere in the membrane (Figure S6). By construction, the free energy change for this process should be zero, assuming a uniform membrane and full equilibration in each step of the process. We employed the simulation protocol detailed in Section S1, with the exception that both ELPs resided within the ELP membrane and we did not apply the entropic correction. We employed harmonic potentials with a force constants of 5000 kJ.mol^{-1}.nm^{-2} and used 41 equally spaced λ windows for the transfer process. We obtained a transfer free energy $\Delta G = (0.024 \pm 0.054)$ kcal.mol^{-1}.residue^{-1}, which spans the expected result of zero within error bars. This result provides additional support to our free energy calculation method to evaluate the transfer free energy of ELP

monomers in ELP bilayer membranes.

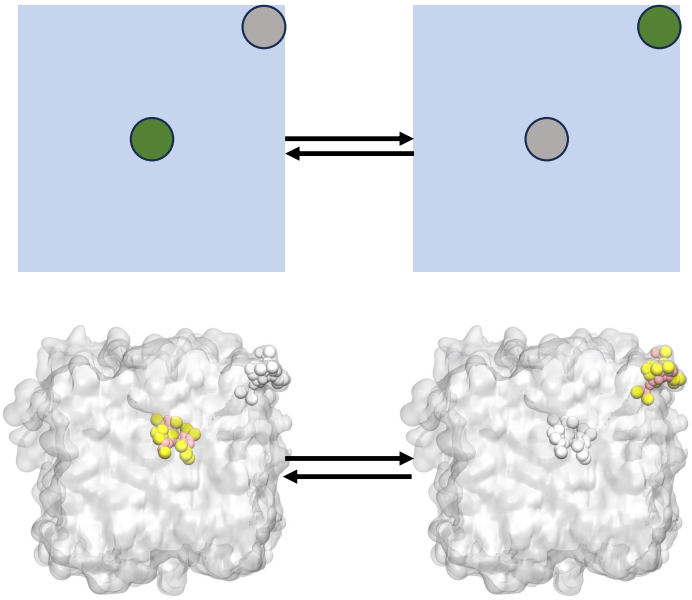

Figure S6: Schematic illustration showing the symmetric decoupling of one ELP monomer from within an ELP membrane and simultaneously coupling a second ELP monomer elsewhere in the membrane. The green circle (upper schematic) and yellow colored ELP molecule (lower visualization) indicates the coupled state, while the gray circle and white ELP molecule indicates the decoupled state. Molecular visualizations were constructed using VMD[26].

S3 Supplementary Figures

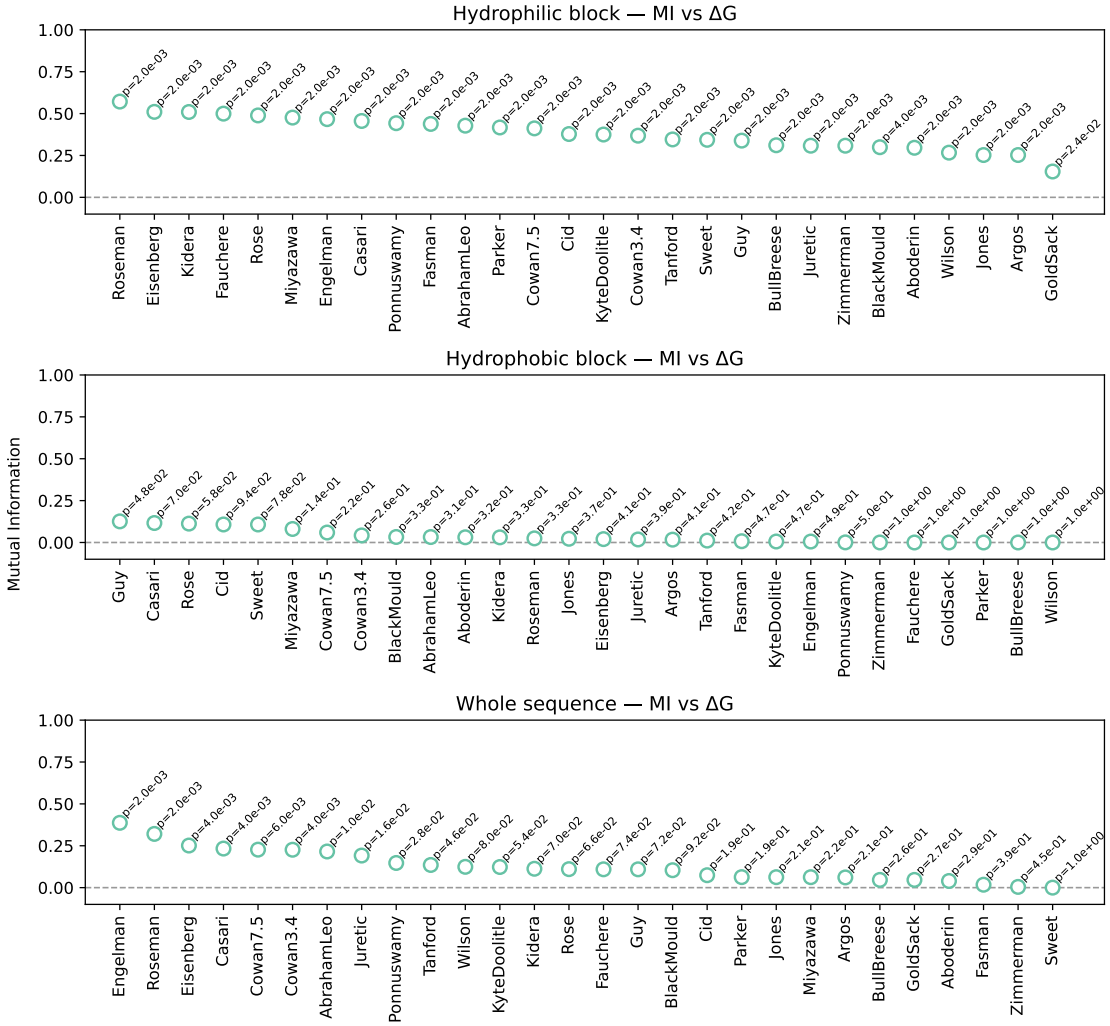

Figure S7: Mutual information of the calculated ΔG and GRAVY[27] values calculated by different hydropathy scales for the hydrophilic block, hydrophobic block and the whole sequence. In total, 28 hydropathy scales were used. The maximum observed values were MI = 0.57 nats ($p = 0.002$) for the Roseman scale for the hydrophilic block and MI = 0.39 nats ($p = 0.002$) for the Engleman scale for the whole sequence. Calculations were performed using scikit-learn[28] and p-values for each MI value computed under random permutations are printed next to each point on the plot. The MI values were computed using the natural log convention so although formally dimensionless can be considered to be reported in nats.

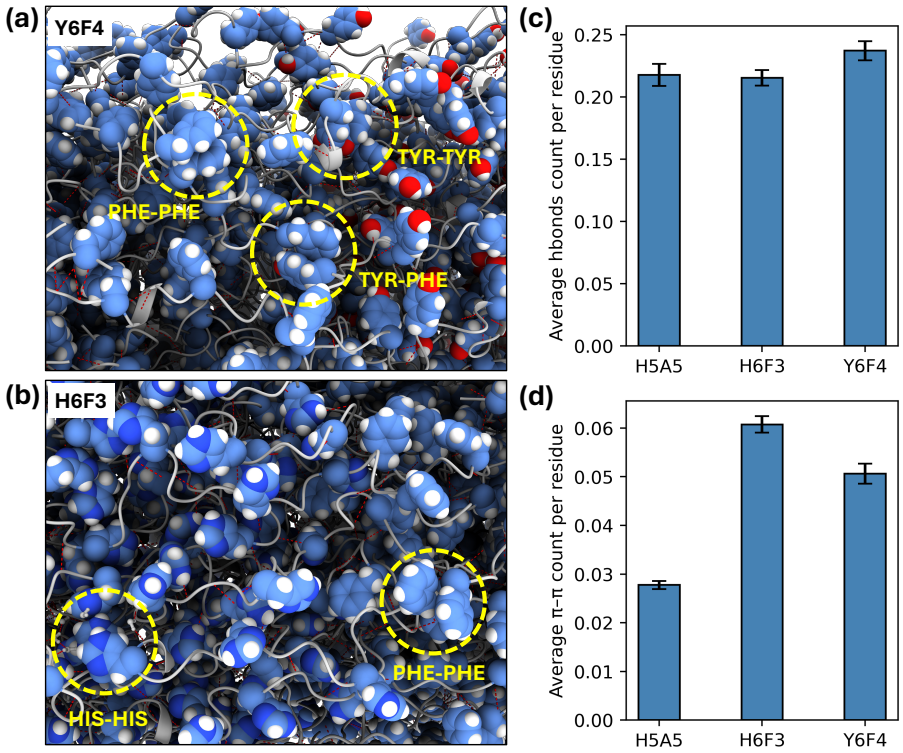

Figure S8: The equilibrated coarse-grained Martini simulations of three selected top candidates – H_5A_5, H_6F_3, Y_6F_4 – were backmapped to CHARMM36 all-atom resolution[29] using the Martini backmap tool[23]. The all-atom structures were energy minimized and equilibrated for 50 ps in the NVT ensemble and 10 ns in the NPT ensemble. Representative snapshots of the all-atom backmapped and equilibrated (a) Y_6F_4 and (b) H_6F_3 ELP membranes. The aromatic residues including Phe, Tyr, and His are represented spheres and the peptide chains are shown in cartoon representation. The π-π stacking among Phe, Tyr, and His residues are highlighted in yellow circles, and the hydrogen bonds between peptides are illustrated by red dashed lines. We harvested 11 frames equally spaced over the 10 ns NPT trajectory and collated counts of (c) the average per residue number of hydrogen bonds and (d) π-π interactions using MDAnalysis[30]. A hydrogen bond is defined as when a hydrogen bond donor and acceptor atom are within a distance of 3.0 Å and the donor–hydrogen–acceptor angle is greater than 150°, and a π-π interaction as when two aromatic rings have centroid–centroid distance less than 6.5 Å and ring-plane orientations consistent with face-to-face or edge-to-face stacking. The molecular renderings were created by ChimeraX[31].

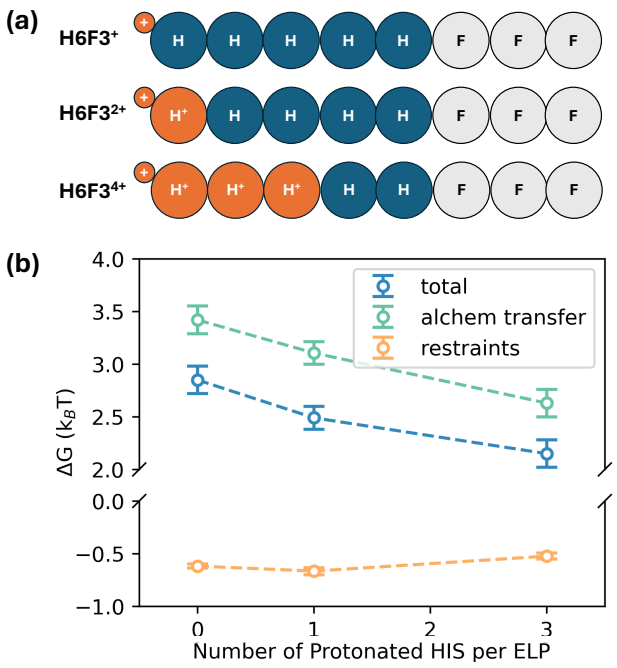

Figure S9: Relationship of ΔG with the protonation states in a membrane consisting of H_6F_3 ELPs. (a) Schematic illustration of the protonation states of the histidine residues in the ELP models with an increasing number of protonated histidines mimicking the decreasing of pH in the chemical environment. (b) The total per-residue ΔG and its components from the alchemical transfer process and the turning on/off restraints for ELPs with different charge states.

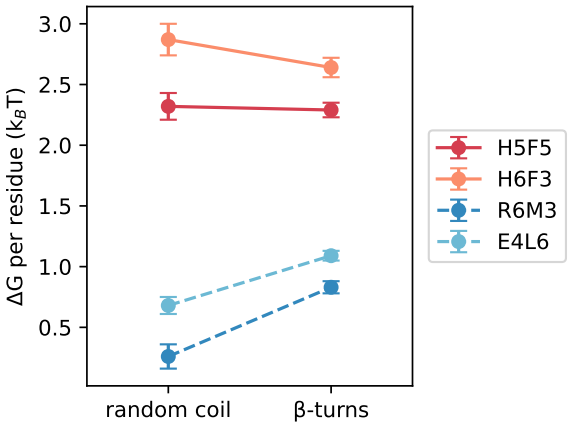

Figure S10: Assessment of the impact of secondary structure assignment to ELP sequences upon membrane stability. We compare the free energy to extract a single ELP from the membrane under a secondary structure assignment of the entire ELP as a random coil (left) and an assignment of a β-turn secondary structure to the hydrophobic block and a random coil to the hydrophilic block (right). The red and orange points correspond to the top performing ELP candidates H_5F_5 and H_6F_3, while the blue and turquoise points correspond to the poor ELP candidates R_6M_3 and E_4L_6. The poor candidates show some limited additional stabilization under the imposition of the β-turn in the hydrophobic block, whereas the good candidates show no change in stability within error bars. The stability rank ordering among the four sequences is maintained under both secondary structure assignments.

S4 Supplementary Tables

Table S1: Summary of free energy estimates for ELP candidates evaluated during active learning. Each entry corresponds to a unique elastin-like polypeptide (ELP) sequence evaluated via molecular dynamics simulations. The table includes the total free energy change (ΔG) for the full ELP, the error in that estimate, and the normalized ΔG per residue along with its corresponding error. The active learning (AL) round in which each sequence was evaluated is also included. Energies are reported in units of $k_{\mathrm{B}}T$. These data are available as a machine readable csv file at `https://github.com/Ferg-Lab/ELP_Simulation` and also accessible via a persistent doi at `https://doi.org/10.5281/zenodo.15778533`.

ELP Sequence	Residue Number	ΔG per ELP ($\mathrm{k_B}T$)	ΔG per residue ($\mathrm{k_B}T$)	AL Round
H6F3	45	129.20 ± 6.04	2.87 ± 0.13	10
H6A3	45	112.87 ± 3.23	2.51 ± 0.07	10
H6A4	50	121.69 ± 5.17	2.43 ± 0.10	10
H5A5	50	118.82 ± 2.79	2.38 ± 0.06	10
H5I4	45	106.48 ± 4.91	2.37 ± 0.11	23
H6F4	50	116.00 ± 5.19	2.32 ± 0.10	10
H5F5	50	116.08 ± 5.35	2.32 ± 0.11	10
T6A4	50	115.36 ± 3.06	2.31 ± 0.06	10
Y6F4	50	110.34 ± 1.68	2.21 ± 0.03	14
H6V4	50	106.78 ± 2.39	2.14 ± 0.05	22
H6I3	45	94.95 ± 3.87	2.11 ± 0.09	0
Y6I4	50	104.19 ± 4.80	2.08 ± 0.10	16
H6C4	50	104.06 ± 3.03	2.08 ± 0.04	1
H6V3	45	91.45 ± 3.51	2.03 ± 0.08	22
Y6I3	45	91.17 ± 3.48	2.03 ± 0.08	21
H3F6	45	90.81 ± 2.43	2.02 ± 0.05	0
H5A4	45	89.57 ± 3.75	1.99 ± 0.08	10
H6M4	50	93.69 ± 3.84	1.97 ± 0.08	12
H4C6	50	96.88 ± 3.30	1.94 ± 0.07	10
H4A6	50	96.55 ± 4.17	1.93 ± 0.08	10
H4I5	45	86.78 ± 3.17	1.93 ± 0.07	21
H6L3	45	86.24 ± 4.82	1.92 ± 0.11	24
H4F6	50	95.13 ± 4.05	1.90 ± 0.08	12
H6L4	50	95.12 ± 3.23	1.90 ± 0.07	24
H4I6	50	93.93 ± 3.39	1.88 ± 0.07	20
Y5A5	50	91.73 ± 3.47	1.83 ± 0.07	10
Y6C4	50	90.99 ± 2.63	1.82 ± 0.04	1
H5F4	45	81.73 ± 2.85	1.82 ± 0.06	11
H6C3	45	81.42 ± 4.50	1.81 ± 0.10	10
T5A5	50	89.98 ± 3.74	1.80 ± 0.08	10
Y3A6	45	80.86 ± 5.63	1.80 ± 0.13	0

Continued on next page

ELP Sequence	Residue Number	ΔG per ELP ($k_B T$)	ΔG per residue ($k_B T$)	AL Round
S6A4	50	87.49 ± 1.37	1.75 ± 0.03	10
H5C5	50	87.27 ± 3.85	1.75 ± 0.08	10
Y6A4	50	87.13 ± 5.14	1.74 ± 0.10	10
T6C4	50	86.84 ± 4.23	1.74 ± 0.09	10
Y4A6	50	86.53 ± 4.31	1.73 ± 0.09	18
H3V6	45	77.80 ± 3.33	1.73 ± 0.07	0
T4C6	50	85.10 ± 4.60	1.70 ± 0.09	0
T6F4	50	84.34 ± 4.58	1.69 ± 0.09	11
Y6V3	45	75.30 ± 3.73	1.67 ± 0.08	17
H6I4	50	83.38 ± 4.26	1.67 ± 0.05	16
T4F6	50	82.96 ± 4.32	1.66 ± 0.09	13
W4A6	50	82.91 ± 2.94	1.66 ± 0.06	13
Y4F6	50	81.69 ± 2.98	1.63 ± 0.06	15
T6A3	45	72.03 ± 1.23	1.60 ± 0.03	15
W6A4	50	79.95 ± 3.32	1.60 ± 0.07	18
W6C4	50	78.30 ± 2.20	1.57 ± 0.04	0
S6F4	50	77.49 ± 1.99	1.55 ± 0.04	19
Y6F3	45	69.76 ± 2.26	1.55 ± 0.05	19
Y5F5	50	76.71 ± 2.88	1.53 ± 0.06	20
W5F5	50	74.55 ± 2.60	1.49 ± 0.05	0
S4C5	45	66.69 ± 3.56	1.48 ± 0.08	0
S4I5	45	62.73 ± 2.88	1.39 ± 0.06	0
Y4C6	50	66.68 ± 5.27	1.33 ± 0.11	10
Q5C5	50	66.40 ± 4.30	1.33 ± 0.09	0
S4A6	50	65.96 ± 3.53	1.32 ± 0.07	14
H5I5	50	65.71 ± 2.44	1.31 ± 0.05	23
Y6M4	50	64.91 ± 2.07	1.30 ± 0.04	17
D6A4	50	64.25 ± 3.55	1.29 ± 0.07	0
H5V5	50	63.55 ± 3.10	1.27 ± 0.06	0
Y3L6	45	56.20 ± 4.37	1.25 ± 0.10	0
H5L4	45	55.57 ± 2.88	1.24 ± 0.06	0
G6A4	50	59.35 ± 2.55	1.19 ± 0.05	0
R4F5	45	50.62 ± 3.15	1.13 ± 0.07	0
G4C5	45	49.72 ± 2.88	1.11 ± 0.06	0
E6A3	45	39.51 ± 1.89	0.88 ± 0.04	0
D5M4	45	38.25 ± 1.80	0.85 ± 0.04	0
K5M5	50	38.55 ± 2.60	0.77 ± 0.05	0
N6M3	45	33.93 ± 2.97	0.75 ± 0.07	0
E4L6	50	34.05 ± 3.30	0.68 ± 0.07	0
R6F4	50	15.90 ± 5.65	0.32 ± 0.11	0
R6M3	45	11.52 ± 4.37	0.26 ± 0.10	0

Table S2: Summary of experimentally realized amphiphilic ELP sequences forming vesicles, with their repeat numbers and residue compositions. The short sequences that are most directly comparable in size to the calculations conducted in this work are indicated in **bold**.

ELP Sequence	# Repeats	Hydrophilic	Hydrophobic	Ref.
E20F20	40	E	F	Huber *et. al.*[32]
$(R5Q5)_2$-F20	40	R, Q	F	Frank *et. al.*[33]
$(E4V)_4(F4V3F4)_4$	64	E	F, V	Vogele *et. al.*[34]
H40I30	70	H	I	Schreiber *et. al.*[35]
H20I30	50	H	I	Schreiber *et. al.*[35]
S40I30	70	S	I	Schreiber *et. al.*[35]
S20I30	50	S	I	Schreiber *et. al.*[35]
H40I30	70	G	I	Schreiber *et. al.*[35]
S20I30	50	S	I	Schreiber *et. al.*[35]
K20I30	50	K	I	Schreiber *et. al.*[35]
K40I30	70	K	I	Schreiber *et. al.*[35]
H10L4	14	H	L	Schreiber *et. al.*[35]
H5L4	9	H	L	Schreiber *et. al.*[35]
H5A5	10	H	A	Schreiber *et. al.*[35]
H5V5	10	H	V	Schreiber *et. al.*[35]
D10V5	15	D	V	Schreiber *et. al.*[35]
S48I48	96	S	I	Sharma *et. al.*[36]

References

(1) F. Sabanés Zariquiey; A. Pérez; M. Majewski; E. Gallicchio; G. De Fabritiis Validation of the Alchemical Transfer Method for the Estimation of Relative Binding Affinities of Molecular Series. *J. Chem. Inf. Model.* **2023**, *63*, 2438–2444.

(2) J. Z. Wu; S. Azimi; S. Khuttan; N. Deng; E. Gallicchio Alchemical Transfer Approach to Absolute Binding Free Energy Estimation. *J. Chem. Theory Comput.* **2021**, *17*, 3309–3319.

(3) S. Azimi; S. Khuttan; J. Z. Wu; R. K. Pal; E. Gallicchio Relative Binding Free Energy Calculations for Ligands with Diverse Scaffolds with the Alchemical Transfer Method. *J. Chem. Inf. Model.* **2022**, *62*, 309–323.

(4) L. Chen; Y. Wu; C. Wu; A. Silveira; W. Sherman; H. Xu; E. Gallicchio Performance and Analysis of the Alchemical Transfer Method for Binding-Free-Energy Predictions of Diverse Ligands. *J. Chem. Inf. Model.* **2023**, *64*, 250–264.

(5) D. Hamelberg; J. A. McCammon Standard Free Energy of Releasing a Localized Water Molecule from the Binding Pockets of Proteins: Double-Decoupling Method. *J. Am. Chem. Soc.* **2004**, *126*, 7683–7689.

(6) M. Aldeghi; A. Heifetz; M. J. Bodkin; S. Knapp; P. C. Biggin Accurate calculation of the absolute free energy of binding for drug molecules. *Chem. Sci.* **2016**, *7*, 207–218.

(7) Y. Qian; I. Cabeza de Vaca; J. Z. Vilseck; D. J. Cole; J. Tirado-Rives; W. L. Jorgensen Absolute Free Energy of Binding Calculations for Macrophage Migration Inhibitory Factor in Complex with a Druglike Inhibitor. *J. Phys. Chem. B* **2019**, *123*, 8675–8685.

(8) T. C. Beutler; A. E. Mark; R. C. . Schaik; P. R. Gerber; W. F. . Gunsteren Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **1994**, *222*, 529–539.

(9) A. z. Kubincová; S. Riniker; P. H. Hünenberger Reaction-field electrostatics in molecular dynamics simulations: development of a conservative scheme compatible with an atomic cutoff. *Phys. Chem. Chem. Phys.* **2020**, *22*, 26419–26437.

(10) S. J. Marrink; H. J. Risselada; S. Yefimov; D. P. Tieleman; A. H. de Vries The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.

(11) J. Hermans; L. Wang Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding. Application to a complex of benzene and mutant T4 lysozyme. *J. Am. Chem. Soc.* **1997**, *119*, 2707–2714.

(12) B. Lai; C. Oostenbrink Binding free energy, energy and entropy calculations using simple model systems. *Theor. Chem. Acc.* **2012**, *131*, 1–13.

(13) N. Goga; A. J. Rzepiela; A. H. de Vries; S. J. Marrink; H. J. C. Berendsen Efficient Algorithms for Langevin and DPD Dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 3637–3649.

(14) M. R. Shirts; J. D. Chodera Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.

(15) Z. Wu; D. L. Dotson; I. Alibay; B. K. Allen; M. S. Barhaghi; J. Hénin; T. T. Joseph; I. M. Kenney; H. Lee; H. Li; V. Lim; S. Liu; D. Marson; P. T. Merz; A. Schlaich; D. Mobley; M. R. Shirts; O. Beckstein alchemlyb: the simple alchemistry library. *J. Open Source Softw* **2024**, *9*, 6934.

(16) J. G. Kirkwood Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3*, 300–313.

(17) C. H. Bennett Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245–268.

(18) D. L. Mobley; J. P. Guthrie FreeSolv: a database of experimental and calculated hydration free energies, with input files. *J. Comput. Aided Mol. Des.* **2014**, *28*, 711–720.

(19) J. Wang; W. Wang; P. A. Kollman; D. A. Case Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247–260.

(20) A. W. Sousa da Silva; W. F. Vranken ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Research Notes* **2012**, *5*.

(21) G. Bussi; D. Donadio; M. Parrinello Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(22) J. J. Uusitalo; H. I. Ingólfsson; P. Akhshi; D. P. Tieleman; S. J. Marrink Martini Coarse-Grained Force Field: Extension to DNA. *J. Chem. Theory Comput.* **2015**, *11*, 3932–3945.

(23) D. H. de Jong; G. Singh; W. F. D. Bennett; C. Arnarez; T. A. Wassenaar; L. V. Schäfer; X. Periole; D. P. Tieleman; S. J. Marrink Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput.* **2012**, *9*, 687–697.

(24) L. Martínez; R. Andrade; E. G. Birgin; J. M. Martínez PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157–2164.

(25) M. Parrinello; A. Rahman Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(26) W. Humphrey; A. Dalke; K. Schulten VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.

(27) J. Kyte; R. F. Doolittle A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105–132.

(28) F. Pedregosa; G. Varoquaux; A. Gramfort; V. Michel; B. Thirion; O. Grisel; M. Blondel; P. Prettenhofer; R. Weiss; V. Dubourg; J. Vanderplas; A. Passos; D. Cournapeau; M. Brusher; M. Perrot; D. Duchesnay Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research* **2011**, *12*, 2825–2830.

(29) J. Huang; A. D. MacKerell Jr CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of Computational Chemistry* **2013**, *34*, 2135–2145.

(30) N. Michaud-Agrawal; E. J. Denning; T. B. Woolf; O. Beckstein MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* **2011**, *32*, 2319–2327.

(31) E. C. Meng; T. D. Goddard; E. F. Pettersen; G. S. Couch; Z. J. Pearson; J. H. Morris; T. E. Ferrin UCSF ChimeraX: Tools for structure building and analysis. *Protein Science* **2023**, *32*, e4792.

(32) M. C. Huber; A. Schreiber; P. von Olshausen; B. R. Varga; O. Kretz; B. Joch; S. Barnert; R. Schubert; S. Eimer; P. Kele; S. M. Schiller Designer amphiphilic proteins as building blocks for the intracellular formation of organelle-like compartments. *Nat. Mater.* **2014**, *14*, 125–132.

(33) T. Frank; K. Vogele; A. Dupin; F. C. Simmel; T. Pirzer Growth of Giant Peptide Vesicles Driven by Compartmentalized Transcription–Translation Activity. *Chem. Eur. J* **2020**, *26*, 17356–17360.

(34) K. Vogele; T. Frank; L. Gasser; M. A. Goetzfried; M. W. Hackl; S. A. Sieber; F. C. Simmel; T. Pirzer Towards synthetic cells using peptide-based reaction compartments. *Nat. Commun.* **2018**, *9*, 3862.

(35) A. Schreiber; M. C. Huber; S. M. Schiller Prebiotic Protocell Model Based on Dynamic Protein Membranes Accommodating Anabolic Reactions. *Langmuir* **2019**, *35*, 9593–9610.

(36) B. Sharma; Y. Ma; H. L. Hiraki; B. M. Baker; A. L. Ferguson; A. P. Liu Facile formation of giant elastin-like polypeptide vesicles as synthetic cells. *Chem. Commun.* **2021**, *57*, 13202–13205.