Supporting Information Methods for "Modeling Sequence-Specific Polymers Using Anisotropic Coarse-Grained Sites Allows Quantitative Comparison with Experiment"

Thomas K. Haxton, Ranjan Mannige, Ronald N. Zuckermann, and Stephen Whitelam
Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

S1. Bonded interactions

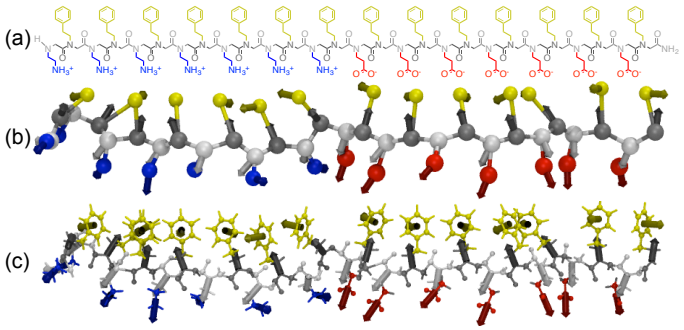

FIG. S1: (a) Chemical diagram, (b) coarse-grained representation, and (c) all-atom representation of a block-28 peptoid.

As shown in Fig. S1, our coarse-grained model consists of chains of coarse-grained sites, each characterized by both a position $\vec{r}$ and the direction $\hat{n}$ of a principle symmetry axis. Block-n peptoids consist of a chain of n backbone sites $X_i^{\text{back}} = (\vec{r}_i^{\text{back}}, \hat{n}_i^{\text{back}})$, $i = 0, ...n-1$, each linked to a single side chain site $X_i^{\text{side}} = (\vec{r}_i^{\text{side}}, \hat{n}_i^{\text{side}})$.

We included two types of bonded interactions, (1) backbone bonded interactions that couple the coordinates of backbone sites with the positions of neighboring backbone sites and (2) side chain bonded interactions that couple the coordinates of backbone sites with the coordinates of side chain sites. To create computationally efficient and accurate interaction potentials, we first identified scalar coordinates that could efficiently describe bonded configurations, then parametrized interaction potentials depending on these coordinates to reproduce potentials of mean force from all-atom simulations.

For the backbone interactions, we chose four scalar coordinates built from the vectors between backbone sites $\vec{r}_{i,i+1}^{\text{back}}$ and the backbone site directors $\hat{n}_i^{\text{back}}$: $|\vec{r}_{i,i+1}^{\text{back}}|$ describing the spacing between monomers, $\hat{r}_{i-1,i}^{\text{back}} \times \hat{n}_i^{\text{back}}$ and $\hat{n}_i^{\text{back}} \cdot \hat{r}_{i,i+1}^{\text{back}}$ describing bending, and $\left(\hat{r}_{i-1,i}^{\text{back}} \times \hat{n}_i^{\text{back}}\right) \cdot \left(\hat{r}_{i,i+1}^{\text{back}} \times \hat{n}_i^{\text{back}}\right)$ describing twist, planarity, and directional persistence. This last coordinate is $+1$ for a planar backbone persisting in a uniform direction, -1 for a planar backbone folding completely back on itself, 0 for a persistent backbone with a 90° twist, and 0 for a backbone with a 90° bend out of plane. For computational efficiency, it can be written in terms of scalar products as $\hat{r}_{i-1,i}^{\text{back}} \cdot \hat{r}_{i,i+1}^{\text{back}} - (\hat{r}_{i-1,i}^{\text{back}} \cdot \hat{n}_i^{\text{back}})(\hat{r}_{i,i+1}^{\text{back}} \cdot \hat{n}_i^{\text{back}})$. We parametrized the backbone interactions by matching potentials of mean force (PMFs) calculated from unconstrained all-atom molecular dynamics simulations of a minimal peptoid dimer (Fig. S2 (a)) solvated in water. We used MFTOID, a CHARMM-based peptoid force field developed in our lab [1], for the all-atom simulations. First, as shown in Fig. S2 (b), we constrained the spacing and twist coordinates by fitting fourth-order polynomials to the two one-dimensional PMFs depending on $|\vec{r}_{i,i+1}^{\text{back}}|$ and $\hat{r}_{i-1,i}^{\text{back}} \cdot \hat{r}_{i,i+1}^{\text{back}} - (\hat{r}_{i-1,i}^{\text{back}} \cdot \hat{n}_i^{\text{back}})(\hat{r}_{i,i+1}^{\text{back}} \cdot \hat{n}_i^{\text{back}})$. Then, we constrained the bend coordinates by fitting harmonic functions for the deviations of $\hat{r}_{i-1,i}^{\text{back}} \times \hat{n}_i^{\text{back}}$ and $\hat{n}_i^{\text{back}} \cdot \hat{r}_{i,i+1}^{\text{back}}$ from central values depending linearly on $|\vec{r}_{i,i+1}^{\text{back}}|$. Fitting to the two two-dimensional PMFs coupling $\hat{r}_{i-1,i}^{\text{back}} \times \hat{n}_i^{\text{back}}$ and $\hat{n}_i^{\text{back}} \cdot \hat{r}_{i,i+1}^{\text{back}}$ with $|\vec{r}_{i,i+1}^{\text{back}}|$ yielded good agreement with those PMFs, as shown in Fig. S2 (c). Without any additional fitting, we found the resulting potential energy functions gave PMFs in good agreement with the two two-dimensional PMFs coupling $\hat{r}_{i-1,i}^{\text{back}} \times \hat{n}_i^{\text{back}}$ and $\hat{n}_i^{\text{back}} \cdot \hat{r}_{i,i+1}^{\text{back}}$ with $\mathcal{F}\left(\hat{r}_{i-1,i}^{\text{back}} \cdot \hat{r}_{i,i+1}^{\text{back}} - (\hat{r}_{i-1,i}^{\text{back}} \cdot \hat{n}_i^{\text{back}})(\hat{r}_{i,i+1}^{\text{back}} \cdot \hat{n}_i^{\text{back}})\right)$ (Fig. S2 (d)). We arrived at a final interaction potential for an N-mer with the form

$$U_{\text{back}} = \sum_{i=1}^{N-1} \sum_{j=0}^{4} K_{1j} |\vec{r}_{i,i+1}^{\text{back}}|^j + \sum_{i=2}^{N-1} \sum_{j=0}^{4} K_{2j} \left(\hat{r}_{i-1,i}^{\text{back}} \cdot \hat{r}_{i,i+1}^{\text{back}} - (\hat{r}_{i-1,i}^{\text{back}} \cdot \hat{n}_i^{\text{back}})(\hat{r}_{i,i+1}^{\text{back}} \cdot \hat{n}_i^{\text{back}})\right)^j + \sum_{i=2}^{N} k_{\text{l}} \left(\hat{r}_{i-1,i}^{\text{back}} \cdot \hat{n}_i^{\text{back}} - \frac{|\vec{r}_{i,i-1}^{\text{back}}| - r_{0\text{l}}}{s_{\text{l}}}\right)^2 + \sum_{i=1}^{N-1} k_{\text{r}} \left(\hat{r}_{i,i+1}^{\text{back}} \cdot \hat{n}_i^{\text{back}} - \frac{|\vec{r}_{i,i+1}^{\text{back}}| - r_{0\text{r}}}{s_{\text{r}}}\right)^2. \quad \text{(S1)}$$

The best-fit interaction parameters are listed in Table SI.

For the side chain interactions, we identified four scalar

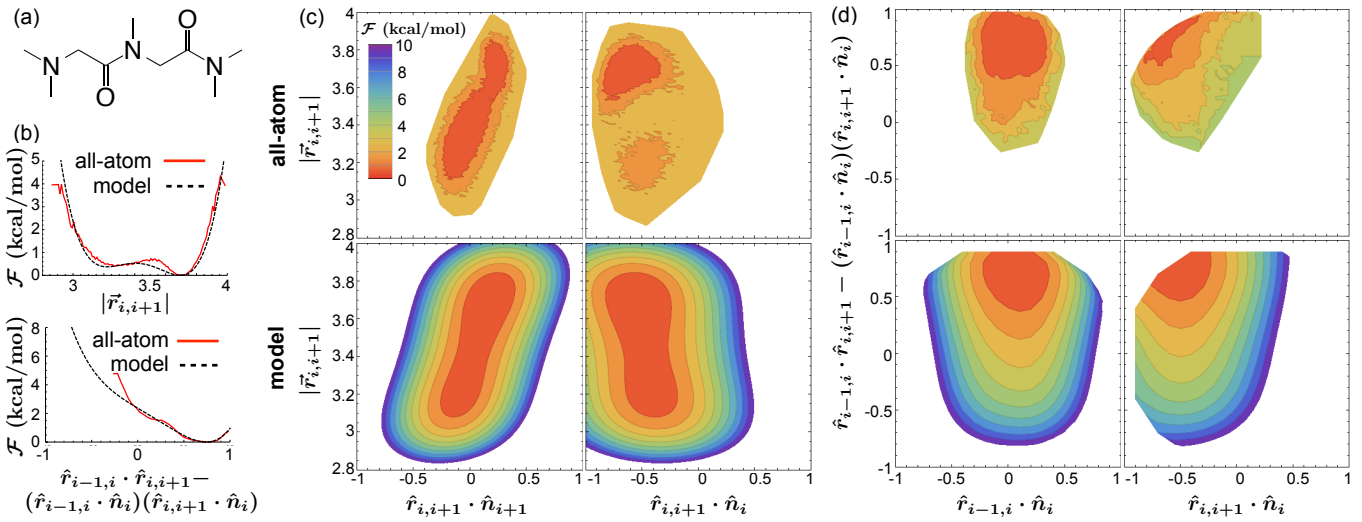

FIG. S2: (a) Molecular model used to parametrize the backbone bonded interactions. We parametrized the interactions by fitting to potentials of mean force calculated from solvated all-atom simulations of the molecular model. (b) First, we fit fourth-order polynomials to potentials of mean force for the spacing coordinate, $|\vec{r}_{i,i+1}^{\,\mathrm{back}}|$, and the twist coordinate, $\hat{r}_{i-1,i}^{\,\mathrm{back}} \cdot \hat{r}_{i,i+1}^{\,\mathrm{back}} - (\hat{r}_{i-1,i}^{\,\mathrm{back}} \cdot \hat{n}_i^{\,\mathrm{back}})(\hat{r}_{i,i+1}^{\,\mathrm{back}} \cdot \hat{n}_i^{\,\mathrm{back}})$. (c) Then, to constrain the bend coordinates, $\hat{r}_{i-1,i}^{\,\mathrm{back}} \times \hat{n}_i^{\,\mathrm{back}}$ and $\hat{n}_i^{\,\mathrm{back}} \cdot \hat{r}_{i,i+1}^{\,\mathrm{back}}$, we fit harmonic potentials of the form given in the last two terms of Eq. S1 to two-dimensional potentials of mean force involving the bend and spacing coordinates. (d) Without any additional fitting, we found that calculated two-dimensional potentials of mean force involving the bend and twist coordinates matched those of the all-atom simulations.

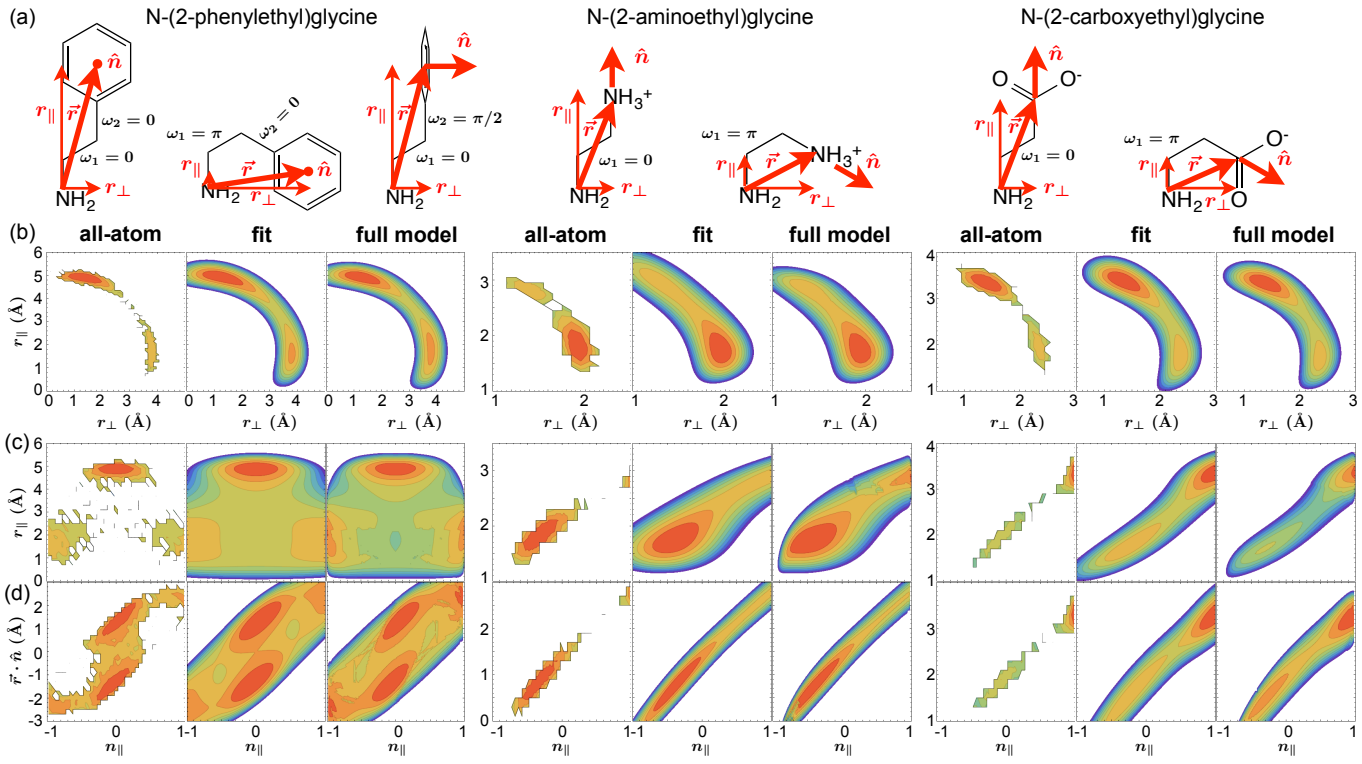

FIG. S3: (a) Molecular models (primary amines) used to paramerize the side chain bonded interactions. We parametrized the interactions by matching potentials of mean force calculated from unconstrained all-atom molecular dynamics simulations of the molecular models. First, we fit two-dimensional PMFs relating the positional coordinates $r_\parallel$ and $r_\perp$ (panel b). Second, subtracting the integrated effect of the first fitting, we fit two-dimensional PMFs relating $r_\parallel$ and $n_\parallel$ (panel c). Third, subtracting the integrated effect of the first and second fitting, we fit two-dimensional PMFs relating $\vec{r} \cdot \hat{n}$ and $n_\parallel$ (panel d). Finally, we numerically computed the PMFs for the full model (columns 3, 6, and 9 of panels b-d), confirming that they agreed well with the all-atom PMFs of columns 1, 4, and 7.

K_{10}	13595.3 kcal/mol	K_{20}	2.36497 kcal/mol
K_{11}	-15881.3 kcal/mol/Å	K_{21}	-3.87853 kcal/mol
K_{12}	6946.51 kcal/mol/Å^2	K_{22}	0.463066 kcal/mol
K_{13}	-1348.29 kcal/mol/Å^3	K_{23}	-3.33253 kcal/mol
K_{14}	97.9797 kcal/mol/Å^4	K_{24}	5.34237 kcal/mol
$k_{\rm l}$	28.2832 kcal/mol	$k_{\rm r}$	12.6891 kcal/mol
$r_{0\rm l}$	3.36096	$r_{0\rm r}$	1.51312
$s_{\rm l}$	1.95114	$s_{\rm r}$	-4.10588

TABLE SI: Best-fit parameters for backbone bonded interactions.

coordinates built from the backbone and side chain directors $\hat{n}_i^{\rm back}$ and $\hat{n}_i^{\rm side}$ and the vectors between backbone and side chain sites, $\vec{r}_i = \vec{r}_i^{\rm side} - \vec{r}_i^{\rm back}$. Dropping the monomer indices, the four coordinates are $r_\parallel = \hat{n}^{\rm back} \cdot \vec{r}$, $r_\perp = \left(|\vec{r}|^2 - {r_\parallel}^2\right)^{1/2}$, $n_\parallel = \hat{n}^{\rm back} \cdot \hat{n}^{\rm side}$, and $\vec{r} \cdot \hat{n} = \vec{r} \cdot \hat{n}^{\rm side}$. We parametrized the side chain interactions for each side chain type by matching PMFs calculated from unconstrained all-atom molecular dynamics simulations of the corresponding primary amines (Fig. S3 (a)) solvated in water, using the CHARMM force field [2]. Since the four coordinates are highly interrelated, we fit and validated the interaction parameters for each side chain type in four steps, as shown in Fig. S3 (b-d). First, we fit the two-dimensional PMF relating the positional coordinates $r_\parallel$ and $r_\perp$ (Fig. S3 (b)). Noting that $r_\parallel$ and $r_\perp$ follow a roughly circular orbit (Fig. S3 (b) as the atomic dihedral angle ω_1 (see Fig. S3 (a)) varies, we included a harmonic term constraining $r_\parallel$ and $r_\perp$ to a circular orbit. We added a fourth-order polynomial of $\arctan(r_\parallel / r_\perp)$ to model the nonuniform preference for dihedral angles ω_1. Second, subtracting the integrated effect of the first fitting, we fit the two-dimensional PMF relating $r_\parallel$ and $n_\parallel$. For the aminoethyl and carboxylethyl side chains, $\hat{n}^{\rm side}$ is parallel to the last carbon bond, so the coupling between $r_\parallel$ and $n_\parallel$ depends mostly on the dihedral angle ω_1, and we could fit to the PMFs using a function that varies harmonically in $n_\parallel$ from a quadratic curve in $(n_\parallel, r_\parallel)$ space. For the phenethyl side chain, $\hat{n}^{\rm side}$ is perpendicular to the last carbon bond, so the coupling between $r_\parallel$ and $n_\parallel$ depends strongly on two atomic dihedral angles, ω_1 and ω_2 (see Fig. S3 (a)). To accurately match the two-dimensional PMF, we fit a bivariate polynomial to the PMF, excluding odd terms in $n_\parallel$ that would destroy the reflection symmetry of $\hat{n}$. Third, subtracting the integrated effect of the first and second fitting, we fit the two-dimensional PMF relating $\vec{r} \cdot \hat{n}$ and $n_\parallel$. Again, the alignment of $\hat{n}^{\rm side}$ with the last carbon bond for aminoethyl and carboxyethyl side chains meant that we could fit their PMFs with a function that varies harmonically in $n_\parallel$ from a quadratic curve in $(n_\parallel, r_\parallel)$ space. For the phenethyl side chain, we again used a bivariate polynomial to fit the complex dependence on $\vec{r} \cdot \hat{n}$ and $n_\parallel$ as ω_1 and ω_2 vary. We included only terms with an even total power to respect the reflection symmetry of $\hat{n}$. Fourth and finally, compiling all terms of the potential energy function, we numerically computed the PMFs for the full model (columns 3, 6, and 9 of Fig. S3 (b-d)) and confirmed that they agreed well with the all-atom PMFs (columns 1, 4, and 7 of Fig. S3 (b-d)). The full form for the side chain interactions are

$$\begin{aligned} U_{\rm side} = k_1 \Big(\left((r_\perp - r_{\perp 0})^2 + (r_\parallel - r_{\parallel 0})^2\right)^{1/2} - r_0 \Big)^2 + \sum_{j=0}^{4} J_{1j} \left(\arctan(r_\parallel / r_\perp)\right)^j \\ + \sum_{j=0}^{6} J_{20j} {r_\parallel}^j + \sum_{j=0}^{2} J_{22j} {n_\parallel}^2 {r_\parallel}^j + J_{240} {n_\parallel}^4 \\ + \sum_{j=0,2,4} J_{30j} {n_\parallel}^j + \sum_{j=1,3} J_{31j} (\vec{r} \cdot \hat{n}) {n_\parallel}^j + \sum_{j=0,2} J_{32j} (\vec{r} \cdot \hat{n})^2 {n_\parallel}^j + J_{331} (\vec{r} \cdot \hat{n})^3 n_\parallel + J_{340} (\vec{r} \cdot \hat{n})^4 \end{aligned} \tag{S2}$$

for the phenylethyl side chain and

$$\begin{aligned} U_{\rm side} = k_1 \Big(\left((r_\perp - r_{\perp 0})^2 + (r_\parallel - r_{\parallel 0})^2\right)^{1/2} - r_0 \Big)^2 + \sum_{j=0}^{4} J_{1j} \left(\arctan(r_\parallel / r_\perp)\right)^j \\ + k_2 \left(n_\parallel - \sum_{j=0}^{2} r_{2j} {r_\parallel}^j) \right)^2 + k_3 \left(\vec{r} \cdot \hat{n} - \sum_{j=0}^{2} n_{3j} {n_\parallel}^j) \right)^2 \end{aligned} \tag{S3}$$

for the aminoethyl and carboxyethyl side chains. The parameters appear in Table SII.

Finally, we note that we coupled the side chain positional coordinates only with the backbone director, $\hat{n}^{\rm back}$, and not with the positions or orientations of neighboring backbone sites. This allowed us to simply decompose the bonded interactions into side chain and backbone terms, as discussed above, and accounts for the flexibility expected in the single the bond between the backbone nitrogen and the first (C_β) carbon of the side chain.

	phenethyl	aminoethyl	carboxyethyl		phenethyl		phenethyl		aminoethyl	carboxyethyl
k_1	23.086	74.463	51.4346	J_{200}	-1.88336	J_{300}	2.84809	k_2	12.8785	30.9392
$r_{\perp 0}$	0.343266	-0.891764	0.389628	J_{201}	-2.02912	J_{302}	-13.0286	r_{20}	-0.0587984	-2.08987
$r_{\parallel 0}$	1.60786	0.999481	1.63953	J_{202}	12.0487	J_{304}	33.5248	r_{21}	-1.024	1.19511
r_0	3.44926	2.93893	1.98906	J_{203}	-10.0584	J_{311}	9.4368	r_{22}	0.476617	-0.081282
J_{10}	13.942	418.586*	193.754	J_{204}	3.44353	J_{313}	-45.2463	k_3	50.2083	23.0679
J_{11}	-82.0459	-1837.99*	-937.961	J_{205}	-0.536649	J_{320}	-2.36074	n_{30}	1.3901	1.94004
J_{12}	186.411	2964.05	1671.22	J_{206}	0.0315484	J_{322}	24.2649	n_{31}	1.55449	1.58382
J_{13}	-164.93	-2079.92	-1279.85	J_{220}	1.61338	J_{331}	-6.30323	n_{32}	-0.155606	-0.210233
J_{14}	49.0435	537.819	355.337	J_{221}	-4.38724	J_{340}	0.711079			
				J_{222}	1.24471					
				J_{240}	0.663689					

TABLE SII: Best-fit parameters for the side chain bonded interactions. *After fine-tuning, J_{10} and J_{11} for aminoethyl sites were adjusted from 418.586 and -1837.99 to 424.643 and -1844.99 (see Section S7).

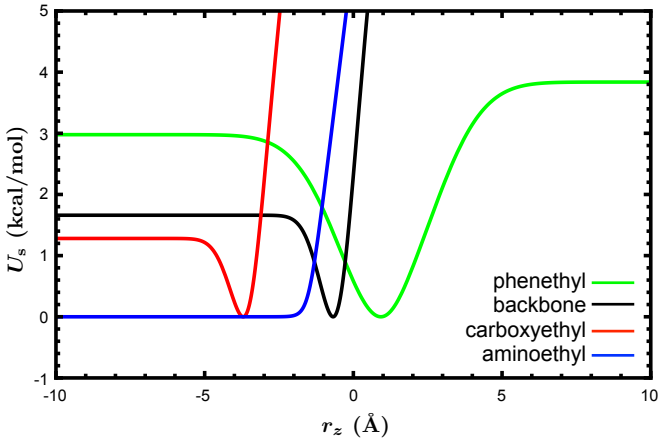

FIG. S4: Solvation interaction energies $U_s(r_z)$ for the four coarse-grained sites, shifted with respect to their minimum values.

S2. Solvation interactions

We modeled the solvation energy as a sum of a sigmoidal term interpolating between bulk and vacuum energies and a Gaussian term modeling the affinity for sites to the interface,

$$U_s(r_z) = U_{bs} - k_B T_r \ln\left(1 + \frac{\exp(U_{bs}/k_B T_r) - 1}{1 + \exp(-4(r_z - z_{int})/w)}\right) + U_{int} \exp\left(-\frac{1}{2}\left(\frac{r_z - z_{int}}{\sigma_{int}}\right)^2\right). \quad \text{(S4)}$$

In Eq. S4 $T_r = 300$ K is room temperature, and we tuned the other parameters to match experimental and/or all-atom simulation data at room temperature, as listed in Table SIII. We set the bulk solvation energy U_{bs} to match experimental values for the analog molecules [3–5] for the analog molecules. We tuned U_{int}, z_{int}, and σ_{int} to match free energy profiles $\mathcal{F}(z)$ calculated with all-atom simulations of the analog molecules where such profiles were available [6–8]. Since free energy profiles for N-methylacetamide were not available, we used data for the related molecule acetamide [6], which has a very similar solvation free energy. We found that free-energy profiles were relatively insensitive to the value of w, so we fixed it at $w = 1$ Å. The best-fit parameters are listed in Table SIII.

In fitting our interfacial parameters to all-atom free energies, one choice would be to directly match $U_s(z) = \mathcal{F}(z)$. However, such an approach would assign an average interfacial free energy to an instantaneous, local potential energy function. Since fluctuations in the position of the air-water interface are large and likely couple to fluctuations in the solute distribution functions, we did not want to average over those fluctuations when constructing a potential energy function. Instead, we fixed the interfacial parameters by minimizing (by eye, using the published plots) the residual of the targeted equality

$$\mathcal{F}(z) = -k_{\mathrm{B}}T \ln\left(\int_{-\infty}^{\infty} dz' P_{\mathrm{i}}(z') \exp\left(-U_{\mathrm{s}}(z-z')/k_{\mathrm{B}}T\right)\right), \tag{S5}$$

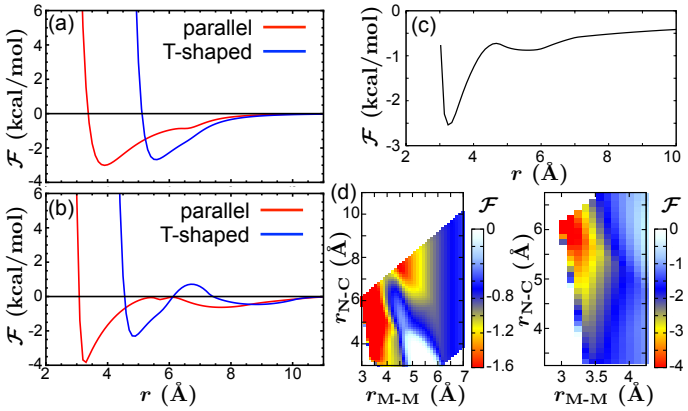

FIG. S5: (a) Potentials of mean force for parallel phenethyl side chain sites, perpendicular phenethyl side chain sites, and oppositely charged side chain sites. Parallel phenethyl potentials of mean force were calculated by fixing site directors to be parallel and perpendicular with respect to each other. The potential of mean force between charged side chain sites was unconstrained but calculated relative to the distance between charged sites at the aminoethyl N and the carboxyethyl carboxyl C. (b) Two-dimensional potentials of mean force for the oppositely charged side chain sites as a function of the charged-charged separation and the separation between methyl carbons.

where $P_{\mathrm{i}}(z)$ is the distribution function for the location of the air-water interface. Assuming that $P_{\mathrm{i}}(z)$ is Gaussian and that the instantaneous, local interface is sharp, we get an error function for the average water density profile. We independently constrained $P_{\mathrm{i}}(z)$ by fitting the error function to the water density profile for each all-atom simulation, each time finding excellent fits with distribution widths between 1.5 and 1.7 Å.

Fig. S4 shows $U_{\mathrm{s}}(r_z)$ for the four coarse-grained sites, shifted with respect to their minimum values. The phenethyl side chain site shows the strongest affinity to the air-water interface, serving to anchor peptoid monolayers to the interface. The backbone and carboxyethyl side chain sites also contribute surface affinity. The relative locations of the minima help to stabilize the amphiphilic pattern, with phenethyl, backbone, and charged sites localizing at increasing depths.

S3. Nonbonded interactions

Analyzing all-atom simulations of peptoid bilayers [12], we found that interactions between phenethyl side chains and between charged side chains are the dominant nonbonded interactions. For computational efficiency, we therefore treated all other nonbonded interactions as isotropically hard-core repulsive with radii described below.

Since the phenethyl side chain site would sample environments in both water and air (the latter in the case of monolayers, with the site sampling mostly near its free energy minimum at the interface), we modeled nonbonded interactions between phenethyl side chain sites in two ways, one representing a gas environment and one representing a solvated environment.

Since we expected that interactions between the uniaxially symmetric π systems should dominate the aromatic-aromatic interactions, we based both the gas and solvated phenethyl-phenethyl interactions on the uniaxial Gay-Berne potential [13], a generalized Lennard-Jones potential that allows for separate variation of shape and energy anisotropy. The Gay-Berne potential has previously been used to model benzene [14–16]. Here, we targeted toluene, the molecular analog of the phenethyl side chain site. The form of the potential is

$$U_{\mathrm{GB}}(X_i, X_j) = 4\epsilon_0 \epsilon(X_i, X_j; \kappa') \left(\left(\frac{\xi\sigma_0}{|r_{ij}| - \sigma_0\sigma(X_i, X_j) + \xi\sigma_0} \right)^{12} - \left(\frac{\xi\sigma_0}{|r_{ij}| - \sigma_0\sigma(X_i, X_j) + \xi\sigma_0} \right)^{6} \right), \tag{S6}$$

where

$$\epsilon(X_i, X_j; \kappa') = \left(1 - \chi(\kappa)^2 \left(\vec{n}_i \cdot \vec{n}_j\right)^2\right)^{-\nu/2} \left(1 - \frac{\chi'(\kappa')}{2} \left(\frac{\hat{r}_{ij} \cdot \vec{n}_i + \hat{r}_{ij} \cdot \vec{n}_j}{1 + \chi'(\kappa') \vec{n}_i \cdot \vec{n}_j} + \frac{\hat{r}_{ij} \cdot \vec{n}_i - \hat{r}_{ij} \cdot \vec{n}_j}{1 - \chi'(\kappa') \vec{n}_i \cdot \vec{n}_j} \right)\right)^{\mu}, \tag{S7}$$

$$\sigma(X_i, X_j) = \left(1 - \frac{\chi(\kappa)}{2} \left(\frac{\left(\hat{r}_{ij} \cdot \vec{n}_i + \hat{r}_{ij} \cdot \vec{n}_j\right)^2}{1 + \chi(\kappa) \vec{n}_i \cdot \vec{n}_j} + \frac{\left(\hat{r}_{ij} \cdot \vec{n}_i - \hat{r}_{ij} \cdot \vec{n}_j\right)^2}{1 - \chi(\kappa) \vec{n}_i \cdot \vec{n}_j} \right)\right)^{-1/2}, \tag{S8}$$

site	analog	$U_{\rm bs}$ (kcal/mol)	$U_{\rm int}$ (kcal/mol)	$z_{\rm int}$ (Å)	$\sigma_{\rm int}$ (Å)
backbone	N-methylacetamide	**-10.1** [3]			
	acetamide	-9.7 [3]	**-1.9**	**-0.5**	**0.6** [6]
Npe side chain	toluene	**-0.76** [4]	**-3.6**	**1.2**	**1.6** [6]
Nae side chain	methylammonium	**-71.3** [5]			
	ammonium	-70 [9, 10]	**0**	**-1.5**	**N/A** [11]
Nce side chain	acetate	**-79.9** [5]	**-1.5**	**-3.5**	**0.6** [7, 8]

TABLE SIII: Solvation parameters used in the model. The third columns lists experimental solvation free energies, along with references. The last three columns list model parameters determined by matching all-atom simulation free energy profiles published in the listed references. Parameters used in the model are listed in bold.

$$\chi(\kappa) = \frac{\kappa^2 - 1}{\kappa^2 + 1}, \tag{S9}$$

and

$$\chi'(\kappa') = \frac{1 - \kappa'^{1/\mu}}{1 + \kappa'^{1/\mu}}. \tag{S10}$$

As in Refs. [14–16] we also included a standard quadrupole electrostatic term in the gas environment,

$$\begin{aligned} U_{\rm q}(X_i, X_j) = \frac{3CQ^2}{4|\vec{r}_{ij}|^5} \Big(1 + 2\left(\hat{n}_i \cdot \hat{n}_j\right)^2 - 5\left(\left(\hat{r}_{ij} \cdot \hat{n}_i\right)^2 + \left(\hat{r}_{ij} \cdot \hat{n}_j\right)^2\right) \\ -20\left(\hat{r}_{ij} \cdot \hat{n}_i\right)\left(\hat{r}_{ij} \cdot \hat{n}_j\right)\left(\hat{n}_i \cdot \hat{n}_j\right) + 35\left(\hat{r}_{ij} \cdot \hat{n}_i\right)^2\left(\hat{r}_{ij} \cdot \hat{n}_j\right)^2\Big), \end{aligned} \tag{S11}$$

ϵ_0	attractive energy scale	1.8 kcal/mol	2.7 kcal/mol
σ_0	size	5.6 Å	6.2 Å
κ	shape anisotropy	0.54	
κ'	energy anisotropy	3.6	
ξ	potential shape parameter	0.5	

TABLE SIV: Gay-Berne parameters for nonbonded interactions between phenethyl side chain sites, adjusted to match all-atom potentials of mean force for toluene [17]. The last column lists fine-tuned values (see text).

where $C = 332.06$ (kcal/mol)Å$/e^2$ is Coulomb's constant and $Q = 0.6$ eÅ^2 is the quadrupole moment of toluene. Quadrupole interactions have been shown to stabilize the favored "slipped-parallel" configuration of benzene [17]. Although toluene also has a significant dipole interaction, we expected that the anti-parallel configurations favored by these interactions would be suppressed in the phenethyl side chains by steric interactions between the aromatic ring of one side chain and the hydrocarbon chain of the other. We therefore did not attempt to model dipole-dipole interactions of the phenethyl side chain sites. This simplification allowed us to treat the phenethyl side chain as a site with a single axis of symmetry, consistent with our coarse-graining scheme.

We set the parameters of the Gay-Berne potential by matching potentials of mean force from all-atom simulations of toluene in vacuum [17]. We fixed the exponents $\mu = 1$ and $\nu = -2$ to maximize computational efficiency. The remaining parameters are listed in Table SIV. We obtained excellent agreement between coarse-grained and atomistic potentials of mean force for both parallel and perpendicular relative orientations, as demonstrated by comparing the red and blue curves of Fig. S5 (a) with the solid and long-dashed curves of Fig. 2 (b) of Ref. [17]. As discussed in section S7, we later fine-tuned two of the parameters, the energy scale and the size, to improve the agreement with experimental X-ray scattering and all-atom simulations. The fine-tuned parameters appear in the last column of Table SIV.

To model the phenethyl-phenethyl interaction in a solvated environment, we included both a short-range repulsive term penalizing voids smaller than a water molecule and a longer-range solvent-separated attractive term, each using the same shape anisotropy as the Gay-Berne potential but with separately adjustable energy anistropy:

$$U_{\rm sm}(X_i,X_j)=\begin{cases}\epsilon_{\rm r}\epsilon(X_i,X_j;\kappa_{\rm r})\left(1-\frac{4}{w^2}\left(|\vec{r}_{ij}|-\sigma_0\sigma(X_i,X_j)-\frac{w}{2}\right)^2\right), & |\vec{r}_{ij}|-\sigma_0\sigma(X_i,X_j)<w\\ -\epsilon_{\rm ss}\epsilon(X_i,X_j;\kappa_{\rm ss})\left(1-\frac{4}{w^2}\left(|\vec{r}_{ij}|-\sigma_0\sigma(X_i,X_j)-\frac{3w}{2}\right)^2\right), & w<|\vec{r}_{ij}|-\sigma_0\sigma(X_i,X_j)<2w\\ 0, & |\vec{r}_{ij}|-\sigma_0\sigma(X_i,X_j)>2w\end{cases} \tag{S12}$$

Because electrostatic interactions are strongly screened in solution, we removed the quadrupole term in the solvated environment. We interpolated the interaction potential between the solvated and gas forms using a sigmoidal function with the same middle ($z_{\rm int}=1.2$ Å) and width ($\sigma_{\rm int}=1.6$ Å) as the interfacial interaction for the phenethyl side chain,

$$S(z)=\frac{1}{1+\exp\left(4(z-z_{\rm int})/\sigma_{\rm int}\right)}, \tag{S13}$$

leading to the total interaction between phenethyl side chains

$$U_{\rm nb}^{\rm p-p}(X_i,X_j)=U_{\rm GB}(X_i,X_j)+\frac{1}{2}\left(S(r_{iz})+S(r_{jz})\right)U_{\rm sm}(X_i,X_j)+\left(1-\frac{1}{2}\left(S(r_{iz})+S(r_{jz})\right)\right)U_{\rm q}(X_i,X_j). \tag{S14}$$

We set the parameters of the solvated potential by matching potentials of mean force from all-atom simulations of toluene in water [17]. The parameters are listed in Table SV. We obtained excellent agreement between coarse-grained and atomistic potentials of mean force for both parallel and perpendicular relative orientations, as demonstrated by comparing the red and blue curves of Fig. S5 (b) with the solid and long-dashed curves of Fig. 3 (b) of Ref. [17]. As discussed in Section S7, we later fine-tuned the size and overall energy scale of the phenethyl side chain sites. The fine-tuned values appear in the last column of Table SV.

We modeled interactions between the charged aminoethyl and carboxyethl side chain sites as a combination of a hard-core repulsive interaction, an excluded solvent interaction, and a screened Coulomb interaction,

$$U_{\rm cc}(X_i,X_j)=U_{\rm ex}\left(|\vec{r}_{ij}|-(R_i+R_j)\right)+U_{\rm el}\left(|(\vec{r}_j+\delta_j\hat{n}_j)-(\vec{r}_i+\delta_i\hat{n}_i)|\right), \tag{S15}$$

where

$$U_{\rm ex}(\Delta r)=\begin{cases}\infty, & \Delta r<0\\ f_{\rm sh}(U_{{\rm bs}i}+U_{{\rm bs}j})\left(1-4\left(\frac{\Delta r-r_{\rm sh}/2}{r_{\rm sh}}\right)^2\right), & 0<\Delta r<r_{\rm sh}\\ 0, & \Delta r>r_{\rm sh},\end{cases} \tag{S16}$$

$$U_{\rm el}(r)=\frac{Cq_iq_j\exp(-r/\lambda_{\rm D})}{\epsilon\left(r-(R_i-\delta_i+R_j-\delta_j)\right)r}, \tag{S17}$$

and

$$\epsilon(r)=\begin{cases}\epsilon_0+(\epsilon_\infty-\epsilon_0)r/\lambda_\epsilon, & r<\lambda_\epsilon\\ \epsilon_\infty, & r>\lambda_\epsilon.\end{cases} \tag{S18}$$

In these equations, R_i is the hard-core radius, $U_{{\rm bs}i}$ is the bulk solvation energy, C is Coulomb's constant, and

ϵ_0	attractive energy scale	2 kcal/mol	3 kcal/mol
σ_0	size	5.6 Å	6.2 Å
κ	shape anisotropy	0.54	
κ'	energy anisotropy	10	
ξ	potential well shape parameter	0.5	
w	solvation width	2.8 Å	
ϵ_r	solvent repulsive energy	1 kcal/mol	1.5 kcal/mol
κ_r	repulsive energy anisotropy	25	
ϵ_{ss}	solvent-separated attractive energy	0.3 kcal/mol	0.45 kcal/mol
κ_{ss}	solvent-separated energy anisotropy	4	

TABLE SV: Parameters for nonbonded interaction between phenethyl side chain sites in a solvated environment, adjusted to match all-atom free energies of toluene in water [17]. The last column represents adjusted parameters due to fine-tuning of the size and overall energy scale of the phenethyl side chain sites (see text).

$\lambda_D = (\epsilon_\infty k_B T/8\pi C c_{salt})$ is the Debye length. We set the sum of the aminoethyl and carboxylethyl hard-core radii equal to 3.88 Å, equal to the center-of-mass separation for the minimum energy configuration in the CHARMM forcefield [2]. We partitioned this distance between aminoethyl (1.82 Å) and carboxyethyl (2.08 Å) according to their mass-weighted radii of gyration. We fixed the charge q_i to be 1 for the aminoethyl side chain site and -1 for the carboxyethyl side chain site. We conducted most of our simulations at a salt concentration of $c_{salt} = 0.2$ M. The distance δ_i reflects the fact that we place the center of the hard sphere at the center of mass of the molecular analog, while we place the charge at the atomic site associated with the greatest charge (N for aminoethyl and carboxyl C for carboxyethyl), allowing for anisotropic interactions favoring contact between the more highly charged ends of the side chains. Using atomic coordinates at rest in the CHARMM force field [2], we fixed $\delta_i = 0.696$ Å for the aminoethyl side chain and $\delta_i = 0.151$ Å for the carboxyethyl side chain. The excluded solvent interaction penalizes charged sites whose hard cores approach within a solvation shell of width r_{sh}, applying a maximum energy penalty equal to f_{sh} times the solvation energy for each site. The Coulomb interaction uses a distance-dependent permittivity that linearly decays from ϵ_0 at contact to the bulk water value, $\epsilon_\infty = 80$, at a separation of λ_D.

We fixed the four remaining parameters of the charged-charged interaction to match free energies calculated using all-atom simulations of acetate and methylammonium in water [18]. We matched both one-dimensional free energies between methylammonium N and acetate carboxyl C and two-dimensional free energies calculated as a function of both this distance and the distance between methyl carbons on each molecule. We did not attempt to reproduce the local minimum corresponding to hydrophobic attraction between methyl groups, since this geometry would be sterically suppressed by the hydrocarbon chains of the side chains. With the exception of this feature, we were able to match potentials of mean force for our model and the all-atom simulations remarkably well, as demonstrated by the agreement of Fig. S5 (c) with Fig. 1 (a) of Ref. [18] and of Fig. S5 (d) with Fig. S2 (d) of Ref. [18]. The best-fit parameters are listed in Table SVI.

r_{sh}	1.5 Å
f_{sh}	0.01
ϵ_0	20
λ_D	4 Å

TABLE SVI: Best-fit parameters for nonbonded interactions between charged side chain sites, adjusted to match all-atom potentials of mean force for methylammonium and acetate in solution [18].

We used the same functional form and parameters for the repulsive interactions between like-charged side chain sites. As far as we know, nobody has conducted all-atom simulations of the analogous like-charged molecular analog systems.

For computational efficiency, we cut off the phenethyl-phenethyl and charged-charged interactions at ranges beyond which interaction energies are never stronger than 0.1 kcal/mol. We used a cutoff of 10.98 Å for phenethyl side chains (11.67 Å after fine-tuning) and a salt-concentration dependent cutoff for charged side chains, equal to 14.12 Å at $c_{salt} = 0.2$ M.

We let the remaining nonbonded interactions (between backbone and side chain sites, and between phenethyl and charged side chain sites) be hard-sphere repulsions. For charged side chain sites, we used the same hard-core radii as in the electrostatic interactions. For phenethyl side chain sites, we used the width of the repulsive part, $\chi\kappa\sigma_0 = 1.512$ Å. For backbone sites, we initially used the radius of the repulsive part of the nitrogen atom's Lennard-Jones interaction in CHARMM [2], 1.65 Å. As discussed in Section S7, we later fine-tuned this parameter (to 2.05 Å) to improve the model's agreement with experimental X-ray scattering and all-atom simulations.

S4. Monte Carlo simulations

We calculated equilibrium properties of solvated peptoids, interfacial peptoids, monolayers, and bilayers by performing Monte Carlo simulations in the appropriate

thermodynamic ensembles, starting from initial conditions described in Sec. S5. In all cases, we used rectangular prism simulation boxes with periodic boundary conditions in three dimensions. When using an interface, as for interfacial peptoids and monolayers, we fixed two interfaces in the xy plane, bounding a slab of (implicit) water, initializing the peptoids or monolayer on one of the two interfaces. We simulated individual solvated and interfacial peptoids at fixed volume and temperature. We simulated monolayers at fixed number, water slab volume, temperature, and in-plane pressure ($NVTp_s$) by using Monte Carlo moves that change the side lengths of the simulation box while maintaining both the surface pressure and the total volume of the water slab. We simulated the bilayer at fixed $NVTp_s$ with $p_s = 0$. In both cases, we sampled the variable-aspect-ratio ensemble by selecting at random a dimension $\alpha \in \{x, y, z\}$, changing the corresponding box side from L_α to $\exp(r)L_\alpha$, where r is a random number between $-\Delta_{\text{aspect}}$ and Δ_{aspect}, changing the other box sides from L_β to $\exp(-r/2)L_\beta$, scaling all coordinates by the same factors, and accepting the move according to a Metropolis criterion for the weight $\exp(-(U + (L_xL_y)p_s)/k_{\mathbf{B}}T)$, where U is the potential energy. We used $\Delta_{\text{aspect}} = 0.002$ and attempted the moves with frequency $0.15/N$, where N is the number of coarse-grained sites. Note that the hydrostatic pressure is not relevant because the effects of water at atmospheric pressure are implicitly included in the coarse-grained potential energy function.

In addition to the box side length Monte Carlo moves, we used two kinds of local moves of single coarse-grained sites: translations within a cube of width 0.8 Å and rotations of the symmetry axes from $\hat{n}$ to $\hat{n}'$, where $\hat{n}' = (\hat{n} + \vec{r})/|\hat{n} + \vec{r}|$ and $\vec{r}$ is a random vector chosen within a box of width 1.6. We attempted translation moves with frequency 0.5, and attempted rotation moves with the remainder of the frequency.

We simulated stacked bilayers used in X-ray scattering experiments by creating a periodic stack of n distinct bilayers as follows. We replicated a bilayer, stacked it in the z direction with 6 nm of (implicit) solvent in between, performed $NVTp_s$ simulations to allow the bilayers to separately evolve, then performed additional simulations including "evaporation moves." In an evaporation move, the space between two bilayers is selected at random, changed by a value between -0.4 Å and 0.4 Å, and accepted according to a normal Metropolis criterion. We found that it was not necessary to bias the bilayers to "evaporate" water; when bilayers diffused into contact via the evaporation moves, they tended to stay in contact due to attractive interactions.

Using Monte Carlo simulations instead of molecular dynamics simulations allowed us to efficiently sample configuration space of a model with complex, anisotropic pair and three-body interactions. Although (in contrast to molecular dynamics) our Monte Carlo simulations provide only a rough approximation of realistic dynamics, an estimate for the equivalent time scales sampled in the simulations can be made by calculating diffusion in the dilute limit. For a hypothetical isolated site, the translational diffusion coefficient is $D_t = p_t{\Delta_t}^2/6\tau_0$, where $p_t = 0.5$ is the probability to make a translation move, $\Delta_t = 0.4$ Å is the maximum step size, and τ_c is the time per Monte Carlo cycle. Equating this with the Stokes-Einstein relation for diffusion of a coarse-grained site of typical radius 2.05 Å (backbone site), this sets the time scale to be $\tau = 1.2 \times 10^{-13}$ sec. Typical simulations on one CPU simulated order 100 peptoids (order 10^4 sites) for 10^8 to 10^9 Monte Carlo cycles, corresponding to an ellapsed time of 10 to 100 μs.

We optimized the speed of our simulations by employing a hybrid neighbor-finding scheme for nonbonded interactions. We found that it was most efficient to use both neighbor lists and a cell structure with linked lists [19]. That is, we identified candidate interacting pairs from neighbor lists, which we generated from pairs in neighboring cells. By using a neighbor list, we were able to search for interacting pairs within a sphere of volume $4\pi r^3/3 \approx 4.2r^3$, where r is the interaction range, rather than within a nine-cell cube of volume $9r^3$. We updated neighbor lists and linked lists after each accepted translation move, since this generated little additional overhead. Because the interaction cutoff between phenethyl side chain sites (11.67 Å) and between charged side chain sites (14.1151 Å at 0.2 M salt) is significantly larger than the interaction cutoff for the more numerous, purely repulsive backbone-backbone and mixed-type interactions (3.71 Å for backbone-carboxyethyl side chain interactions), we found that it was most efficient to use separate neighbor lists and linked lists for phenethyl-phenethyl, charged-charged, and short-range interactions.

After optimizing the neighbor-finding scheme, the calculation of energies between the long-range-interacting anisotropic side chain sites became the costliest part of our simulations, comprising around 63% of the total simulation time. Even compounding this cost with our need to perform both rotations and translations, we estimate that our model is no more than six times slower than a similarly optimized simulation of a system with simple isotropic interactions of a similar range (e.g. screened Coulomb).

S5. Initialization

Although the equilibrium properties of peptoid phases should in principle be independent of initial conditions, for computational efficiency we chose to start from well-ordered initial conditions, since melting unstable ordered structures typically occurs faster than the nucleation and growth of ordered phases. We initialized peptoids, monolayers, and bilayers in low-energy, ordered configurations as sketched in Fig. S6 (a-i). Anticipating that extended peptoid backbones in monolayers and bilayers would zigzag every two monomers to allow greater sep-

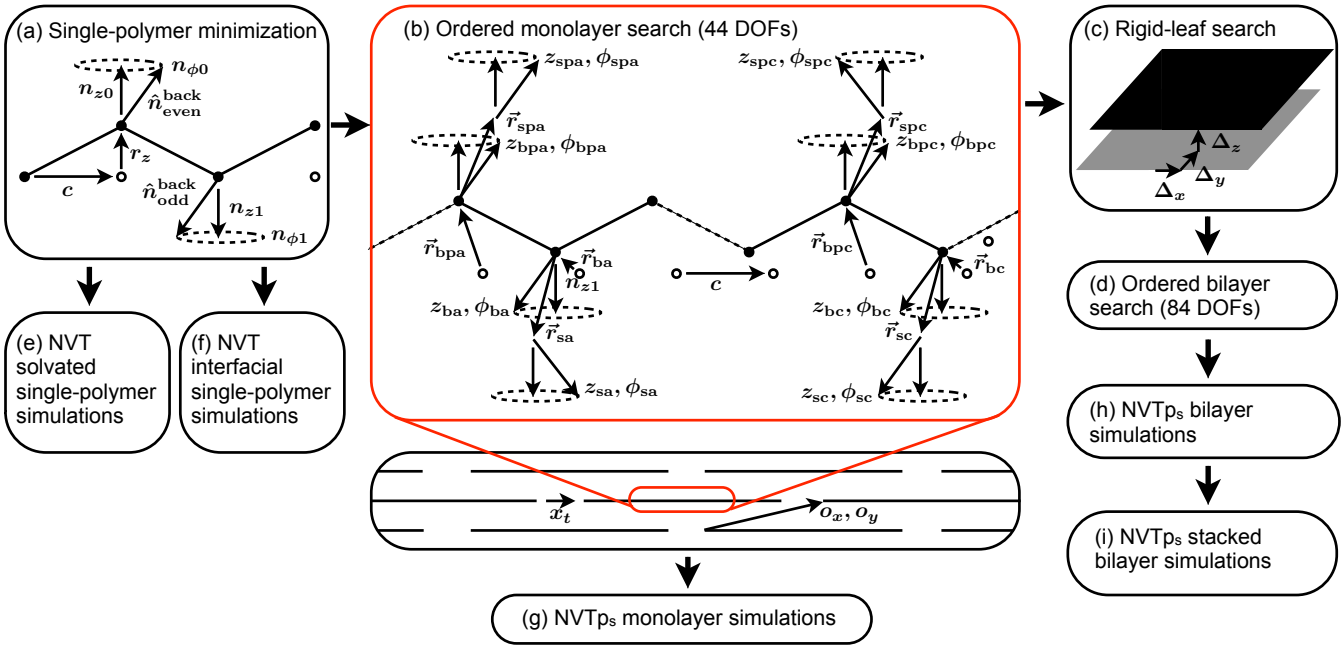

FIG. S6: Flow chart showing how we initialized our simulations. (a) First, we minimized the potential energy of a single peptoid constrained to remain in an extended, ordered configuration with a two-monomer repeat unit. (b) Starting from a perfect brick arrangement of these minimized peptoids, we found a low-energy ordered monolayer by performing a Monte Carlo search within a space of ordered monolayers. (c) Starting from two copies of the minimum-energy ordered monolayer, we rigidly moved the two leaves to find the lowest-energy bilayer. (d) Starting from this bilayer configuration, we then found a low-energy ordered bilayer by performing a Monte Carlo search within a space of ordered bilayers. Finally, starting from the appropriate energy-minimized configurations, we performed equilibrium simulations of (e) isolated solvated peptoids, (f) monolayers, (g) isolated bilayers, and (h) stacked bilayers.

aration between alternating charged and nonpolar side chains, we first searched for low-energy backbone configurations subject to this zigzag constraint, shown in Fig. S6 (a). Numerically minimizing over the six degrees of freedom $(c, r_z, n_{z0}, n_{\phi 0}, n_{z1}, n_{\phi 1})$ characterizing the zigzag pattern, we found a backbone configuration with an energy only 0.25 kcal/mol/monomer greater than the energy of the unconstrained minimum. This low value demonstrates the ability of the peptoid backbone to accommodate a planar, extended, amphiphilic pattern. We attached side chains to this backbone in their fully energy-minimized configurations relative to the backbone directors $\hat{n}_i$.

Next, as illustrated in Fig. S6 (b), we constructed low-energy monolayer by performing a Monte Carlo search on an ordered monolayer of block-28 peptoids. We initialized the monolayer by replicating a low-energy extended peptoid in the xy plane, alternately offsetting rows of peptoids by half a peptoid length to satisfy the "perfect brick" configuration favored by charged interactions. Then, we performed a Monte Carlo search for the lowest energy monolayer within a restricted space of ordered monolayers. We allowed four lattice parameters to vary: the monomer spacing c, the offset in x and y between neighboring rows of peptoids, and the spacing x_t between termini in a row of peptoids. In addition, for each of four monomer types (aminoethyl, carboxyethyl, and separate types for phenethyl sites in aminoethyl and carboxyethl blocks), we varied the five degrees of freedom (three translational and two rotational) characterizing the backbone site and the five degrees of freedom characterizing the configuration of the side chain site relative to the backbone. We performed the search for a range of temperatures and used the lowest-energy configuration sampled in all the searches as the starting conditions for our equilibrium simulations.

Once we found a low-energy ordered monolayer, we turned it into a bilayer by replicating two copies and sandwiching them into bilayers offset by translations in three dimensions (Fig. S6 (c)). First, we found a good starting point by rigidly moving the two leaves to find the lowest energy possible. We implemented this initial search by directly enumerating a three-dimensional grid of energies. Then, we performed a Monte Carlo search on the ordered bilayer (Fig. S6 (d)). This search was similar to the ordered monolayer search, except that now we considered eight monomer types instead of four, treating the monomers on the two leaves separately. Fig. S6 (l) shows the lowest energy found after x Monte Carlo steps for the various temperatures. Fig. S6 (m) shows the final lowest energy for each temperature found after 5×10^7 Monte Carlo steps.

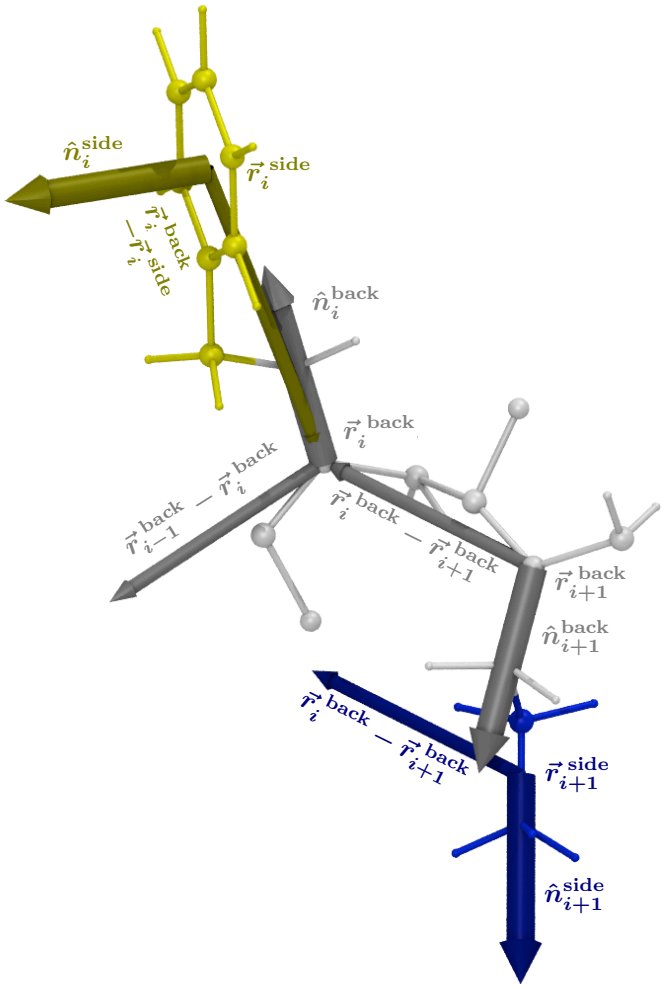

FIG. S7: Schematic showing how symmetry axis vectors and separation vectors are used to generate estimated all-atom configurations (see text). For each coarse-grained site, a principle symmetry axis (thick arrow) and a complementary displacement vector (thin arrow) describe an orientation used to construct positions for atoms associated with the site.

Starting from these initial conditions, we conducted equilibrium simulations of solvated and interfacial peptoids, monolayers, and peptoids. We used system sizes of 48 (2×24) peptoids for the monolayers, 96 ($2 \times 2 \times 24$) peptoids for the bilayers, and 192 (2×96) peptoids for the stacked bilayers.

S6. All-atom configurations and scattering spectra

Keeping track of principle axes of symmetry for each coarse-grained site allows us to combine coarse-grained configurations with locally favored covalent chemistry to generate estimated all-atom configurations. Fig. S7 illustrates our method. For each site, we define a plane from two vectors, the symmetry axis $\hat{n}$ and a complementary displacement vector, which is either a backbone-backbone separation $\vec{r}_{i-1}^{\text{back}} - \vec{r}_i^{\text{back}}$ (for backbone sites or charged side chains sites) or a backbone–side chain separation $\vec{r}_i^{\text{back}} - \vec{r}_i^{\text{side}}$ (for phenethyl side chain sites). Charged and phenethyl side chain sites are treated differently because their symmetry axes are parallel and perpendicular, respectively, to their side chain hydrocarbon chains. Once a plane has been defined for each site, we place atoms associated with each site in positions relative to these planes according to their low-energy configurations in the MFTOID extension [1] of the CHARMM forcefield [2]. Under this procedure, most atoms are placed on their respective planes, while amino and methylene hydrogens are placed out of their planes due to their tetrahedral coordination. In contrast to our assignment of molecular analogs for interaction parametrization, we associate the backbone carbonyl group with its nearest backbone nitrogen site, to its right, rather than the farther site to its left. This allows for a better constrained estimate for the carbonyl atomic positions, especially considering the relative rigid ω bond separating the carbonyl and the nitrogen [20].

Choosing a displacement vector in addition to each site's principle symmetry axis allows us to fully constrain the orientation of each site and generate all-atom configurations. The choice of displacement vector for the backbone sites (the backbone-backbone separation) is motivated by the largely planar and extended configurations found in all-atom simulations of peptoid bilayers [12]. The choice of displacement vector for the side chain sites is largely arbitrary. For charged side chain sites the choice only dictates the angular placement of methylene hydrogens, amino hydrogens, and carboxyl oxygens about the long axis of the site. For phenethyl side chain sites the choice ensures that the gamma carbon of the hydrocarbon chain points back toward the backbone.

Once we generated all-atom configurations, we calculated X-ray scattering spectra at wavevector q via

$$I(\vec{q}) = |\sum_j f_j \exp(i\vec{q} \cdot \vec{r}_j)|^2, \tag{S19}$$

where the sum runs over all atoms for atom j. Eq. S19 assumes that electrons are localized at the atomic sites. We set the atomic scattering factor f_j equal to the atomic number, since this is nearly exact for the experimental 11 keV X-rays [21]. Because of periodic boundary conditions, the values of $\vec{q}$ must be discretized to be commensurate with the periodic box.

We compared to experimental solution X-ray scattering by taking the radial average, $I(q) = \langle I(\vec{q}) \rangle_{|\vec{q}|=q}$. We calculated two-dimensional in-plane scattering spectra by taking a slice at $q_z = 0$, $I(q_x, q_y) = I(\{q_x, q_y, 0\})$. We calculated radially averaged in-plane spectra by taking a radial average, $I(q_{xy}) = \langle I(q_x, q_y) \rangle_{{q_x}^2 + {q_y}^2 = {q_{xy}}^2}$. We calculated transverse spectra for the stack of bilayers by taking a transverse slice, $I(q_z) = I(\{0, 0, q_z\})$. We normalized all experimental and simulated spectra. For solution scattering, we normalized by I_0 in the Guinier fit (see main text). For in-plane spectra, we normal-

ized by the peak of the radially averaged spectra near $q_{xy} = 2\pi/(4.5\text{Å})$. For transverse spectra, we normalized by the peak near $q_z = 2\pi/(5.6\text{Å})$.

S7. Fine-tuning

Using our coarse-grained model parametrized from small-molecule analogs, we found qualitative agreement with experimental [22] and all-atom [12] results: peptoids collapsed into hydrophobic globules in solution and adsorbed to an air-water interface, and monolayers and bilayers were stable with dominant X-ray peaks associated with spacing between parallel peptoids. However, after looking in detail at various pair distribution functions for coarse-grained and all-atom bilayers, as well as experimental, all-atom, and coarse-grained X-ray scattering spectra, we found that we could improve the agreement of the coarse-grained model by fine-tuning a few parameters. Throughout the fine-tuning procedure, we calculated all-atom and coarse-grained in-plane X-ray scattering spectra from free-floating bilayers, rather than stacks of bilayers. We checked for the original and final version of the model that these spectra were very similar to those obtained from stacks of bilayers.

First, as shown by comparing the magenta curves (original model) to the blue curves (all-atom) in Fig. S8 (a), we found that aminoethyl side chain sites stayed closer to the peptoid backbone than in all-atom simulations. In contrast, distribution functions for phenethyl and carboxyethyl side chain sites agreed well with those found in all-atom simulations (Fig. S8 (b-c)). As shown in Fig. S8 (d), the tendency for aminoethyl side chain sites to stay near the backbone was inherited from the monomer from which the side chain bonded potential energy function was parametrized: in all-atom simulations, the phenethyl primary amine preferentially populates the *cis* configuration, while the phenthyl monomer in a bilayer preferentially populates the *trans* configuration. To correct the discrepancy between coarse-grained and all-atom simulations in the bilayer, we found that it was convenient to add -7 kcal/mol to the term of the side chain interaction (Eq. S3) linear in $\arctan(r_\parallel/r_\perp)$. After shifting the potential to keep its minimum at 0, this resulted in adjusting the parameters J_{10} and J_{11} in Eq. S3 and Table SII from 418.586 and -1837.99 to 424.643 and -1844.99.

Second, as shown by comparing the magenta curves (original model) to the black (experiment) and blue (all-atom) curves in Fig. S8 (e), we found that the X-ray scattering peak corresponding to the spacing between peptoids was too weak and occurred at too large a wavenumber (too small a distance). We expected that increasing the sharpness of the X-ray peak would require increasing some or all of the nonbonded interaction strengths. We tried all combinations of models created by modulating the bonded, phenyl-phenyl, and amino-carboxy interaction strengths by the factors $\{0.8, 1, 1.2\}$, $\{1, 1.2\}$, and $\{1, 1.2, 1.5\}$, respectively. We found that varying the bonded and amino-carboxy interaction strengths had little effect on the X-ray peak, but varying the phenyl-phenyl interaction strength by a factor of 1.5 was sufficient to yield an X-ray peak with the same shape as the experimental and all-atom peaks, albeit still shifted to larger wave numbers. In Fig. S8, the brown curves ("version 2") denotes the version of the model with the phenyl-phenyl interaction increased by 50% and the amino side chain interaction biased toward *trans*.

After these two corrections, we still needed to correct the location of the polymer-spacing X-ray peak. After comparing several distribution functions between coarse-grained and all-atom simulations, including the selections shown in Fig. S8 (f-h), we suspected that the discrepancy was arising from a combination of two factors: the backbone sites and/or the phenethyl side chain sites were too small. For instance, Fig. S8 (f) shows that while backbone sites in the coarse-grained model were approaching as close as their hard-core diameter, 3.3 Å, backbone nitrogens in the all-atom simulations did not approach closer than 4 Å. Not surprisingly, using nitrogen's van der Waals radius for the backbone site was insufficient to account for the steric repulsion between entire coarse-grained sites. Similarly, Fig. S8 (f) shows that while phenethyl side chain sites in the coarse-grained model were approaching as close as the width of the repulsive part of the Gay-Berne potential (Eq. S6), $2\chi\kappa\sigma_0 = 3.024$ Å, the centers of the aromatic rings in the all-atom simulations did not approach closer than 4 Å.

In contrast, the comparison of relative orientations between the coarse-grained and all-atom bilayers was relatively good. The first column of fig. S8 (g) shows the radial distribution function for phenethyl side chain sites on the same polymer (first row), the same leaf but different polymers (second row), and different leaves (third row). Selecting neighbors within approximately the first shell of these distribution functions, we plot the orientation-dependent scalar coordinates that enter the Gay-Berne potential in the next three columns. The trends in all three of these coordinates match those for the all-atom simulations, although the all-atom simulations show some fine detail that is missing in the coarser model. Similarly, Fig. S8 (h) shows that the relative orientations of aminoethyl and carboxyethyl sites in contact (top plot) and separated by (implicit) solvent (bottom plot) agreed reasonably well after we biased the aminoethyl side chain to *trans*.

To simulataneously improve the agreement with X-ray scattering and distribution functions, we therefore adjusted the backbone and phenethyl side chain sizes while keeping all orientational geometry and interaction energies fixed. We varied the backbone radius r_{back} and the phenethyl side chain size σ_0 on a grid, $r_{\text{back}} = 1.65$ Å, 1.75 Å, ..., 2.35 Å and σ_0=5.6 Å, 5.8 Å, ..., 7.0 Å. We found that $r_{\text{back}} = 2.05$ Å and $\sigma_0 = 6.2$ Å yielded the best results, shown as red curves in Fig. S8. We kept these values for the version of the model we used to produce all of the results in the main text.

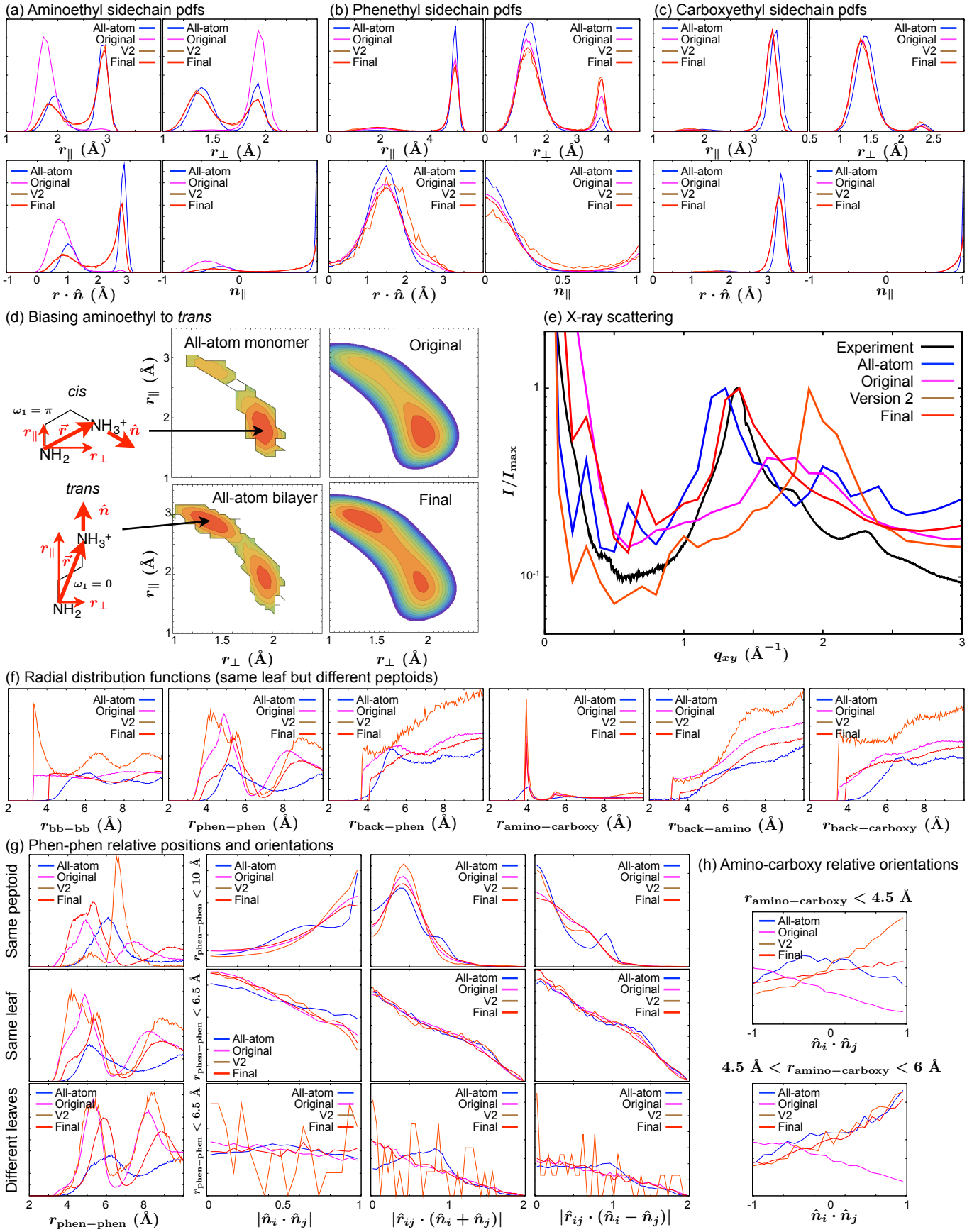

FIG. S8: Selection of distribution functions and X-ray scattering spectra used to fine-tune the coarse-grained model. Experimental (black) and all-atom (blue) results are compared to the original version (magenta), a version (v2) with an increased phenethyl side chain nonbonded interaction strength and aminoethyl side chain interactions biased towards the *trans* configuration (brown), and the final version with the size of backbone and phenethyl sites increased (red). See text for discussion of individual panels.

[1] D. T. Mirijanian, R. V. Mannige, R. N. Zuckermann, and S. Whitelam, J. Comput. Chem. **35**, 360 (2014).
[2] A. D. Mackerell, M. Feig, and C. L. Brooks, J. Comput. Chem. **25**, 1400 (2004).
[3] R. Wolfenden, Biochemistry **17**, 201 (1978).
[4] A. Radzicka and R. Wolfenden, Biochemistry **27**, 1664 (1988).
[5] Y. K. Kang, G. Némethy, and H. A. Scheraga, J. Phys. Chem. **91**, 4118 (1987).
[6] A. K. Shaytan, V. A. Ivanov, K. V. Shaitan, and A. R. Khokhlov, J. Comput. Chem. **31**, 204 (2010).
[7] B. Minofar, R. Vácha, A. Wahab, S. Mahiuddin, W. Kunz, and P. Jungwirth, J. Phys. Chem. B **110**, 15939 (2006).
[8] P. Jungwirth, private communication.
[9] P. Kebarle, Jerusalem Symp. Quantum Chem. Biochem. (1975) **8**, 81 (1976).
[10] R. Wolfenden, L. Andersson, P. M. Cullis, and C. C. B. Southgate, Biochemistry **20**, 849 (1981).
[11] S. Gopalakrishnan, P. Junwirth, D. J. Tobias, and H. C. Allen, J. Phys. Chem. B **109**, 8861 (2005).
[12] R. V. Mannige, T. K. Haxton, C. Proulx, G. L. Butterfoss, R. N. Zuckermann, and S. Whitelam, submitted.
[13] J. G. Gay and B. J. Berne, J. Chem. Phys. **74**, 3316 (1981).
[14] S. Gupta, W. B. Sediawan, and E. McLaughlin, Mol. Phys. **65**, 961 (1988).
[15] T. R. Walsh, Mol. Phys. **100**, 2867 (2002).
[16] I. Cacelli, G. Cinacchi, G. Prampolini, and A. Tani, J. Chem. Phys. **120**, 3648 (2004).
[17] C. Chipot, R. Jaffe, B. Maigret, D. A. Pearlman, and P. A. Kollman, J. Am. Chem. Soc. **118**, 11217 (1996).
[18] S. Zhu and A. H. Elcock, J. Chem. Theory Comput. **6**, 1293 (2010).
[19] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, New York, 1987).
[20] F. P. Gasparro and N. H. Kolodny, J. Chem. Ed. **54**, 258 (1977).
[21] C. T. Chantler, J. Phys. Chem. Ref. Data **24**, 71 (1995).
[22] B. Sanii, T. K. Haxton, G. K. Olivier, B. Barton, C. Proulx, S. Whitelam, and R. N. Zuckermann, ACS Nano **8**, 11674 (2014).