Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supporting Information: Shearing Friction Behaviour of Synthetic Polymers Compared to a Functionalized Polysaccharide on Biomimetic Surfaces: Models for the Prediction of Performance of Eco-designed Formulations

by Benjamin J Coscia[1*], John C Shelley[1], Andrea R Browning[1], Jeffrey M. Sanders[2], Robin Chaudret[3], Roger Rozot[4], Fabien Léonforte[4], Mathew D. Halls[5] and Gustavo S Luengo[4*]

1. Schrödinger, Inc., Portland, OR 97204, United States
2. Schrödinger, Inc., New York, NY 10036, United States
3. Schrödinger, Inc., 80538 München, Germany
4. L'Oréal Research and Innovation, Aulnay-Sous Bois, France
5. Schrödinger, Inc., San Diego, California 92121, United States

* corresponding authors

S1. Reference All-Atom Simulations

The simulations presented in the main text are Coarse-grained models based on reference all-atom simulations. This section describes the construction and simulation of the all-atom models of both bulk formulations and formulations in contact with model hair surfaces. The information in this section is essential for fully reproducing the study in the main text.

S1.1 Bulk Formulations

We studied five atomistic models of formulations in bulk solution whose compositions are detailed in **Table S1**. For M100 and M2003, we studied two concentrations: a low concentration representative of those experienced while using the shampoo formulation, and a high concentration (5 times higher than the low concentration system) intended to increase sampling of interactions for use in parameterizing our CG models. For PS, we only simulated one concentration about 10 times higher than expected during shampoo usage. The size of the PS monomer precluded us from studying lower concentration systems without using even shorter chains or considerably larger atomistic systems. However, we believe the single high concentration is sufficient to parameterize our CG models.

As shown in **Table S1**, we chose to split the polymers into multiple chains of approximately equal length, rather than simulate single chains. We reason that using multiple chains facilitates more sampling of a greater diversity of intermolecular interactions in solution.

We built initial configurations using the *Disordered System Builder* and then equilibrated them according to the following procedure:

1. 100 ps Brownian dynamics in the NVT ensemble at 10K with a 1 fs time step
2. 24 ps molecular dynamics in the NVT ensemble at 300K with a 1 fs time step
3. 240 ps molecular dynamics in the NVT ensemble at 700K with a 1 fs time step
4. 24 ps molecular dynamics in the NPT ensemble at 300K, 1 atm and a 1 fs time step
5. 200 ps molecular dynamics in the NPT ensemble at 300K, 1 atm and a 2 fs time step
6. 10 ns molecular dynamics in the NPT ensemble at 300K, 1000 atm and a 2 fs time step
7. 10 ns molecular dynamics in the NPT ensemble at 300K, 1 atm and a 2 fs time step

Polymer Type	**Target wt%**			**# monomers**	**# chains**	**# chains of exp. MW**	**Number of Components**				**Total Atoms**
	Polymer	SLES + CAPB	NaCl				CAPB	SLES	NaCl	Water	
M100	0.50%	1%	0.7%	77	2	0.12	24	44	297	134312	409532
M100	2.50%	5%	0.7%	408	4	0.66	128	234	316	134312	435458
M2003PR	0.25%	1%	0.7%	48	2	0.017	24	44	297	134312	408480
M2003PR	1.25%	5%	0.7%	248	4	0.088	126	235	316	134312	429713
PS	2.00%	10%	0.3%	29	2	0.023	269	489	142	134312	456970

Table S1. *Compositions of bulk formulations studied via all-atom MD simulations.*

Following equilibration, we ran a 200 ns production simulation in the NPT ensemble at 300K and 1 atm, with a 2 fs timestep. In order to determine equilibration, we used the *Cluster Analysis* tool, as described in Section 2.1 of the main text, in order to identify aggregates. We considered these systems to be equilibrated once the total number of clusters versus time reached a plateau. Based on **Figure S1**, we chose the last 50 ns of the simulation time as the equilibrated region.

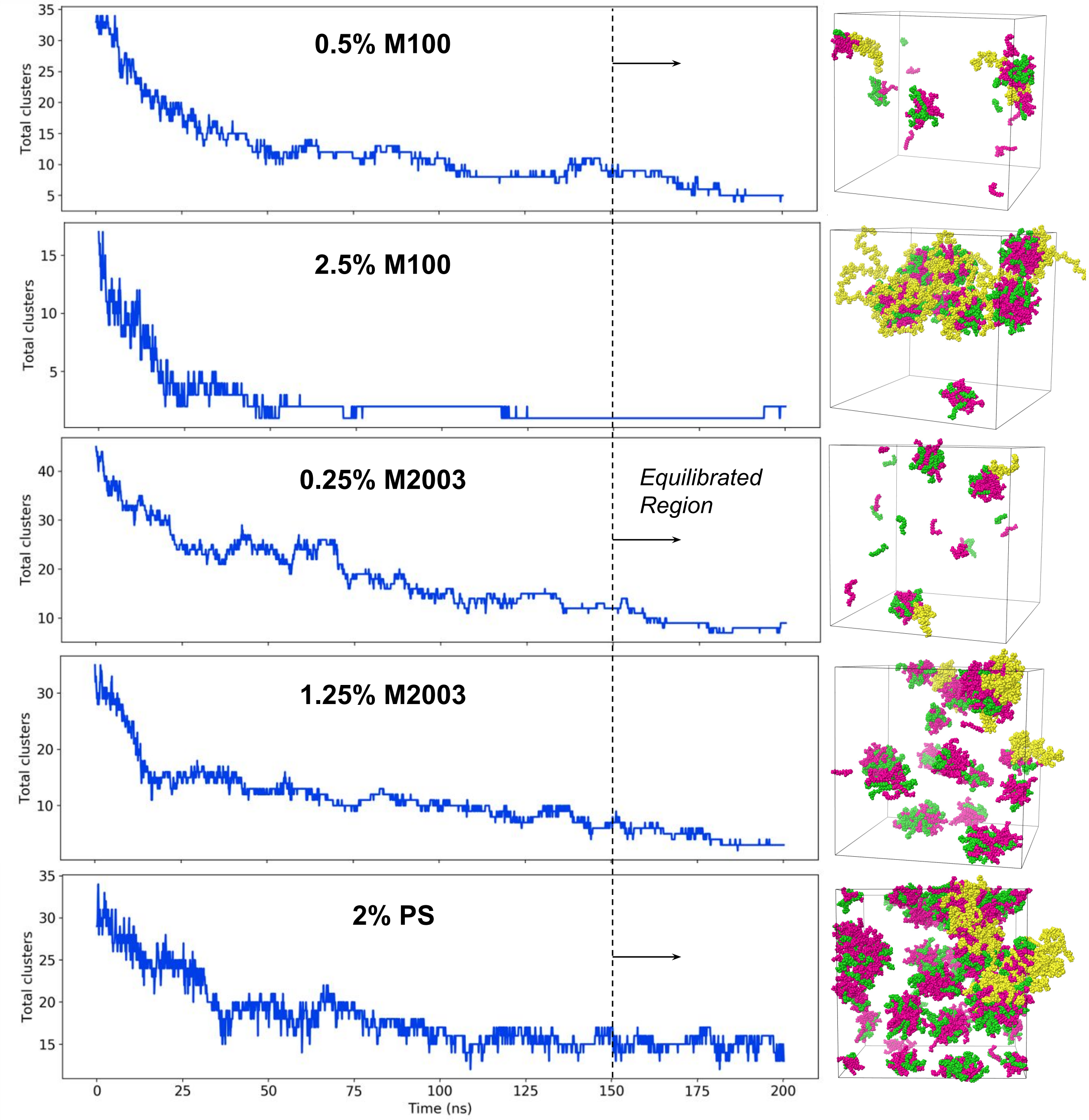

Figure S1 Number of clusters versus time for each AA bulk formulation studied. Snapshots of the final frame are placed next to each time series. We determined system equilibration based on when the number of clusters versus time plateaued. For simplicity, we define the last 50 ns of the production simulation as the equilibrated region.

S1.2 Formulations in contact with model hair surface

For each of the three polymers, we studied atomistic models of formulations in contact with two different hair conditions: healthy and extremely damaged. The compositions of the solution components of each model are detailed in **Table S2**. We chose the number of water molecules needed to reach a specific distance between hair substrate surfaces. Our approach to determining this number is explained in **Section S3**.

Our AA hair models are heavily inspired by the work of Cheong et al.[25] As shown in **Figure S2**, we use the *Nanostructure Builder* to create an inert substrate composed of five layers of graphene stacked 0.335 Å apart, thick enough to prevent significant non-covalent interactions of species across the barrier. The surfaces were equal in their x and y dimensions with side lengths of 18.4 nm, 13.9 nm and 13.9 nm for M100, M2003 and PS respectively. We covalently attach surface components to carbon atoms on the two outer layers of the graphene substrate with attachment points chosen on an equispaced hexagonal grid. The structure of graphene allows equispaced grafting of surface components 0.425 nm or 0.491 nm apart which corresponds to surface densities of 6.42 and 4.79 attachments/nm^2 respectively. For healthy hair, we used the 0.491 Å grid spacing in order to target a density of 4.26 18-MEA molecules/nm^2. To achieve the exact density, we randomly removed 18-MEA molecules. Note that in our coarse-grained model discussed in the main text, we did not include this random removal step. For damaged hair, we used the 0.425 Å grid spacing and targeted a density of sulfonate groups of 2.2 sulfonates/nm^2. We filled the remaining sites with tethered ethanol molecules.

When we attach groups to the graphene sheets, we must change the sp^2 hybridization of the associated surface carbons to sp^3. Consequently, the bonded parameters at the attachment sites are updated which allows the configuration to deviate from the original flat sheet. This behavior is incompatible with the middle three graphene sheets. We mitigate these effects by transforming the hybridization of an equal number of graphene carbon atoms in the three middle sheets.

Polymer Type	**Hair Type**	**# Polymer Monomers**	**# Polymer Chains**	**Chains of exp. MW**	**Number of Components**				**Total Atoms**
					CAPB	SLES	NaCl	Water	
M100	Healthy	818	4	1.32	127	233	373	45301	492084
	Damaged	818	4	1.32	127	233	373	45301	304991
M2003PR	Healthy	2018	4	0.718	178	324	591	34858	352583
	Damaged	2018	4	0.718	178	324	591	73838	363018
PS	Healthy	120	4	0.094	151	276	253	36805	351156
	Damaged	120	4	0.094	151	276	253	70066	344158

Table S2. *Compositions of formulations in contact with model hair studied via all-atom MD simulations*

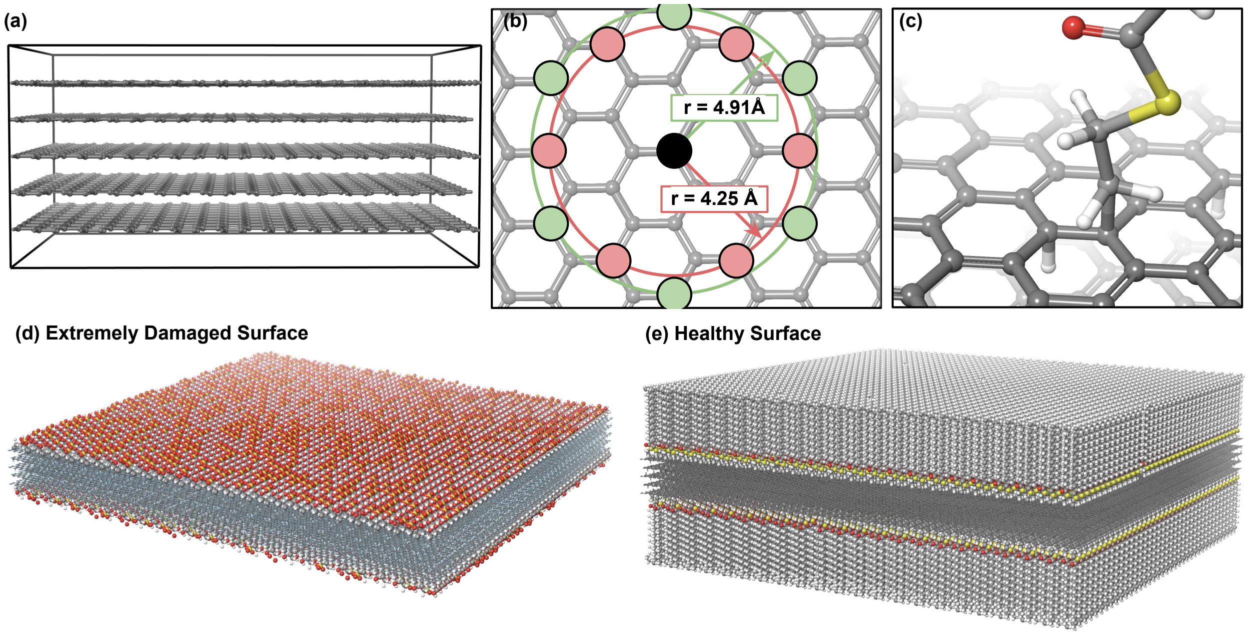

Figure S2 Model hair surface build procedure. (a) We stack five graphene sheets 3.35 Å apart then (b) locate an equispaced hexagonal grid of carbon atoms to be used as surface attachment points. (c) We attach surface components at the grid points and modify the hybridization of the graphene carbon from sp^2 to sp^3. This also requires that we modify the hybridization of an adjacent carbon atom, to which we also add a hydrogen atom. (d) The extremely damaged surface incorporates sulfonate and ethanol moieties tethered on a grid of graphene carbon atoms spaced 4.25 Å apart. (e) The healthy hair surface incorporates 18-MEA molecules placed on a grid spaced 4.91 Å apart.

We built initial configurations on top of the hair surfaces using the *Disordered System Builder* and then equilibrated them according to the following procedure:

1. 100 ps Brownian dynamics in the NVT ensemble at 10K with a 1 fs time step
2. 1 ns molecular dynamics in the NPT-anisotropic ensemble at 100K, 1 atm with a 1 fs time step
3. 1 ns molecular dynamics in the NPT-anisotropic ensemble at 300K, 1 atm with a 1 fs time step
4. Modify configuration to have box vectors that are averaged over the last 20% of frames from 3.
5. 1 ns molecular dynamics in the NPAT ensemble at 300K, 1 atm and a 2 fs time step

NPT-anisotropic refers to pressure coupling that allows the orthorhombic box vectors to change independently in each dimension. NPAT refers to a pressure coupling scheme that holds the *x* and *y* box vectors constant, but allows the *z*-dimension to change.

Following equilibration, we ran a 200 ns production simulation in the NPAT ensemble at 300K and 1 atm, with a 2 fs timestep. As with the bulk systems, we used the *Cluster Analysis* tool in order to identify aggregates and determine equilibration. The hair surface is excluded from the clustering calculations. Additionally, we exclude the polymers from this set of calculations. Since these systems are more concentrated than the bulk systems, polymers fill much of the space and result in identification of one or two large clusters that persist throughout the whole simulation. Limiting the calculation to surfactants allows us to resolve equilibration of aggregates within the polymer. Based on the plots in **Figure S3**, we chose the last 100 ns of the simulation time as the equilibrated region.

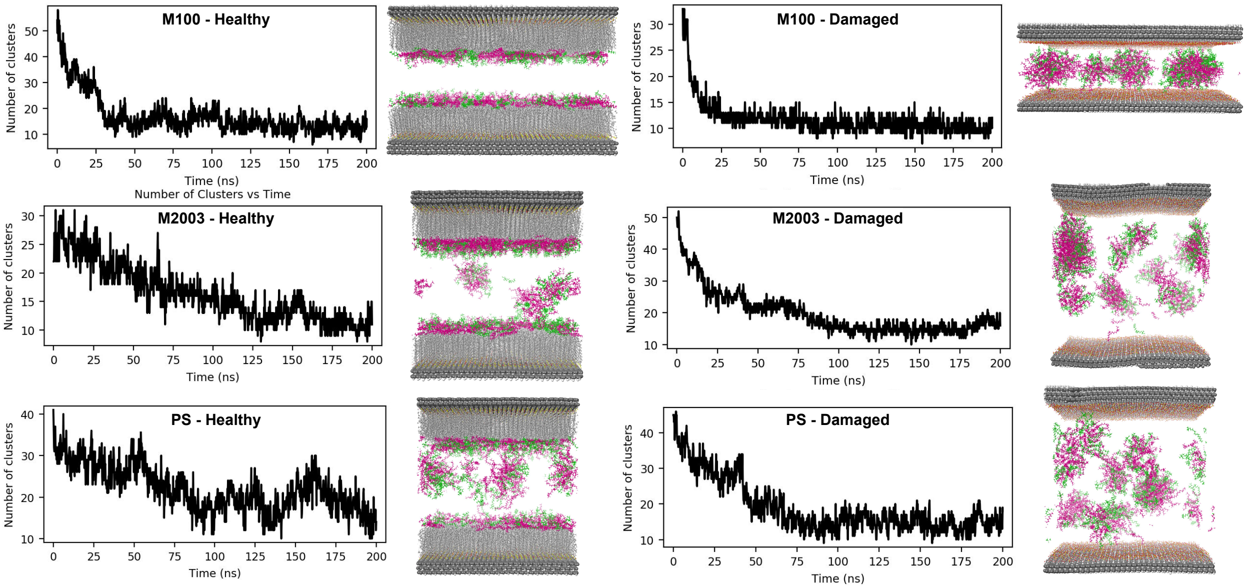

Figure S3 Number of clusters versus time during production simulations of AA formulations in contact with model hair surfaces. Clustering calculations only consider the positions of surfactant molecules. Note that, in the healthy systems, distinct clusters are formed by surfactants adsorbed to the surface in disjoint patches. Snapshots of the final frame from the trajectories are shown to the right of each plot. Surfactants are colored green (CAPB) and pink (SLES). Water molecules and polymers are hidden for ease of visualization. All other components belong to the model hair surfaces. All systems are reasonably equilibrated after 100 ns of simulation time.

S2. Coarse-Grained Parameterization

We developed a Martini-style set of coarse-grained parameters by comparison of coarse-grained test systems to reference all-atom simulations discussed in **Section S1**. We begin by mapping the atomistic trajectories to coarse-grained representations based on the bead definitions shown in **Figure S4**. From the all-atom reference trajectories, we extract reference bond length and angle distributions, as well as all pairwise radial distribution functions (RDFs).

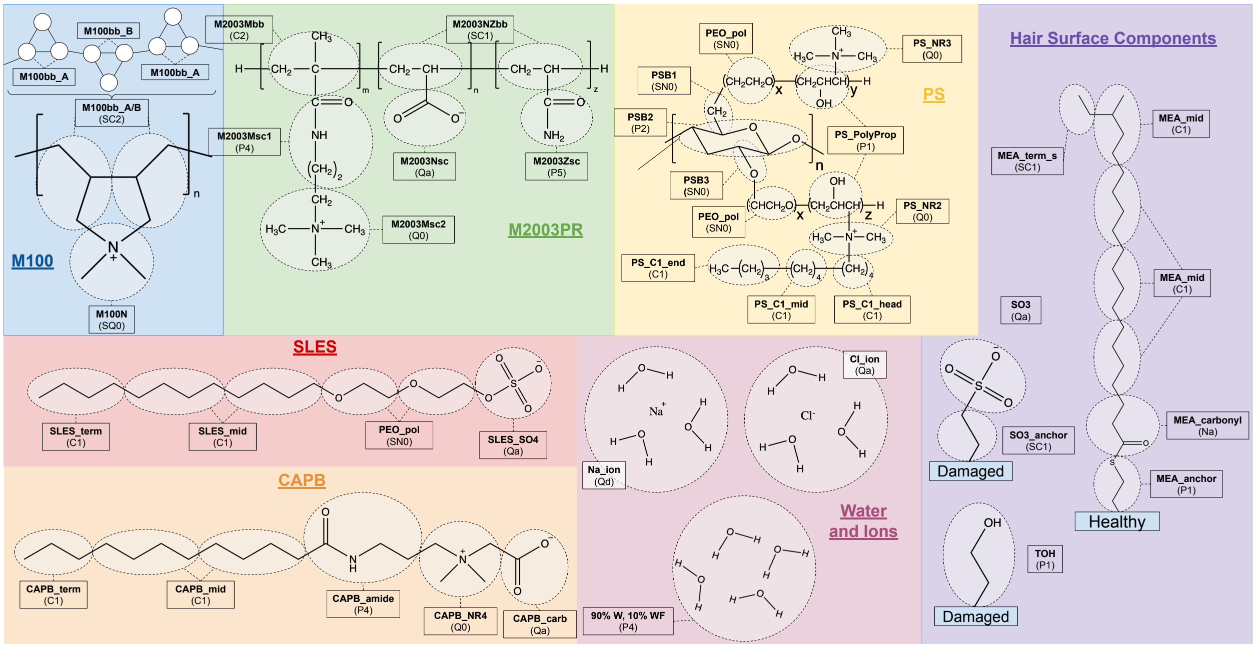

Figure S4 Coarse-grained bead mapping definitions. We mapped each group of encircled atoms to a single coarse-grained site with initial martini site-types listed in parentheses below the bolded bead names.

We optimize our CG parameters against the AA reference distributions iteratively. We initialize a CG model with the exact same composition as the all-atom system using the Martini site types listed in **Figure S4**. We build ingredients using the *Coarse Grained Sketcher*, and polymers using *Polymer Builder*. We build the initial configuration in a disordered state by passing the ingredients to the *Disordered System Builder*. We apply default Martini parameters using the *Coarse Grained Force Field Assignment* tool. Next, we run an equilibration procedure followed by a 200 ns production simulation according to the following steps:

1. Brownian dynamics 0.1 ns, NVT, 10 K, 1 fs time step
2. Molecular dynamics 0.1 ns, NVT, 10 K, 1 fs time step
3. Molecular dynamics 0.1 ns, NVT, 300K, 1 fs time step
4. Molecular dynamics 0.1 ns, NPT, 1 atm, 10 K, 1 fs time step
5. Molecular dynamics 2 ns, NPT, 10 bar, 300 K, 2 fs time step
6. Molecular dynamics 0.5 ns, NPT, 1 atm, 300 K, 10 fs time step
7. Molecular dynamics 200 ns, NPT, 1 atm, 300 K, 15 fs time step

From the output of these simulations, we collect bond length and angle distributions as well as RDFs and then compare them directly to the AA reference simulations. Based on these comparisons, we update the parameters of the bonded and nonbonded parameters given by the potential used for each type of interaction:

$$V_{bond} = (k_{bond}/2)(r - r_{eq})^2 \tag{S1}$$

$$V_{angle} = (k_{angle}/2)(\theta - \theta_{eq})^2 \tag{S2}$$

$$V_{nonbonded}(r) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right] \tag{S3}$$

We repeat the above process iteratively until we reach satisfactory agreement with AA simulations. Anecdotally, we observed relatively fast convergence of the bonded parameters while the nonbonded parameters generally required many more iterations in order to appropriately balance competing interactions.

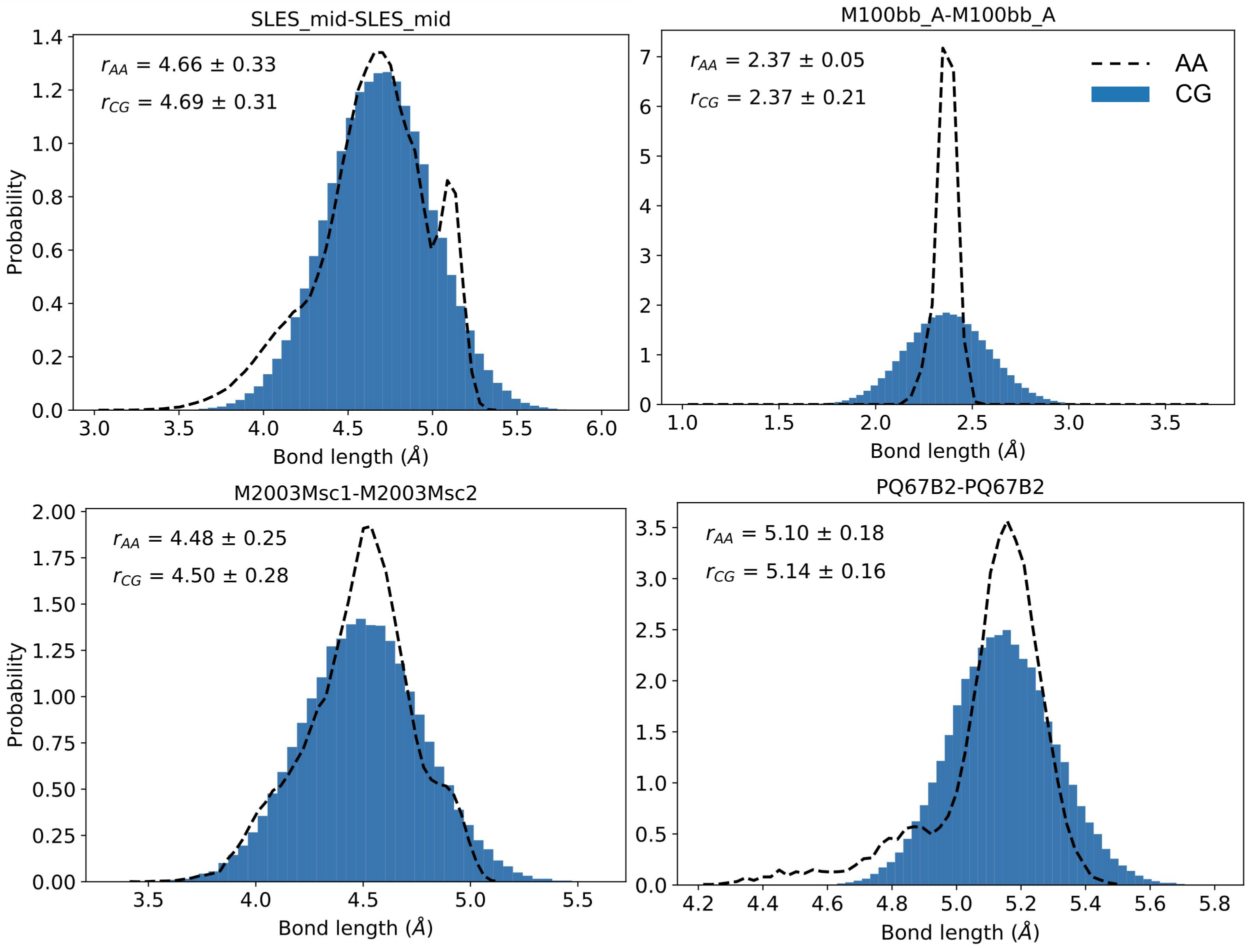

Figure S5 Selected bond length distributions from the reference AA and final fit CG potential. See Figure S4 for bead definitions. In many cases, the width of the AA distributions were difficult to reproduce without excessively high force constants (k_{bond}) which leads to system instability. Therefore we conceded to lower force constant and consequently broader bond length distributions.

Here we feel it is appropriate to emphasize that we determined convergence while weighing what is reasonable for manual optimization of hundreds of RDFs simultaneously. Clearly, we would benefit from an automated parameterization procedure. Approaches such as iterative Boltzmann weighting are attractive towards this end, but are typically used with numerical potentials rather than the two-parameter shifted Lennard Jones model implemented by Martini. Future work will focus on automating our procedure for faster and more accurate parameter sets.

A natural consequence of eliminating degrees of freedom via coarse-graining is the loss of some details in the CG bonded and nonbonded parameter distributions relative to AA. **Figures S5-S7** demonstrate this point and illustrate the types of choices we made in order to capture the relevant structural features. For bonded parameters we attempt to maximize overlap between the distributions. For nonbonded parameters, we focused on matching the integrals of the RDFs because the integrals quantify the number of neighboring beads. For ion and water beads, the AA RDFs are based on individual ions and water molecules, causing the peaks to shift to lower values of r.

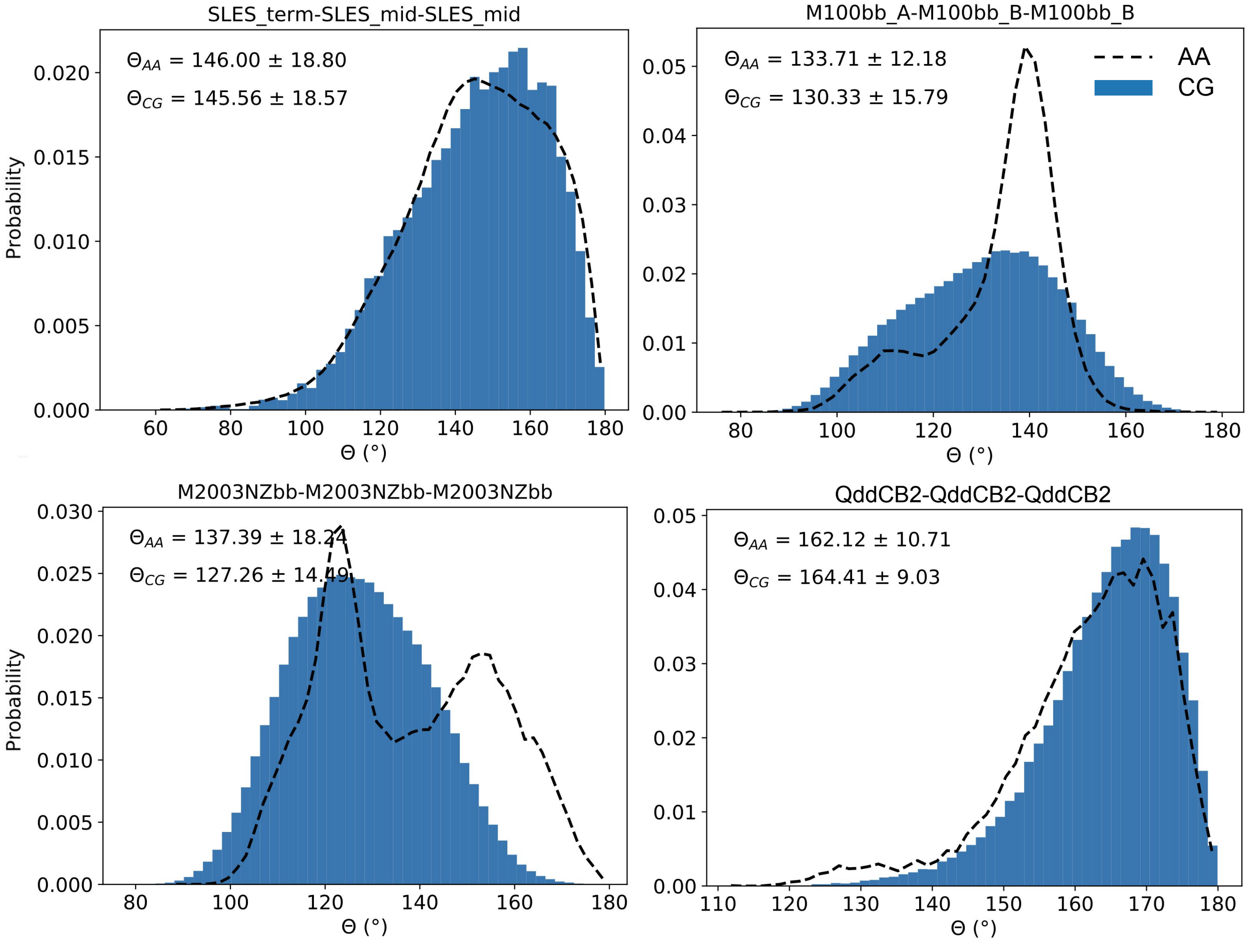

Figure S6 Selected angle distributions from the reference AA and final fit CG potential. See **Figure S4** for bead definitions. Surfactant tail parameters (e.g. top left) were generally parameterized with Θ=180 and k_{angle} chosen to reproduce the width of the AA distribution. The backbone of PS (bottom left) was subject to a similar approach. The multimodal angle distributions exhibited by the M100 and anionic M2003 backbones (top right and bottom left respectively) could only be approximated by CG potentials.

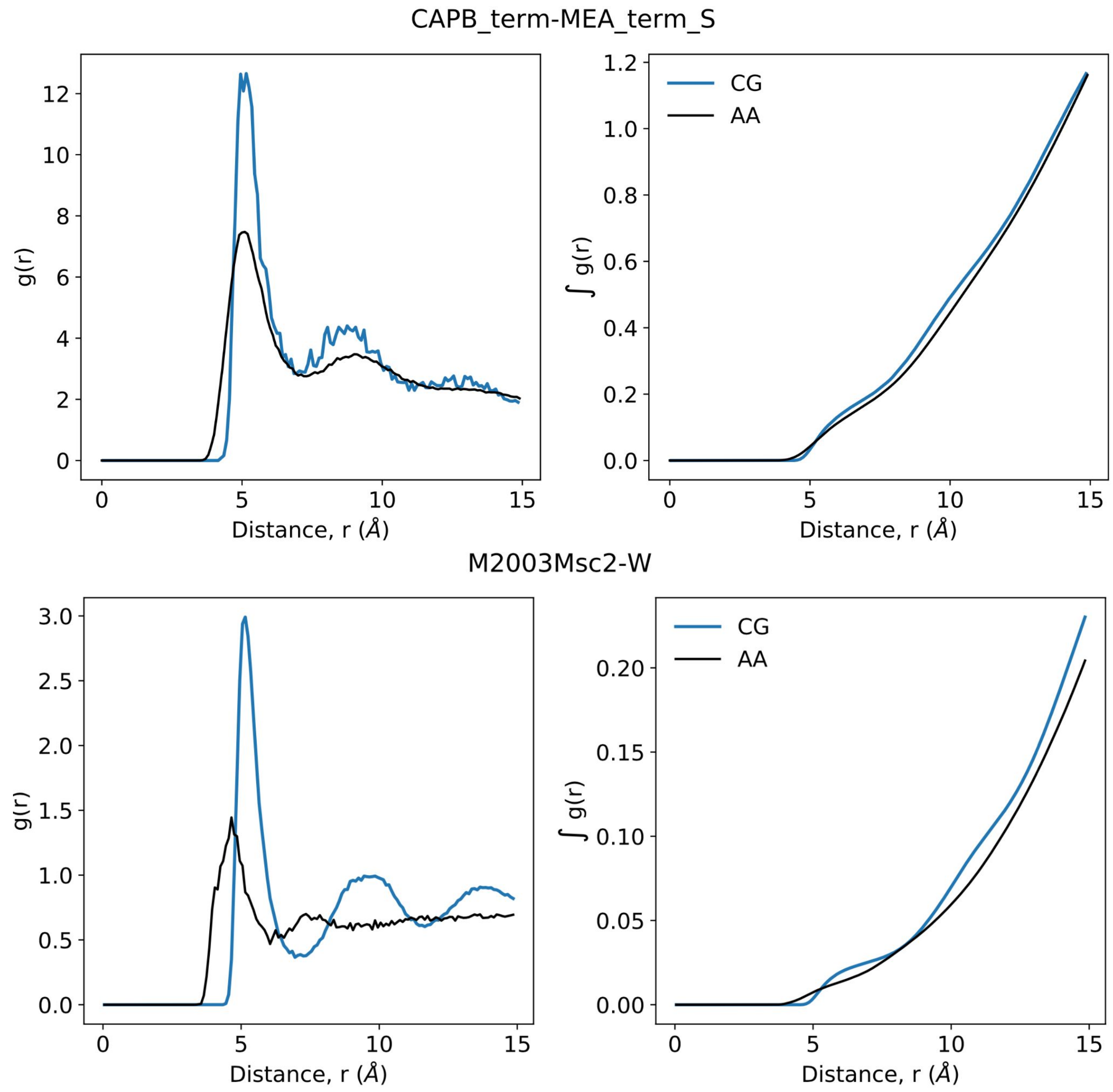

Figure S7 Selected radial distribution functions (g(r)) from the reference AA and final fit CG potential.. See **Figure S4** for bead definitions. We focused our efforts on matching the integrals of the RDFs for r < 15 Å. Since Martini represents 4 water molecules with one 'W' bead, the peak in the CG RDF is expected to be offset from the peak in the AA distribution, as shown in the bottom plots. However, the integral communicates that the total number of water molecules in the vicinity of the M2003Msc2 beads is about right.

In addition to the quantitative validation of parameters, we found it useful to visualize the trajectories associated with each iteration. **Figure S8** illustrates the refinement in aggregation behavior that we observed while optimizing the nonbonded parameters for M100. These types of visualization helped us to understand which parameters required tuning.

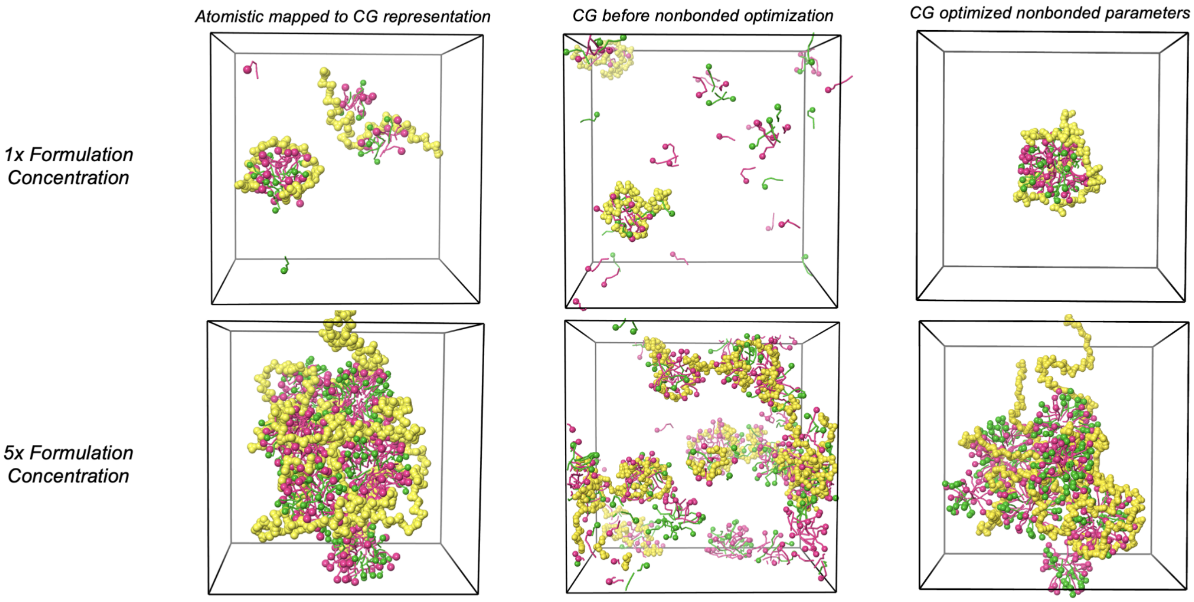

Figure S8 There is a clear qualitative mismatch between the mapped AA model of the M100 systems and our initial CG models, prior to any optimization. After iterative optimization of the nonbonded parameters, we see CG aggregation behavior that is consistent with AA.

We developed our nonbonded Lennard-Jones parameters in a way that emphasizes transferability in order to facilitate expansion of the parameter set to include new molecules and different conditions. Therefore, we were careful to parameterize a single set of parameters for components common to all systems. These include CAPB, SLES, water, ions and the hair surfaces.

In bulk systems, we optimized CAPB, SLES, water and ion nonbonded parameters simultaneously. We did not change interactions between water molecules (i.e. (W, W), (W, WF) and (WF, WF)) from the base set of Martini parameters, however we did modify the strength of interactions between water and all other beads.

To optimize the nonbonded parameters for systems in contact with the model hair surface, we held all parameters from the bulk system optimization constant, and only modified pairwise nonbonded interactions between the surface and solution components.

S3. Equilibration of Coarse-Grained Systems

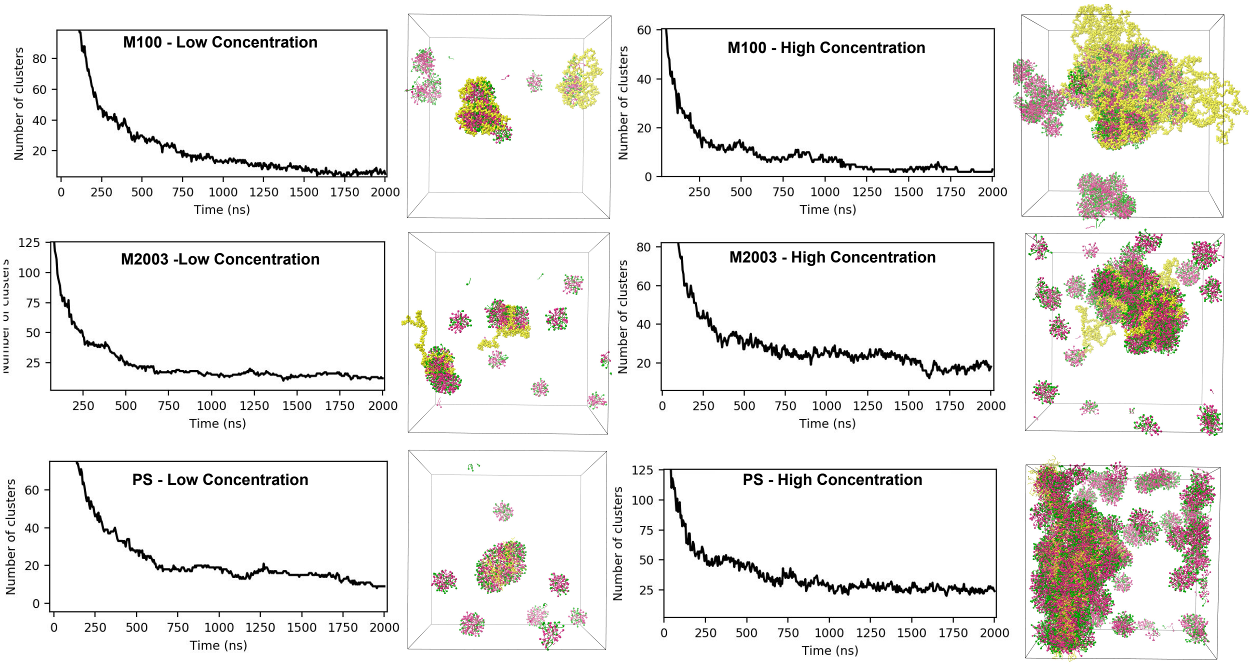

Figure S9 Number of clusters versus time during production simulations of Coarse-grained simulations of bulk formulations.

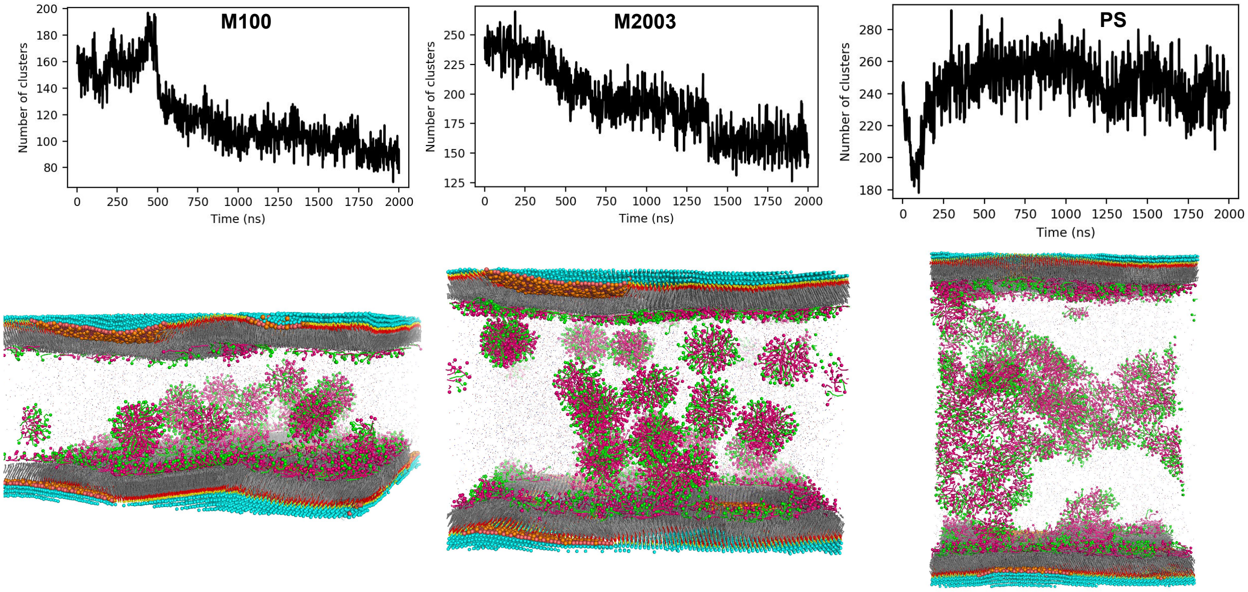

Figure S10 Number of clusters versus time during production simulations of partially damaged hair systems.

S4. Confined System Water Content Determination

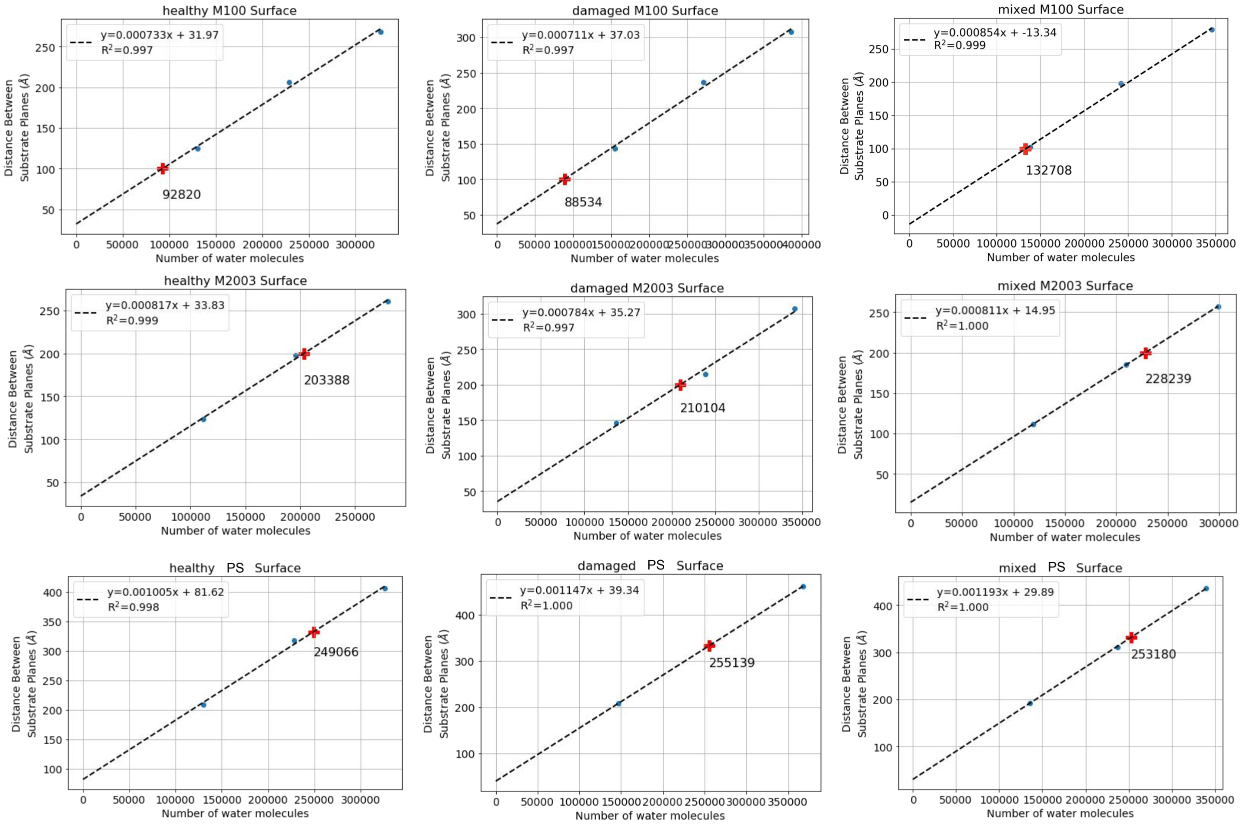

Figure S11 We estimated the number of water molecules needed to achieve a specific z separation distance between solvent-adjacent faces of the model hair substrates by linearly interpolating the resulting separation of 3 systems with different water contents.

Procedure:

1. Filled initial cells with water beads up to 500,000 total atoms
2. Create 2 additional configurations by reducing total number of water molecules to 70% and 40% of original number, via random deletion of water molecules
3. Equilibrate all configurations (27 total)
4. Measure interplanar distance after equilibration
5. Interpolate precise number of water molecules needed from a linear fit

S5. Surface Configuration for Shearing

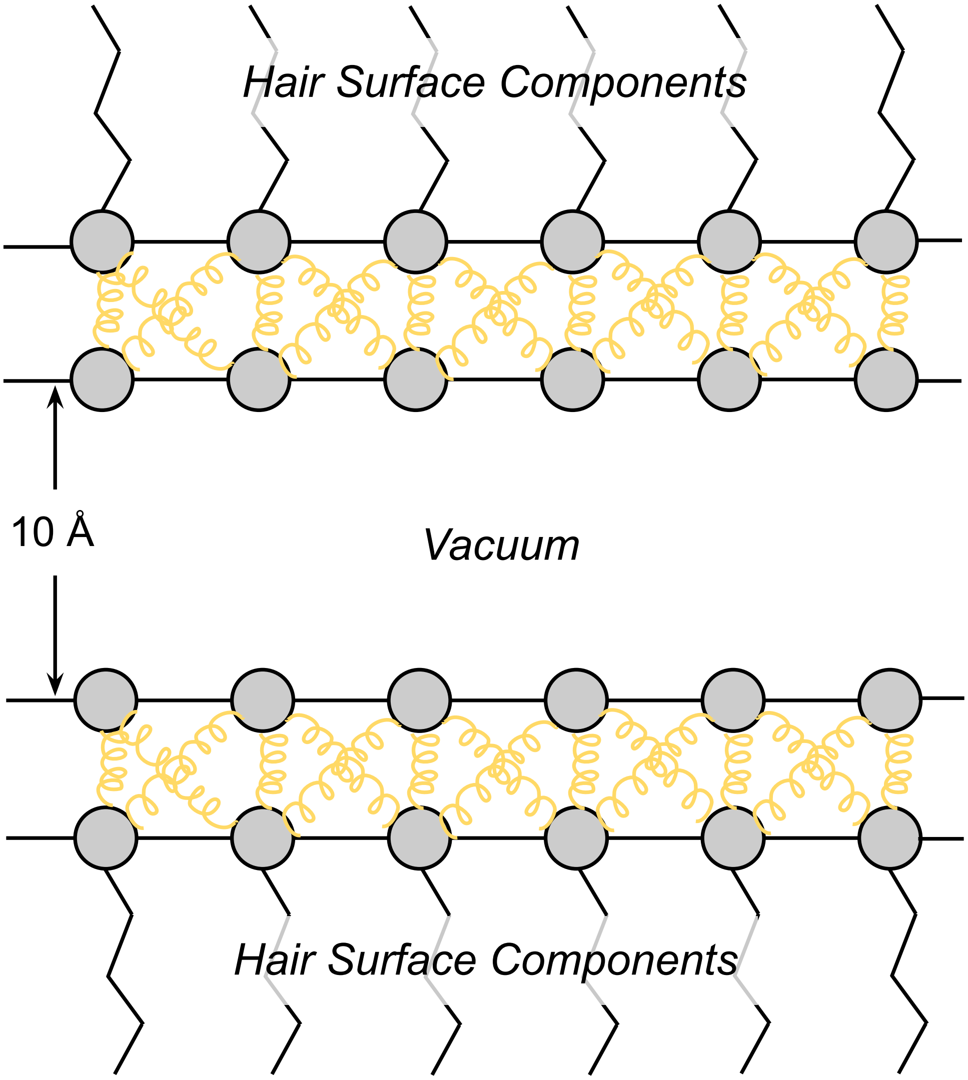

Figure S12 In order to measure drag force without frictional contributions from the four graphene-like substrate layers used to anchor the surface components, we added a vacuum gap between the two middle layers and applied harmonic restraints (yellow spring-like drawings) between each set of outer layers between each site and its nearest neighbors. The restraints add mechanical strength which prevents significant bending of the surface.

S6. Density-Order Parameter Plots

The plots in this section are analogous to **Figure 6** of **Section 3.2** of the main text and may be useful in supporting the statement made about M100 and PS.

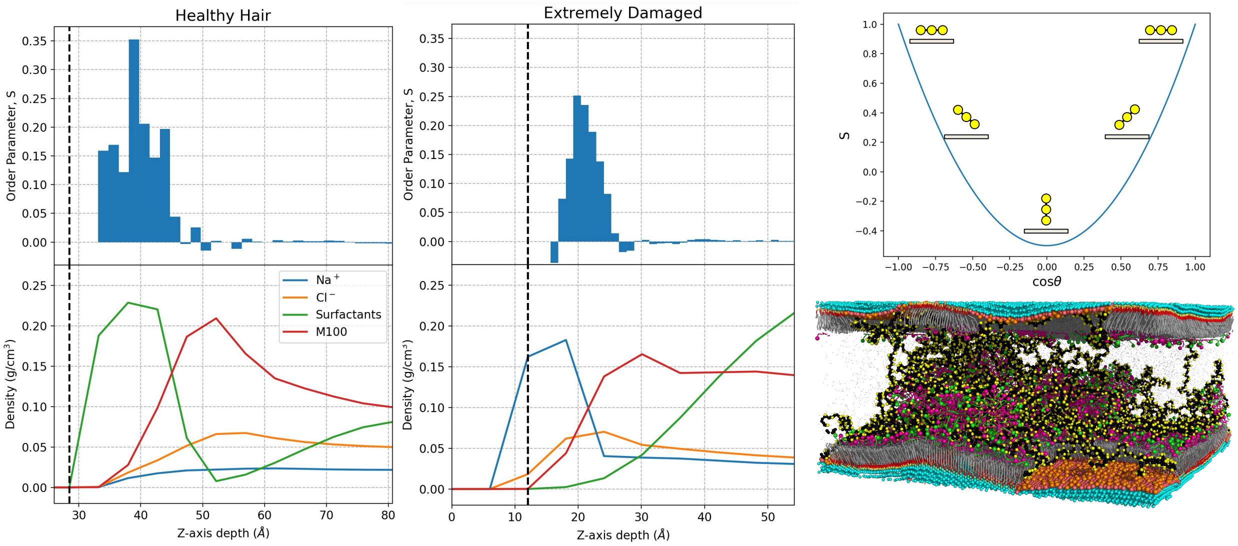

Figure S13 Density profiles aligned with *z*-dependent polymer order parameters for healthy and extremely damaged model hair surfaces in contact with M100 formulations. The bottom right image is of a partially damaged system which has structural features consistent with a mixture of the healthy and extremely damaged surfaces. See the main text for further discussion.

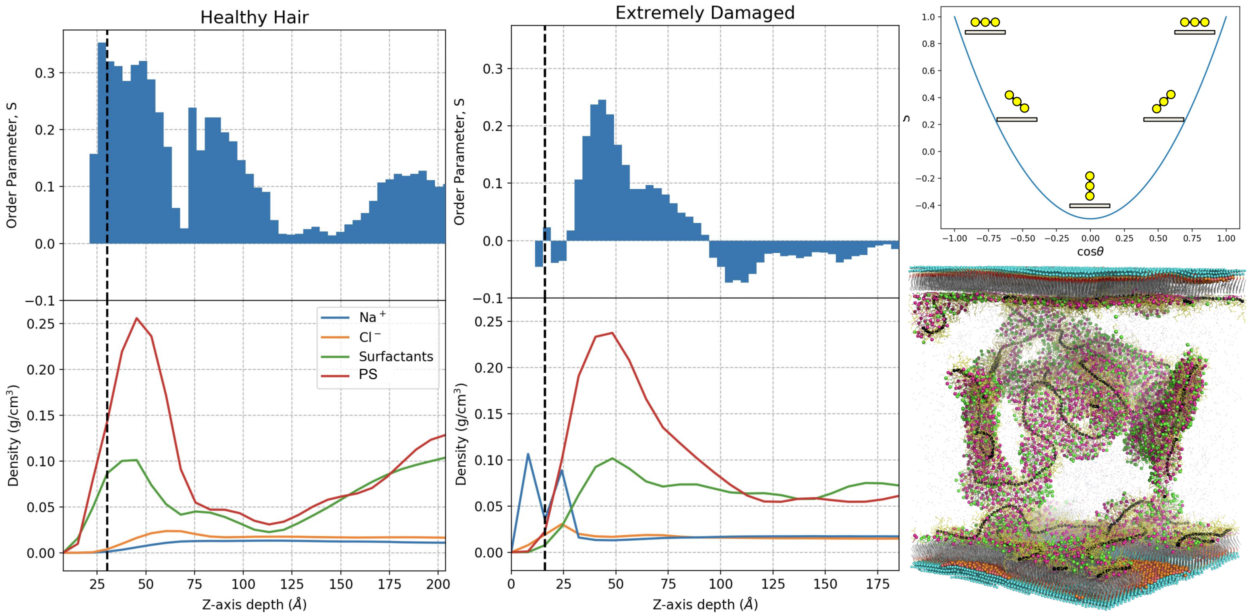

Figure S14 Density profiles aligned with *z*-dependent polymer order parameters for healthy and extremely damaged model hair surfaces in contact with PS formulations. The bottom right image is of a partially damaged system which has structural features consistent with a mixture of the healthy and extremely damaged surfaces. In contrast to M100 and M2003, the high persistence length of PS causes surface-induced order to persist across the bulk solution. The high persistence length is visually apparent in the bottom right image where the black PSB2 beads, representative of the polymer's backbone, undulate gradually throughput space. See the main text for further discussion.

S7. Normal Force During Shearing

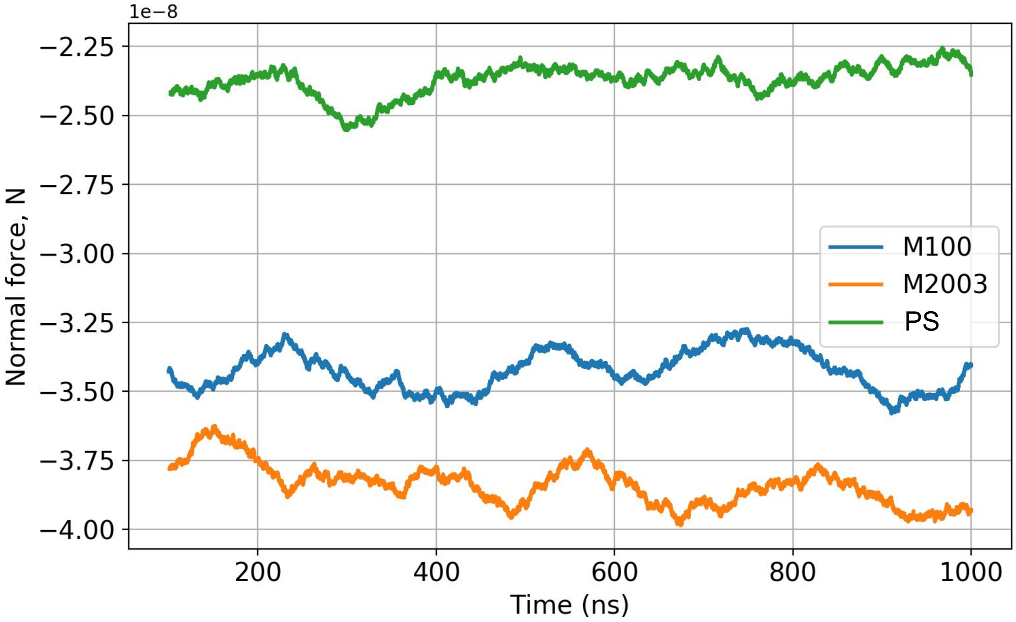

Figure S15. The force normal to the hair surfaces, measured throughout the simulation, stays approximately constant. This implies that the majority of the contributions to the fluctuations of the coefficient of friction in **Figure 6** of the main text are a consequence of forces exerted by the solution while being sheared.