Supplementary Information for

Learning the Relationship between Nanoscale Chemical Patterning and Hydrophobicity

Nicholas B. Rego, Andrew L. Ferguson, and Amish J. Patel

Andrew Ferguson.
E-mail: andrewferguson@uchicago.edu
Amish Patel.
E-mail: amish.patel@seas.upenn.edu

This PDF file includes:

Supplementary text
Figs. S1 to S17
SI References

1. Generating Libraries of Diverse Chemical Patterns

Given that binary self-assembled monolayer (SAM) patches span a large and high-dimensional chemical pattern space, with 2^{36} patterns possible for the 6×6 patches, here we seek to sample a small but diverse subset of such patches. If the patches are chosen randomly, certain polar contents and patterning motifs are much more likely to appear than others; in particular, randomly generated patterns are likely to favor roughly equal numbers of polar and nonpolar end groups with the groups being well-dispersed rather than clustered. To generate a library of diverse patches that span a wide range of polar contents and degrees of dispersion/clustering, we quantify the former using the number of polar groups, n_o, and the latter using the degree of polar clustering, d_o, which is defined as the root mean squared deviation of the polar groups from their center of mass (Figure S1A). For every value of n_o, we then estimated the number of patterns, $\Omega_{n_o}(d_o)$, as a function of d_o either by exhaustive sampling (for n_o-values with fewer than 60,000 patterns) or using the Wang-Landau (1) algorithm with a tolerance of 10^{-10} and a maximum of $60,000$ iterations per round. As illustrated in Figure S1B for $n_o = 18$, $\Omega_{n_o}(d_o)$ is sharply peaked, highlighting that randomly-generated patterns would span a narrow range of d_o-values. To engender pattern diversity, we instead sample uniformly across all possible values of n_o and d_o and obtain 442 distinct patches. To obtain patterns that span a wide range, not only in their d_o-values, but also in their degrees of nonpolar clustering, we swap the identities of polar and nonpolar end groups in each of our 442 patches to generate another 442 'inverted' patches (Figure S1C) and arrive at a total of $N = 884$ patches for our training library of 6×6 patches. This procedure was repeated to additionally generate diverse 4×4 patches ($N = 228$) and 4×9 patches ($N = 682$) for inclusion in the expanded dataset (Figure 4C).

2. Molecular Simulations of Self-Assembled Monolayer (SAM) Surfaces

Simulation set-up. Self-assembled monolayer (SAM) surfaces composed of hexagonally-packed alkyl chains were prepared following references (2, 3). Each chain contains a sulfur atom, 10 united-atom methylene groups, and either a methyl (nonpolar) or hydroxyl (polar) end group that is exposed to solvent. Molecular visualizations of the methyl- and hydroxyl-terminated chains are shown in Figure S2A. The majority of our calculations were performed using a SAM surface, consisting of $12 \times 12 = 144$ alkyl chains, with chemically patterned patches embedded in a polar background (Figure 1A). An elongated SAM surface, composed of $6 \times 36 = 216$ alkyl chains, was used to estimate the hydrophobicity of the 2×12, 2×18, 2×32 and 3×12 uniform nonpolar patches (Figure 4B). Moreover, a larger SAM surface, composed of 16×16 alkyl chains, was used to study the 12×8 nonpolar patch, the non-rectangular nonpolar patches (Figure S12), and the 12×12 patterned patches (Figure S15). SAM surfaces were solvated using a roughly 4 nm thick slab of water and placed normal to the z-dimension in a $5.2 \times 6.0 \times 8.5$ nm^3 simulation box, as shown in Figures 1C and S2B; for the elongated and larger SAM surfaces, simulation boxes of dimensions, $2.6 \times 18 \times 8.5$ nm^3 and $6.93 \times 8.0 \times 8.5$ nm^3, respectively, were used. Following reference (4), a repulsive wall of spherical atoms was positioned at $z = 8.2$ nm to induce a buffering vapor layer, which serves to accommodate any water molecules displaced from the vicinity of the SAM surfaces during the Indirect Umbrella Sampling (INDUS) calculations.

Simulation parameters. All simulations were performed using GROMACS version 4.5.3 (5), suitably modified to incorporate the biasing potentials used in the INDUS calculations. The equations of motion were integrated using the leap-frog integrator (6) with a time step of 2 fs and periodic boundary conditions were used in all dimensions. All systems were simulated in the canonical ensemble with temperature maintained at $T = 300$ K using the stochastic velocity-rescale thermostat (7) and a coupling constant of $\tau = 0.5$ ps. Before performing biased (INDUS) simulations, every patterned SAM system was energy-minimized using the steepest-descent algorithm and equilibrated for 100 ps; biased simulations were run for 500 ps. Water molecules were represented explicitly using the three-point SPC/E model (8). The united-atom methylene groups were modeled using Model UAc from reference (9). The Lennard-Jones parameters and bonded potentials for the methyl and hydroxyl end groups were modeled using the Amber99SB force field (10), whereas their partial charges were taken from the OPLS-AA force field (11). Lennard Jones parameters were combined using the Lorentz-Bertholet mixing rules. Van der Waals and short-range electrostatic interactions were truncated at 1 nm, and long-range electrostatics were treated using the Particle Mesh Ewald method (12). All bonds to hydrogen atoms were constrained; SAM end group bonds were constrained using the LINCS algorithm (13), whereas the bonds in water were constrained using the SETTLE algorithm (14). The sulfur atoms were restrained to lie on a 0.5 nm hexagonal lattice using harmonic restraints with a spring constant of $40,000$ kJ/mol/nm^2. The third methylene groups from the end of the alkyl chain were also position-restrained with the same spring constant (4) to hinder the slow circular motion of the alkyl chains.

3. Characterizing Hydrophobicity using Indirect Umbrella Sampling (INDUS)

The hydrophobicity, f, of a patterned patch was quantified using the free energetic cost, ΔG_{cav}, of displacing all waters from a cuboidal probe volume, v, placed adjacent to the patch (Figure 1B).

Probe volume dimensions. The dimensions, l_x and l_y, of the cuboidal probe volume, v, in the x and y directions, respectively, were chosen to span all end groups within the patch, as illustrated in Figure S3A. In particular, these dimensions were determined by running a 5 ns equilibrium simulation of a purely polar patch, and padding the average x and y coordinates of the top-right and bottom-left hydroxyl end groups in the patch with a buffer distance of $\delta = 0.275$ nm; see Figure S3B) The dimensions, l_x and l_y, of v are related to the patch dimensions, p and q, as follows:

  Nicholas B. Rego, Andrew L. Ferguson, and Amish J. Patel

$$l_x(p) = (2\delta) + r_x(p-1), \text{ and}$$
$$l_y(q) = (2\delta) + r_y(q-0.5), \tag{1}$$

where $r_y = 0.5$ nm is the lattice spacing in the y direction and $r_x = \frac{\sqrt{3}}{2}r_y$ is the spacing between end groups in the x direction.

The width, l_z, of v was chosen to include waters within the first hydration shell of the SAM patches. The lower bound of v was chosen to be $z_{min} = 1.7$, but to ensure that patches of the same size/shape have the same average numbers of waters in v, the upper bound, z_{max}, of v was varied across different patches. An inspection of the normalized density of water oxygens, $g(z)$, adjacent to the uniform 6×6 patches (Figure S3 C-F) makes it clear that a choice of $z_{max} = 2.175$ nm for the polar patch and $z_{max} = 2.3$ nm for the nonpolar patch results in an average of 100 waters in v for both patches. For patterned patches, z_{max} was chosen to be between 2.175 nm and 2.3 nm such that the corresponding v contained roughly 100 waters.

Estimating ΔG_{cav} using INDUS. We use the Indirect Umbrella Sampling (INDUS) method (15, 16) to calculate the probability, $P_v(N)$, of observing N water molecules in a probe volume v, placed adjacent to a patch pattern of interest, and estimate ΔG_{cav} through the relationship, $\Delta G_{cav} = -k_B T \ln P_v(0)$, where $k_B T$ is the thermal energy and k_B is Boltzmann's constant. In particular, we indirectly sampled the number of waters in v by sampling the coarse-grained number, $\tilde{N}_v$, of waters in v, defined using a coarse-graining length of 0.01 nm and a cutoff of 0.02 nm (16). To systematically sample $\tilde{N}_v$, we used a series of harmonic biasing potentials, $\mathcal{U}_{\kappa,N^*}(\tilde{N}_v) = \frac{\kappa}{2}(\tilde{N}_v - N^*)^2$, where the values of κ and N^* were chosen following reference (17). The free energetics, $-k_B T \ln P_v(N)$, of water density fluctuations in v were computed following reference (16), and are shown in Figure S4 for three 6×6 patches along with the corresponding values of ΔG_{cav}.

4. Training and Evaluating Linear Models

All linear models (M1, M2, M3, and M7) were trained using ordinary least squares regression using the Scikit-Learn Python package (18). The performance of each model was evaluated by five-fold randomized cross-validation (CV), and quantified using the root-mean squared error (RMSE), ε, over the five CV rounds. Five-fold CV was repeated 1000 times to obtain error bars for model errors ε and for the parameters of models M2 and M3 (Figure 4D and Figure S14B).

Performance of one-feature models. To asses the usefulness of polar content, n_o, as a predictor of hydrophobicity, we compared the linear model M1 (Figure S5A, $\varepsilon = 3.67$ $k_B T$) to non-linear functional forms of n_o, such as quadratic (Figure S5B, $\varepsilon = 3.42$ $k_B T$) or higher order polynomials. We also considered a 'conditional average' model (Figure S5C, $\varepsilon = 3.53$ $k_B T$), where the hydrophobicity of a patch with n_o polar groups is predicted to be the average hydrophobicity, $\bar{f}_{n_o}$, of all patches (in the training set) with n_o polar groups. The non-linear models perform only marginally better than model M1 (Figure S5D), suggesting that the primary limitation on the accuracy of these one-feature models is their inability to account for the variation in the hydrophobicity of patches with the same polar content, n_o.

5. Artificial and Convolutional Neural Network Models

Artificial (ANN) and convolutional (CNN) neural networks were trained on the $N = 884$ dataset of 6×6 patch patterns using the PyTorch (19) Python library. All ANN and CNN models were trained on model M1 residuals, $f(\boldsymbol{x}) - \hat{f}_{M1}(n_o(\boldsymbol{x}))$.

Data augmentation and featurization of patch patterns. To account for the invariance of patch hydrophobicity under translation, rotation, and reflection, we augmented our $N = 884$ training set of 6×6 patches with isomorphic variants, as shown in Figure S6. Every patch in the augmented dataset ($N = 74,256$) was embedded in a 12×11 SAM surface with a background of polar groups, and was represented using a 132-component feature vector, $\boldsymbol{x}$. The components of $\boldsymbol{x}$ correspond to end group positions, and are assigned a value of $(+1)$ for nonpolar patch groups, (-1) for polar patch groups, and (0) for polar groups outside the patch.

Training neural network models. The ANN and CNN models were trained by minimizing the mean squared error (MSE) using stochastic gradient descent with the Adam (20) optimization algorithm included in the PyTorch package. For the Adam optimizer, the learning rate was set to 1×10^{-3}, β values were chosen to be 0.9 and 0.999, and $\epsilon = 1 \times 10^{-8}$. All models were trained with mini-batch gradient descent using a batch size of 200 samples. Early-stopping was used to terminate training if model MSE did not improve for 10 consecutive epochs. Rectified linear unit (ReLU) activation functions were employed in all nodes. CNNs were constructed using the Hexagdly (21) PyTorch plugin to enable convolutions over the hexagonally-packed lattice of end groups. Convolutional filters used a kernel size of 1 with a stride length of 1 and max pooling filters employed a kernel size of 1 with a stride length of 2; the kernel size employed corresponds to the nearest-neighbor kernel shown in Figure 2B. A padding length of 1 was used for all filters. All CNN architectures employed two convolutional layers applied in sequence. In each layer, the same number of convolutional filters was applied, in parallel, to each patch pattern; the convolved patterns were passed through a ReLU activation filter and then through a max-pooling operation to reduce the dimensionality of the patterns. The output of the second convolutional layer was passed to a fully connected ANN.

117 **Evaluating ANN and CNN model performance.** As with the linear models, the performance of the ANN and CNN models was
118 evaluated using 5-fold CV. To determine optimal network architectures for models M1A and M1C, we trained models with
119 different hyper-parameter combinations, and evaluated their performance as an average over five independent CV rounds.

120 **Determining the optimal ANN architecture.** The optimal number of hidden layers and nodes per hidden layer in the ANN model
121 were determined by training a series of models with different architectures. The performance, ε, of different ANN architectures
122 is shown in Figure S7. The best performing ANN architecture had 1 hidden layer with 48 neurons and an output layer with
123 one neuron, employing a total of 6,433 trained parameters. This architecture was chosen for model M1A, and was trained over
124 five independent CV rounds to evaluate its performance ($\varepsilon = 2.75\ k_{\mathrm{B}}T$).

125 **Determining the optimal CNN architecture.** The number of convolutional filters, hidden ANN layers, and nodes per hidden layer
126 of the CNN model were similarly determined by training a variety of CNN models with different hyper-parameter combinations
127 (Figure S8). We chose a near-optimal CNN model that had 5 convolutional filters, 1 hidden layer with 4 nodes and an output
128 layer with one neuron, employing a total of 349 trained parameters. Although more complicated CNN architectures performed
129 slightly better, we selected this architecture for model M1C because it corresponds to a very high performing model with the
130 fewest parameters. The error of model M1C, estimated over five independent CV rounds, was $\varepsilon = 2.72\ k_{\mathrm{B}}T$.

6. Two-body Feature Constraints and Isomorphic Forms of Model M2

132 To incorporate two-body or nearest-neighbor end group correlations into our models, we must consider the number of polar
133 (n_{oo}), nonpolar (n_{cc}) and mixed (n_{oc}) neighbors. Although the number of neighbors of each type varies with the patch pattern,
134 $\boldsymbol{x}$, the total number of nearest neighbors, $n_{\mathrm{xx}} \equiv n_{\mathrm{oo}} + n_{\mathrm{oc}} + n_{\mathrm{cc}}$ depends only on patch dimensions, p and q, for a rectangular
135 patch:

$$n_{\mathrm{oo}} + n_{\mathrm{oc}} + n_{\mathrm{cc}} = n_{\mathrm{xx}}(p,q) = 3pq + 2(p+q) - 1. \tag{2}$$

137 Moreover, stoichiometric and geometric considerations dictate that the number of polar and mixed neighbors must be related
138 to the total number of polar end groups through:

$$2n_{\mathrm{oo}} + n_{\mathrm{oc}} = 6n_{\mathrm{o}} + [2n_{\mathrm{xx}}(p,q) - 6pq], \tag{3}$$

140 where the last term on the right hand side depends only on the patch dimensions. Thus, once the patch dimensions and n_{o} are
141 specified, the nearest-neighbor numbers, n_{oo}, n_{oc}, and n_{cc} are related to one another through equations 2 and 3 with only one
142 of the three variables being independent. By choosing n_{oo} to be the independent variable (along with n_{o}), we obtain the linear
143 model M2 for 6×6 patches as:

$$\hat{f}_{\mathrm{M2}}(n_{\mathrm{o}}, n_{\mathrm{oo}}) = 105.68 + 5.39\ n_{\mathrm{o}} - 0.38\ n_{\mathrm{oo}} \tag{4}$$

145 By equivalently choosing n_{oc} and n_{o} as the independent variables, we obtain the following isomorphic model M2:

$$\hat{f}_{\mathrm{M2}}(n_{\mathrm{o}}, n_{\mathrm{oc}}) = 96.86 + 4.24\ n_{\mathrm{o}} + 0.19\ n_{\mathrm{oc}} \tag{5}$$

147 This form of model M2 emphasizes that placing polar and nonpolar groups next to one another increases patch hydrophilicity.
148 We can similarly choose n_{cc} and the number of nonpolar groups, $n_{\mathrm{c}} = pq - n_{\mathrm{o}}$ as the independent variables to obtain:

$$\hat{f}_{\mathrm{M2}}(n_{\mathrm{c}}, n_{\mathrm{cc}}) = 249.57 - 3.09\ n_{\mathrm{c}} - 0.38\ n_{\mathrm{cc}} \tag{6}$$

150 We emphasize that Equations 4, 5 and 6 contain the same information and that other isomorphic models can also be constructed
151 by choosing one feature each from the sets, $\{n_{\mathrm{o}}, n_{\mathrm{c}}\}$ and $\{n_{\mathrm{oo}}, n_{\mathrm{oc}}, n_{\mathrm{cc}}\}$ with different feature choices providing different
152 perspectives on the chemical determinants of hydrophobicity. For instance, the intercept in equation 6 corresponds to the
153 hydrophobicity of a uniform polar patch (for which $n_{\mathrm{c}} = n_{\mathrm{cc}} = 0$), whereas the coefficients capture the extent to which the
154 presence of nonpolar groups, as quantified by n_{c}, and their nonpolar patterning, as quantified by n_{cc}, contribute to patch
155 hydrophobicity. This form of model M2 clarifies that the addition of a nonpolar end group to a purely polar patch increases its
156 hydrophobicity (i.e., decreases $\hat{f}$ by 3.09 $k_{\mathrm{B}}T$), its effect is smaller than the addition of a polar group to a purely nonpolar
157 patch (which increases $\hat{f}$ by 5.39 $k_{\mathrm{B}}T$ according to equation 4).

7. Three-body Feature Constraints

There are 16 possible three-body features (Figure S9) that span different chemistries ('ooo', 'ccc', 'ooc', 'cco', 'oco', 'coc') and
shapes ('collapsed', 'bent', or 'extended') with certain features considered equivalently due to symmetry; e.g., $n_{\mathrm{ooc}}^{\mathrm{compact}}$ counts
the number of both 'ooc' and 'oco' compact clusters. The total number of 'compact', 'bent', and 'straight' three-body terms
($n_{\mathrm{xxx}}^{\mathrm{compact}}$, $n_{\mathrm{xxx}}^{\mathrm{bent}}$, and $n_{\mathrm{xxx}}^{\mathrm{straight}}$, respectively) must be constant for patches of a given size and shape:

$$n_{\mathrm{ooo}}^{\mathrm{compact}} + n_{\mathrm{ccc}}^{\mathrm{compact}} + n_{\mathrm{ooc}}^{\mathrm{compact}} + n_{\mathrm{occ}}^{\mathrm{compact}} = n_{\mathrm{xxx}}^{\mathrm{compact}}(p,q) = 2pq + 2(p+q)$$

$$n_{\mathrm{ooo}}^{\mathrm{bent}} + n_{\mathrm{ccc}}^{\mathrm{bent}} + n_{\mathrm{ooc}}^{\mathrm{bent}} + n_{\mathrm{oco}}^{\mathrm{bent}} + n_{\mathrm{occ}}^{\mathrm{bent}} + n_{\mathrm{coc}}^{\mathrm{bent}} = n_{\mathrm{xxx}}^{\mathrm{bent}}(p,q) = 6pq + 7p + 8q - 2$$

$$n_{\mathrm{ooo}}^{\mathrm{straight}} + n_{\mathrm{ccc}}^{\mathrm{straight}} + n_{\mathrm{ooc}}^{\mathrm{straight}} + n_{\mathrm{oco}}^{\mathrm{straight}} + n_{\mathrm{occ}}^{\mathrm{straight}} + n_{\mathrm{coc}}^{\mathrm{straight}} = n_{\mathrm{xxx}}^{\mathrm{straight}}(p,q) = 3pq + 4(p+q) - 2$$

 Nicholas B. Rego, Andrew L. Ferguson, and Amish J. Patel

We chose to eliminate the purely nonpolar terms $n_{\text{ccc}}^{\text{compact}}, n_{\text{ccc}}^{\text{bent}}$, and $n_{\text{ccc}}^{\text{straight}}$. Furthermore, by incorporating polar content, n_{o}, and the total number of polar-polar neighbors, n_{oo}, we can construct 8 additional constraint equations, enumerated in Figure S10:

$$n_{\text{ooc}}^{\text{compact}} + 3n_{\text{ooo}}^{\text{compact}} - 2n_{\text{oo}} = 2(p+q) + 2$$

$$n_{\text{ooc}}^{\text{bent}} + 2n_{\text{ooo}}^{\text{bent}} - 4n_{\text{oo}} = 6p + 8q$$

$$n_{\text{ooc}}^{\text{straight}} + 2n_{\text{ooo}}^{\text{straight}} - 2n_{\text{oo}} = 4(p+q) - 2$$

$$n_{\text{occ}}^{\text{compact}} - 3n_{\text{ooo}}^{\text{compact}} + 4n_{\text{oo}} - 6n_{\text{o}} = 2(p+q) - 4$$

$$n_{\text{occ}}^{\text{bent}} + 2n_{\text{oco}}^{\text{bent}} + 4n_{\text{oo}} - 12n_{\text{o}} = 8(p+q) - 4$$

$$n_{\text{occ}}^{\text{straight}} + 2n_{\text{oco}}^{\text{straight}} + 2n_{\text{oo}} - 6n_{\text{o}} = 4(p+q) - 2$$

$$n_{\text{coc}}^{\text{bent}} - n_{\text{ooo}}^{\text{bent}} + 4n_{\text{oo}} - 6n_{\text{o}} = p - 2$$

$$n_{\text{coc}}^{\text{straight}} - n_{\text{ooo}}^{\text{straight}} + 2n_{\text{oo}} - 3n_{\text{o}} = 0$$

These constraints result in the set of 7 linearly independent features, $\{n_{\text{o}}, n_{\text{oo}}, n_{\text{ooo}}^{\text{compact}}, n_{\text{ooo}}^{\text{bent}}, n_{\text{ooo}}^{\text{straight}}, n_{\text{oco}}^{\text{bent}}, n_{\text{oco}}^{\text{straight}}\}$, which were used to construct model M7.

8. Learning Curves: Model Performance vs Size of the Training Dataset

All our models for predicting the hydrophobicity of 6×6 patterned SAM patches were trained using a library comprising $N = 884$ patch patterns. To assess how model performance depends on the size of the training dataset, we randomly select $M = 177$ (roughly one-fifth of N) patches from the library and designate them to be the test-set. This procedure is repeated a total of five times using a fifth of the dataset as the test set each time. Each model is then trained on increasingly larger subsets of the remaining $N - M = 767$ patches, and model error, ε, is reported using the RMSE calculated on the five different test sets. As the selection of the testing set is randomized, this procedure is repeated 4 times to estimate uncertainties in ε for models M1A and M1C, and 1000 times to estimate the uncertainties in ε for the linear models. The learning curves, which plot model performance, ε, as a function of training set size, are shown for both the linear (Figure S11A) and neural network (Figure S11B) models. The performance of the linear models appears to saturate as the size of the training set is increased, whereas the performance of the neural network models continues to improve (for $< 10^3$ patches), suggesting that increasing the size of the training set is unlikely to significantly improve the performance of models M1 to M7, whereas larger libraries are likely to improve the performance of models M1A and M1C. In Figure S11B, we also plot the performance of the best-performing model as a function of training set size (multi-colored line); model M2 (orange segment) initially out-performs all other models for small libraries (< 30 patches), whereas the model M3 performs optimally when modest training sets (< 300 patches) are used, and model M7 performs best when trained on the largest datasets. These results suggest that as the size of the training dataset is increased, the neural network models will continue to display improved performance by learning higher order correlations in the data; in contrast, the linear models will need to employ additional descriptors that capture increasingly higher order correlations to optimally utilize larger datasets.

9. Explaining Models M1A and M1C using Surrogate Models

To shed light into why our neural network (NN) models perform well as they do, we draw inspiration from the Local Interpretable Model-agnostic Explanations (LIME) technique (22), proposed by Ribeiro and coworkers to explain predictions made by black-box models. LIME relies on training interpretable 'surrogate' models to explain the predictions of more complex, un-interpretable models. Based on the excellent performance of our linear models (M2, M3 and M7), we propose using them as interpretable surrogates for explaining our NN models (M1A and M1C). To this end, we first consider a substantially expanded dataset with 88,002 patches, which contains not only the 884 patches included in the original training set, but also an additional $87,118$ patterns generated using Wang-Landau sampling. We then used a NN model M1Y (where $Y \in \{A,C\}$) to predict the hydrophobicity of these 88,002 patches, and using these predictions, we trained a surrogate model MX (where $X \in \{2,3,7\}$). We found that all surrogate models MX| M1Y explain the underlying NN model remarkably well with errors of roughly 1 $k_{\text{B}}T$, and that the coefficients of the surrogate models are in good agreement with those of linear models MX obtained by training with the simulation dataset of 884 patches. The surrogate models and their performances are as follows:

Name	Model	ε ($k_{\text{B}}T$)
$\hat{f}_{\text{M2}}$	$105.7 + 5.39\,n_{\text{o}} - 0.384\,n_{\text{oo}}$	$-$
$\hat{f}_{\text{M2}|\text{M1A}}$	$105.59 + 5.313\,n_{\text{o}} - 0.3545\,n_{\text{oo}}$	0.99
$\hat{f}_{\text{M2}|\text{M1C}}$	$106.07 + 5.319\,n_{\text{o}} - 0.3637\,n_{\text{oo}}$	1.00
$\hat{f}_{\text{M3}}$	$105.8 + 5.40\,n_{\text{o}} - 0.49\,(n_{\text{oo}} + n_{\text{ooo}}^{\text{compact}}) + 0.13\,(n_{\text{ooo}}^{\text{bent}} + n_{\text{ooo}}^{\text{straight}})$	$-$
$\hat{f}_{\text{M3}|\text{M1A}}$	$106.00 + 5.154\,n_{\text{o}} - 0.292\,(n_{\text{oo}} + n_{\text{ooo}}^{\text{compact}}) + 0.0547\,(n_{\text{ooo}}^{\text{bent}} + n_{\text{ooo}}^{\text{straight}})$	1.02
$\hat{f}_{\text{M3}|\text{M1C}}$	$106.42 + 5.191\,n_{\text{o}} - 0.331\,(n_{\text{oo}} + n_{\text{ooo}}^{\text{compact}}) + 0.0693\,(n_{\text{ooo}}^{\text{bent}} + n_{\text{ooo}}^{\text{straight}})$	0.99
$\hat{f}_{\text{M7}}$	$105.7 + 5.89\,n_{\text{o}} - 0.64\,n_{\text{oo}} - 0.51\,n_{\text{ooo}}^{\text{compact}} + 0.16\,n_{\text{ooo}}^{\text{bent}} + 0.11\,n_{\text{ooo}}^{\text{straight}} - 0.05\,n_{\text{oco}}^{\text{bent}} - 0.04\,n_{\text{oco}}^{\text{straight}}$	$-$
$\hat{f}_{\text{M7}|\text{M1A}}$	$105.08 + 5.703\,n_{\text{o}} - 0.658\,n_{\text{oo}} - 0.013\,n_{\text{ooo}}^{\text{compact}} + 0.045\,n_{\text{ooo}}^{\text{bent}} + 0.078\,n_{\text{ooo}}^{\text{straight}} - 0.023\,n_{\text{oco}}^{\text{bent}} - 0.032\,n_{\text{oco}}^{\text{straight}}$	0.93
$\hat{f}_{\text{M7}|\text{M1C}}$	$105.91 + 5.695\,n_{\text{o}} - 0.582\,n_{\text{oo}} - 0.202\,n_{\text{ooo}}^{\text{compact}} + 0.103\,n_{\text{ooo}}^{\text{bent}} + 0.025\,n_{\text{ooo}}^{\text{straight}} - 0.044\,n_{\text{oco}}^{\text{bent}} + 0.014\,n_{\text{oco}}^{\text{straight}}$	0.93

193 These findings suggest that the trained NN models, which operate on the entire chemical pattern of the patch, $\boldsymbol{x}$, extract
194 information pertaining to local chemical correlations between patch end groups that is encoded in the descriptors (e.g., n_{oo})
195 used to construct the linear models (M2, M3 and M7).

10. Uniform nonpolar Patches of Various Shapes and Sizes

197 To construct a physically-motivated model for the hydrophobicity of a purely nonpolar patch, f_c, as a function of its size and
198 shape, we note that cavity formation is expected to be driven by interfacial physics, and hence its free energetic cost should be
199 proportional to the cross-sectional area, $l_x l_y$, of each cuboidal probe volume v. To account for possible edge-effects due to the
200 finite size of the patches considered here, we additionally included the perimeter of the v in the x and y dimensions, ($2l_x$ and
201 $2l_y$, respectively, for a rectangular $p \times q$ patch) to arrive at the following linear model:

$$\hat{f}_c(l_x, l_y) = c_1(l_x l_y) + c_2(l_x) + c_3(l_y) \tag{7}$$

We trained this model on a dataset of 62 purely nonpolar patches of different sizes and shapes, resulting in the following learned
coefficients:

$$c_1 = 5.63 \ k_B T/\text{nm}^2$$
$$c_2 = 11.85 \ k_B T/\text{nm}$$
$$c_3 = 7.54 \ k_B T/\text{nm}$$

To express $\hat{f}_c$ as a function of patch dimensions p and q, we substituted the relationships, $l_x(p)$ and $l_y(q)$ (Equation 1), into
Equation 7 to obtain:

$$\begin{aligned}
\hat{f}_c(p, q) &= \hat{f}_c(l_x(p), l_y(q)) \\
&= [c_1 r_x r_y]pq + [(2\delta c_1 - 0.5 r_y c_1 + c_2)r_x]p + [(2\delta c_1 - r_x c_1 + c_3)r_y]q \\
&\quad + \left[c_1(2\delta - r_x)(2\delta - 0.5 r_y) + c_2(2\delta - r_x) + c_3(2\delta - 0.5 r_y) \right]
\end{aligned}$$

The coefficients and intercept of $\hat{f}_c(p, q)$ were then determined from the trained coefficients $c_1, c_2,$ and c_3 and the known values
of δ, r_x, and r_y to obtain:

$$\hat{f}_c(p, q) = 1.22pq + 5.86p + 4.10q + 3.85$$

203 This model is able to capture the hydrophobicity of rectangular nonpolar patches with an error of $\varepsilon = 1.87 \ k_B T$.

Extending $\hat{f}_c$ to non-rectangular patches. To generalize $\hat{f}_c$ to non-rectangular patches, we recast the model in terms of the total
number of patch end groups, n_{tot}, and the number of end groups along the perimeter of patch in the x and y dimensions, n_x
and n_y, respectively (Figure S12A). Because $n_{tot} = pq$, $n_x = 2p$ and $n_y = 2q$ for a rectangular $p \times q$ patch, our generalized
model for nonpolar patch hydrophobicity becomes:

$$\hat{f}_c(n_{tot}, n_x, n_y) = 1.22 n_{tot} + 2.93 n_x + 2.05 n_y + 3.85$$

204 To test this model, we estimated the hydrophobicity, f_c, of the 7 non-rectangular nonpolar patches shown in Figure S12B, and
205 found that the generalized model $\hat{f}_c(n_{tot}, n_x, n_y)$ is able to accurately predict f_c with an error of $\varepsilon = 4.75 \ k_B T$.

11. Uniform Polar Patches of Various Shapes and Sizes

We also construct a physically-motivated model for the hydrophobicity of a purely polar rectangular patch, f_o, as a function of
its dimensions, p and q. To do so, we estimate f_o for 62 patches of different sizes and shapes, and use this dataset to train a
model $\hat{f}_o(p, q)$ that is linear in pq, p and q, resulting in:

$$\hat{f}_o(p, q) = 5.81pq + 5.12p + 1.79q + 1.51$$

207 As shown in Figure S13, this model is able to predict f_o quite well with an error of $\varepsilon = 4.01 \ k_B T$.

208 We note that the generalized model M2: $\hat{f}_{M2}(p, q, n_o, n_{oo}) = \alpha_c(p, q) + \alpha_o n_o + \alpha_{oo} n_{oo}$, discussed in the main text, can also
209 be used to predict the hydrophobicity, $f_o(p, q)$, of polar rectangular patches of all shapes and sizes. In particular, focusing on
210 the leading order terms (i.e., those proportional to pq), and noting that for a purely polar patch, $n_o = pq$ and n_{oo} grows as $3pq$,
211 we obtain the constraint: $5.81 = 1.22 + \alpha_o + 3\alpha_{oo}$, where 5.81 and 1.22 are the coefficients of pq in the models for $\hat{f}_o(p, q)$ and
212 $\hat{f}_c(p, q)$, respectively. Plugging the values of $\alpha_o = 5.07$ and $\alpha_{oo} = -0.32$, obtained for model M2, into the right hand side of
213 the constraint equation, we obtain 5.33, which is within 10% of 5.81 (left hand side of the constraint).

Nicholas B. Rego, Andrew L. Ferguson, and Amish J. Patel

12. Generalizing Model M3 to Rectangular Patches of all Shapes and Sizes

To generalize model M3 $[\hat{f}_{M3} = \alpha_c + \alpha_o n_o + \alpha_{oo}(n_{oo} + n_{ooo}^{\text{compact}}) + \alpha_{ooo}(n_{ooo}^{\text{straight}} + n_{ooo}^{\text{bent}})]$ to rectangular patches of all shapes and sizes, we first note that the intercept, α_c, in model M3 corresponds to the hydrophobicity, f_c, of a uniform nonpolar patch for which $n_o = n_{oo} = n_{ooo}^{\text{compact}} = n_{ooo}^{\text{straight}} = n_{ooo}^{\text{bent}} = 0$. Thus, we expect α_c to depend on patch dimensions according to: $\alpha_c(p,q) = \hat{f}_c(p,q) = 1.22pq + 5.86p + 4.10q + 3.85$. To assess whether the model M3 coefficients, α_o, α_{oo} and α_{ooo}, are independent of patch size and shape, we compare their values obtained by training the original 6×6 and the expanded datasets (Figure S14). We find good qualitative agreement across all the trained coefficients, and reasonable quantitative agreement, especially for the coefficients that are larger in magnitude (and therefore more important). Our generalized model M3, trained on the expanded dataset, which can be used to predict the hydrophobicity of heterogeneous rectangular patches of all shapes and size, and is thus given by: $\hat{f}_{M3} = \alpha_c(p,q) + \alpha_o n_o + \alpha_{oo}(n_{oo} + n_{ooo}^{\text{compact}}) + \alpha_{ooo}(n_{ooo}^{\text{straight}} + n_{ooo}^{\text{bent}}) = 1.22pq + 5.86p + 4.10q + 3.85 + 4.98n_o - 0.28(n_{oo} + n_{ooo}^{\text{compact}}) + 0.06(n_{ooo}^{\text{straight}} + n_{ooo}^{\text{bent}})$.

13. Patterned Patches of Different Sizes

To interrogate the influence of chemical patterning on hydrophobicity for patches of different sizes, we studied 6×6, 8×8, 10×10 and 12×12 patterned patches. For each patch size, we estimated the hydrophobicity, f, of two judiciously chosen patches that had the same number of polar groups but different chemical patterns; both patches had an equal number of polar and nonpolar groups, but their nonpolar end groups were either clustered or dispersed, as shown in Figure S15. We chose dispersed and clustered patches with the same n_o because model M1 would predict that these patches should have the same hydrophobicity. In contrast, we found that differences in chemical patterning gave rise to differences in hydrophobicity for all patch sizes, and that these differences increased with patch size (Figure S15). Moreover, our generalized models M2 and M3 correctly predict that patches with well-dispersed end groups are more hydrophilic than those with clustered end groups, and they capture the differences in hydrophobicity with chemical patterning reasonably well across the different patch sizes.

14. Rational Design of Patterned Patches using Model M3

We used model M3 to perform the hot-spot analysis and the greedy design described in the main text, and found that the locations of the hot-spot end groups as well as the characteristics of the maximally hydrophobic or hydrophilic patches (Figure S16) are qualitatively similar to those discovered by model M2 (Figure 5). Additionally, we compare the hydrophobicity of the maximally hydrophilic ($\hat{f}^{\text{max}}$) and the maximally hydrophobic ($\hat{f}^{\text{min}}$) patches with a certain polar content, n_o, obtained using models M2 and M3 (Figure S17A). In both cases, the maximum range of attainable hydrophobicities, i.e., $\hat{f}^{\text{max}} - \hat{f}^{\text{min}}$, is realized at roughly 50% polar coverage (Figure S17B).
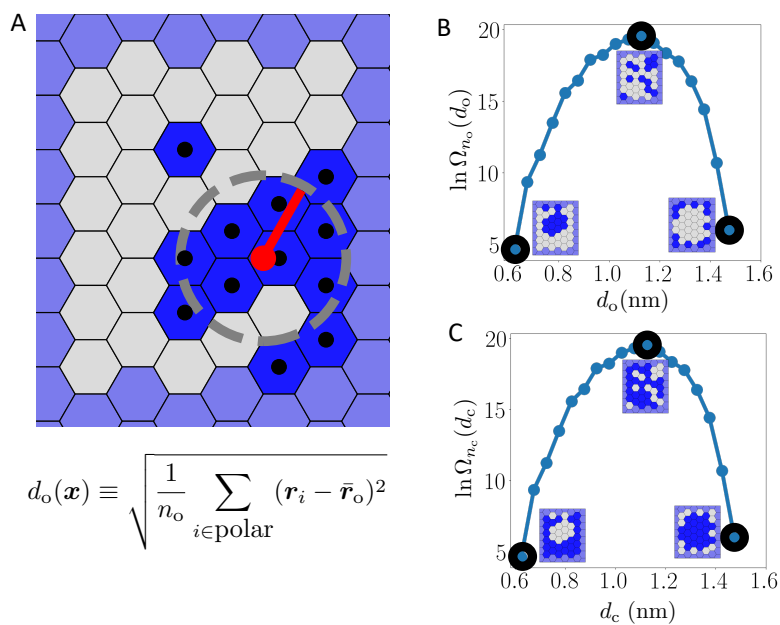
$$d_{\mathrm{o}}(\boldsymbol{x}) \equiv \sqrt{\frac{1}{n_{\mathrm{o}}} \sum_{i \in \mathrm{polar}} (\boldsymbol{r}_i - \bar{\boldsymbol{r}}_{\mathrm{o}})^2}$$

Fig. S1. Using polar end group clustering to construct a library of diverse patterns. (A) The degree of polar clustering, d_{o}, of a patch pattern, $\boldsymbol{x}$, is defined as the root mean squared deviation of the positions, $\boldsymbol{r}_i$ (black dots), of all n_{o} polar end groups (hydroxyl, blue) in the patch around their centroid, $\bar{\boldsymbol{r}}_{\mathrm{o}}$ (red dot). The radius (red) of the illustrated circle (gray) corresponds to d_{o} of the pattern shown here. nonpolar (methyl) end groups are shown in white and polar groups outside of the patch are shown in light blue. (B) The number of 6×6 patches with n_{o} polar groups, $\Omega_{n_{\mathrm{o}}}(d_{\mathrm{o}})$, shown here for $n_{\mathrm{o}} = 12$, is peaked at $d_{\mathrm{o}} \approx 1.1$ nm, highlighting that highly clustered (lower left) or dispersed (lower right) patch patterns are improbable. (C) To obtain patterns spanning a wide range of d_{c} (degree of nonpolar clustering, defined analogously to d_{o}) values, we swap the identities of polar (hydroxyl, blue) and nonpolar (methyl, white) end groups to generate 'inverted' patches. The number of 6×6 patches with n_{c} nonpolar groups, $\Omega_{n_{\mathrm{c}}}(d_{\mathrm{c}})$, is shown here for $n_{\mathrm{c}} = 12$.

Nicholas B. Rego, Andrew L. Ferguson, and Amish J. Patel
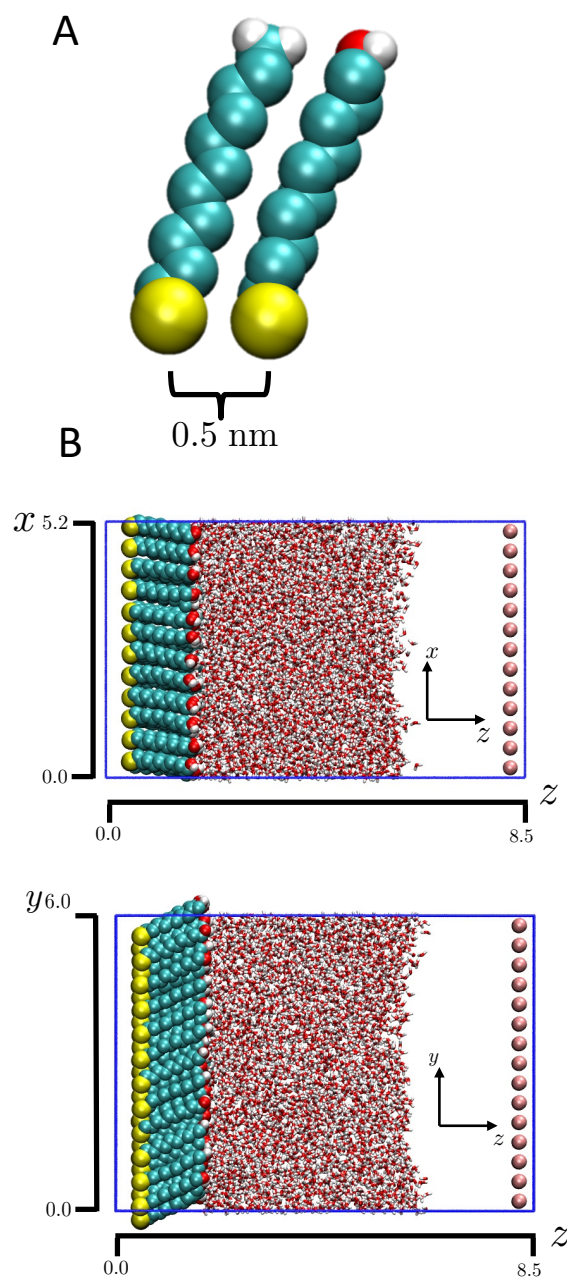
Fig. S2. Simulating self-assembled monolayer (SAM) surfaces. (A) The SAM surfaces are comprised of hexagonally-packed alkyl chains, which contain a sulfur atom (yellow), 10 united-atom methylene groups (teal), and either a nonpolar methyl (left, teal/white) or polar hydroxyl (right, red/white) end group (B) The sulfur atoms are spaced 0.5 nm apart and are arranged on a hexagonal lattice in the $x - y$ plane. The SAM surface (space-fill representation) is solvated using a roughly 4 nm thick slab of water (red/white, licorice representation) A repulsive wall is placed at $z \approx 8.2$ nm (pink spheres) to induce a buffering vapor layer.

A top

l_y

l_x

B $\delta = 0.275$ nm

$r_y = 0.5$ nm

$r_x = \dfrac{\sqrt{3}}{2} r_y \approx 0.43$ nm

$l_y(q) = (2\delta) + r_y(q - 0.5)$

$l_x(p) = (2\delta) + r_x(p - 1)$

C pure polar

D pure non-polar

E

$l_z = 0.475$ nm

$g(z)$

z (nm)

F

$l_z = 0.600$ nm

z (nm)
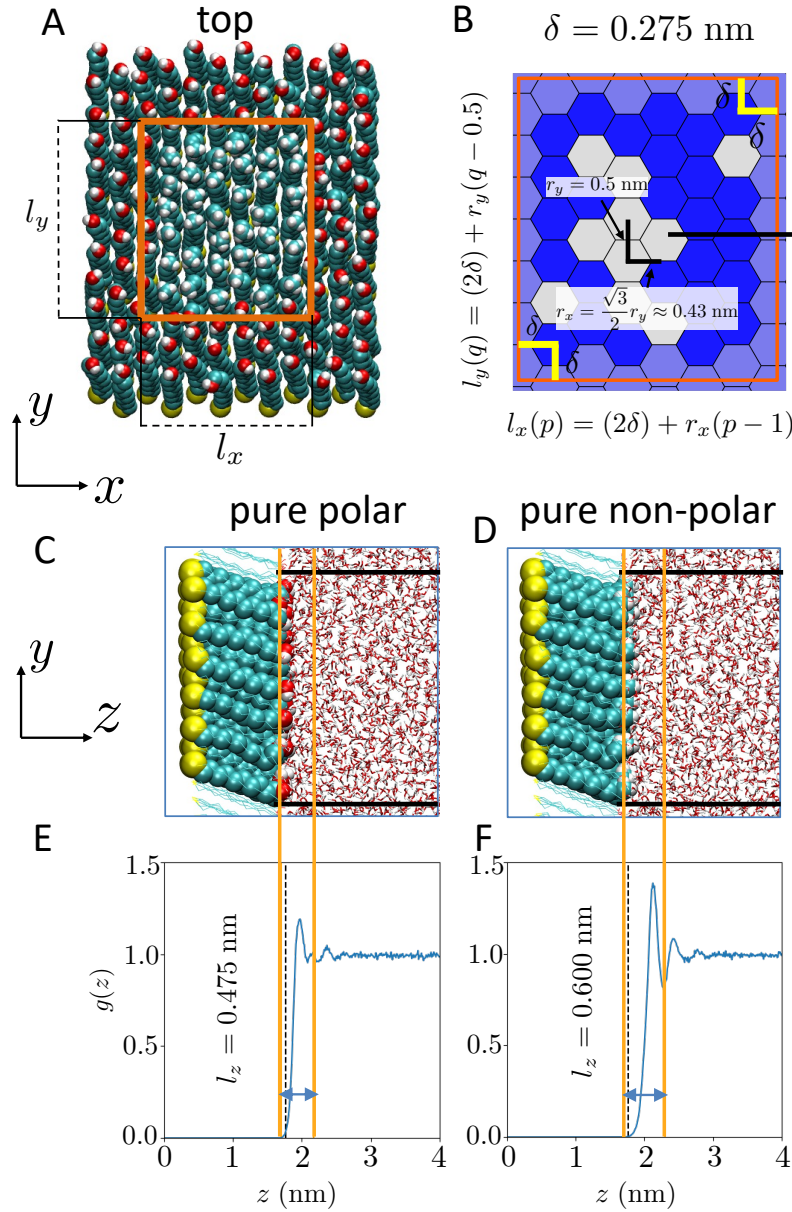
Fig. S3. Defining the cuboidal probe volume, v. (A) A top-down view illustrating a nonpolar 6×6 pattern and the dimensions, l_x and l_y, of v (orange) in the x and y directions, respectively. (B) The dimensions, l_x and l_y, are determined by calculating the average positions of the lower-left and upper-right patch end groups in a patch and padding them with a buffer, δ (yellow lines) to ensure that v encompasses all end groups in the patch. The relationships between the dimensions of v, i.e., l_x and l_y, and the patch dimensions, p or q are shown, where r_x and r_y are the average end group spacings in the x and y directions, respectively. (C, D) Simulation snapshots of water near the purely polar and nonpolar 6×6 patches, and (E, F) the corresponding normalized water density profiles, $g(z)$. To ensure that probe volumes adjacent to all patches have roughly the same average number of waters, their lower z-bounds were fixed at $z_{\min} = 1.7$ nm, and their upper bounds, $z_{\max}$, were varied. Choosing $z_{\max} = 2.175$ nm for the polar patch and $z_{\max} = 2.3$ nm for the nonpolar patch results in 100 waters in both probe volumes; for patterned surfaces containing a mixture of polar and nonpolar end groups, $z_{\max}$ was chosen to be between these two extremes such that roughly 100 waters were present within the probe volume.

Nicholas B. Rego, Andrew L. Ferguson, and Amish J. Patel
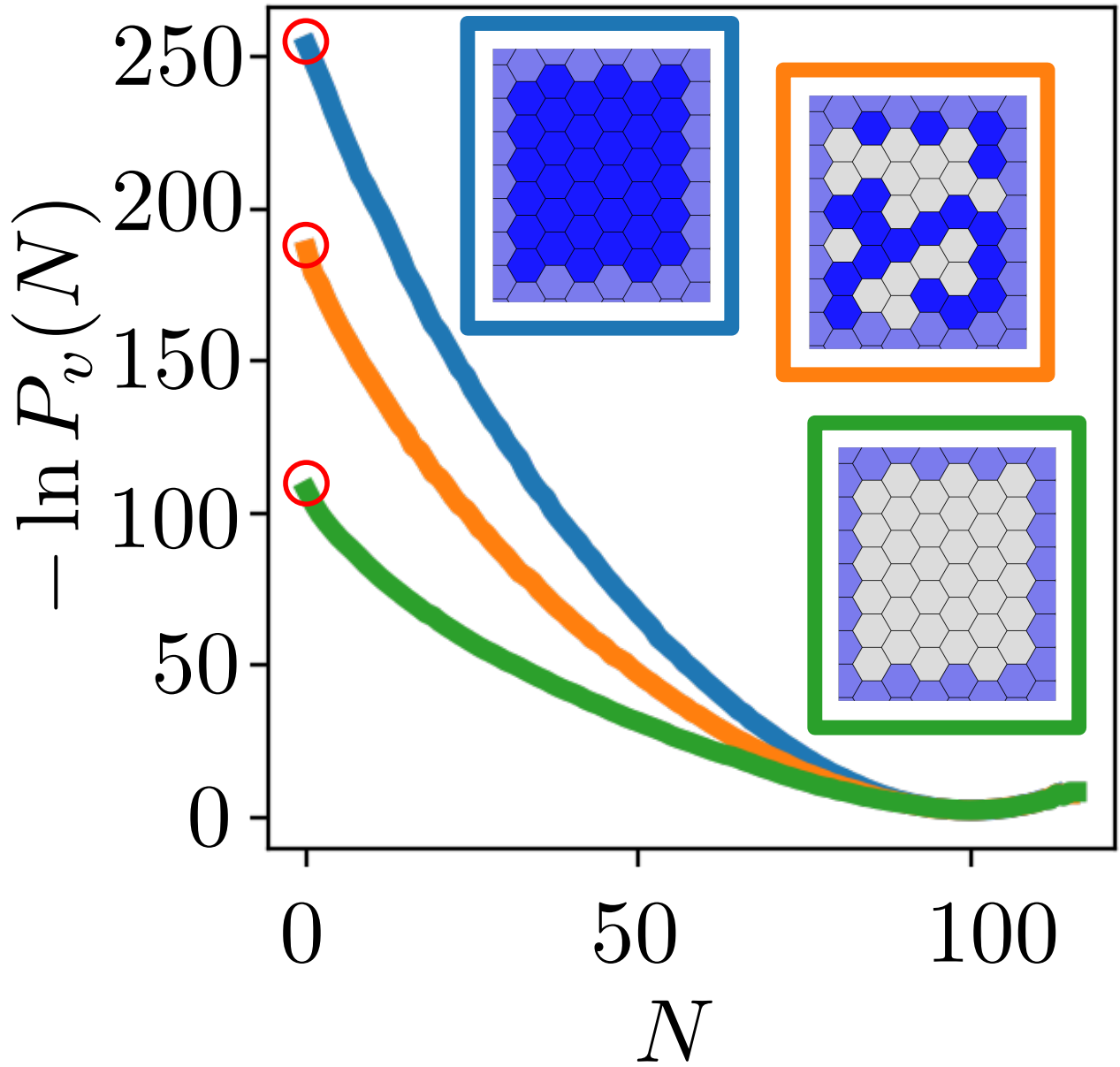
Fig. S4. The free energetics, $-\ln P_v(N)$, of observing N waters in a probe volumes, v, adjacent to polar (blue), nonpolar (green) and mixed (orange) 6×6 patches are shown (in units of $k_{\mathrm{B}}T$), and the free energy of cavity creation, $-\ln P_v(0)$, which serves to quantify patch hydrophobicity, f, is highlighted (red circles).
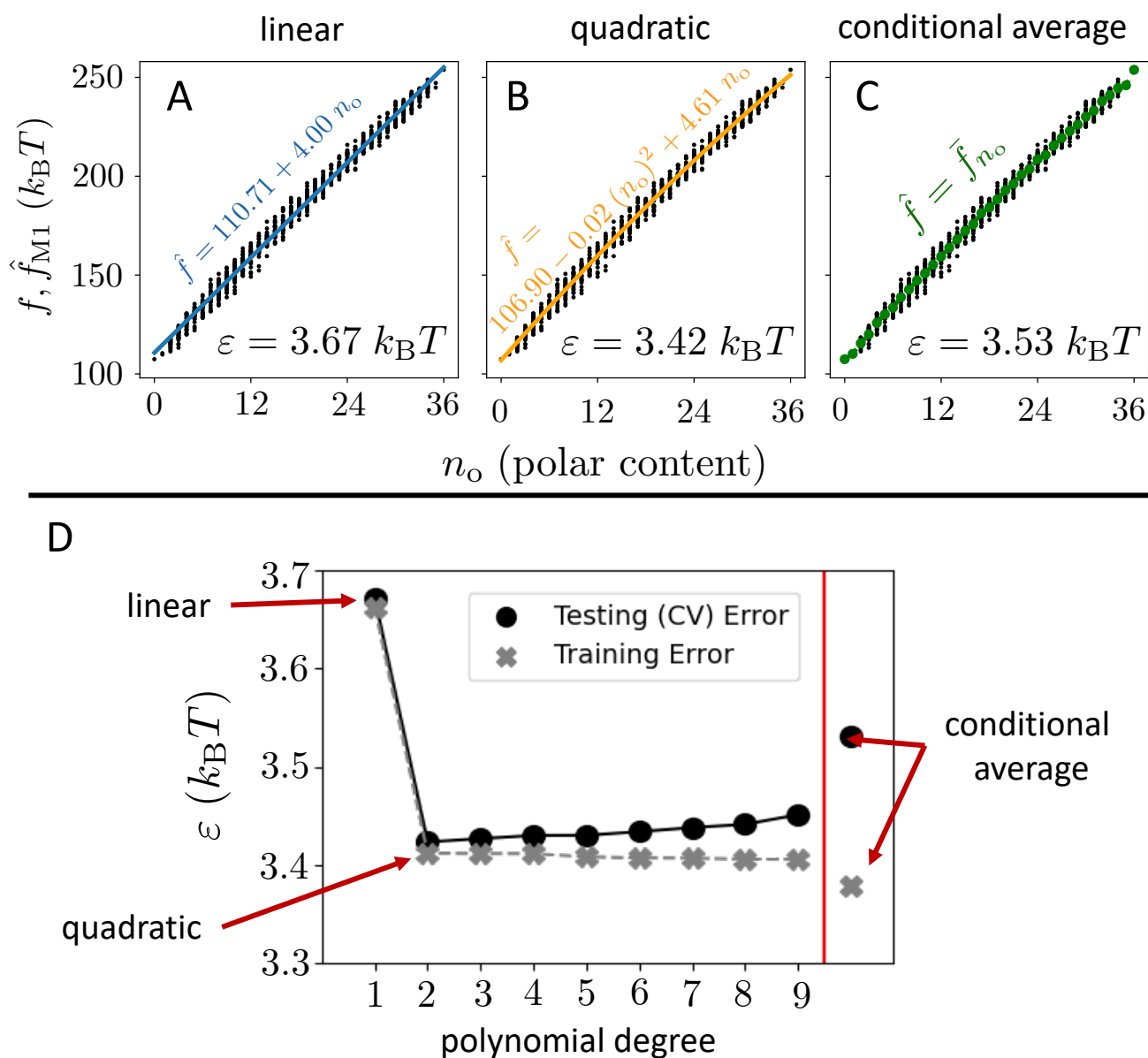
Fig. S5. Non-linear functions of polar content, n_o, do not perform significantly better than model $M1$, which is linear in n_o. (A-C) The hydrophobicity, f, of every patch in the library is plotted as a function of the polar content, n_o, of the patch, and the data is fit to (A) linear, (B) quadratic and (C) conditional average models. The cross-validated errors, ε, are also shown for each model. Although the models capture the trend of hydrophilicity increasing with polar content, they fail to capture the variation in f for patches with the same n_o; therefore, the non-linear models perform only marginally better than the linear model $M1$. (D) The testing (black) and total in-sample training (gray) error of different one-feature (n_o) models are shown. As expected, increasing model complexity (e.g., polynomial degree) results in a decrease in the in-sample training error, but the testing error increases as more complex models overfit the available data. In fact, the conditional average model, which provides the best-possible in-sample training performance, displays as substantial gap between in-sample and testing performance, highlighting the susceptibility of complex models to over-fitting.

Nicholas B. Rego, Andrew L. Ferguson, and Amish J. Patel

Translational

Rotational

Mirror
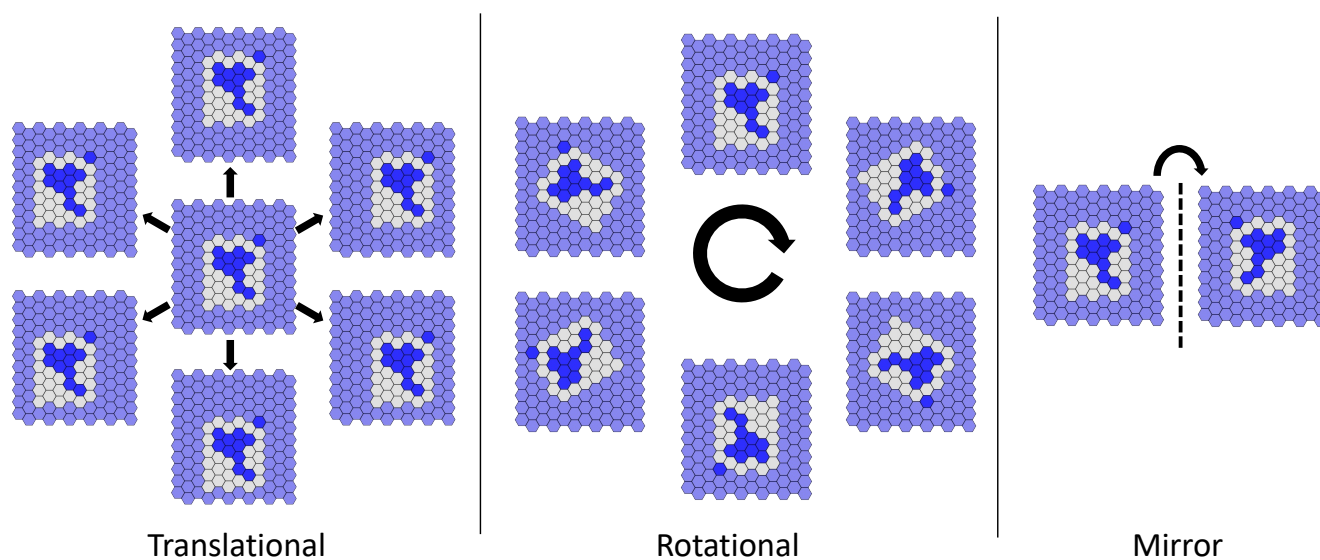
Fig. S6. Augmenting the training library of 6×6 patches with translated, rotated and reflected variants. To account for the invariance of patch hydrophobicity under translation, rotation and reflection, we augmented the training dataset with 83 isomorphic variants for every patch ($7 \times 6 \times 2$ for translational, rotational, and reflection factors, respectively); select variants are shown for a representative patch pattern. To accommodate the variants, every patch was embedded in a 12×11 SAM surface with a background of polar groups; polar end groups outside the patch are shown in light blue.
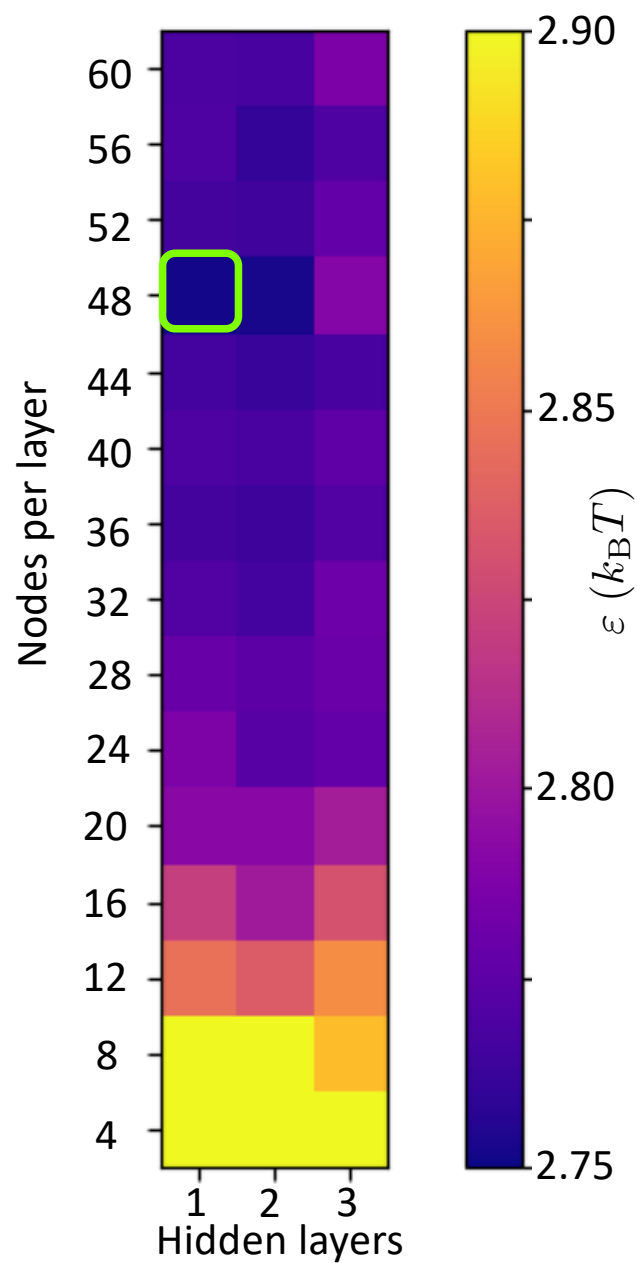
Fig. S7. Determining the optimal ANN architecture for model $\mathrm{M1A}$. A series of models with different numbers of hidden layers and nodes per hidden layer was systematically trained and the variation of model performance, ε, with the choice of hyper-parameters is shown. The model with the best performance ($\varepsilon = 2.75~k_{\mathrm{B}}T$) contains one hidden layer with 48 nodes and is highlighted in green.

Nicholas B. Rego, Andrew L. Ferguson, and Amish J. Patel
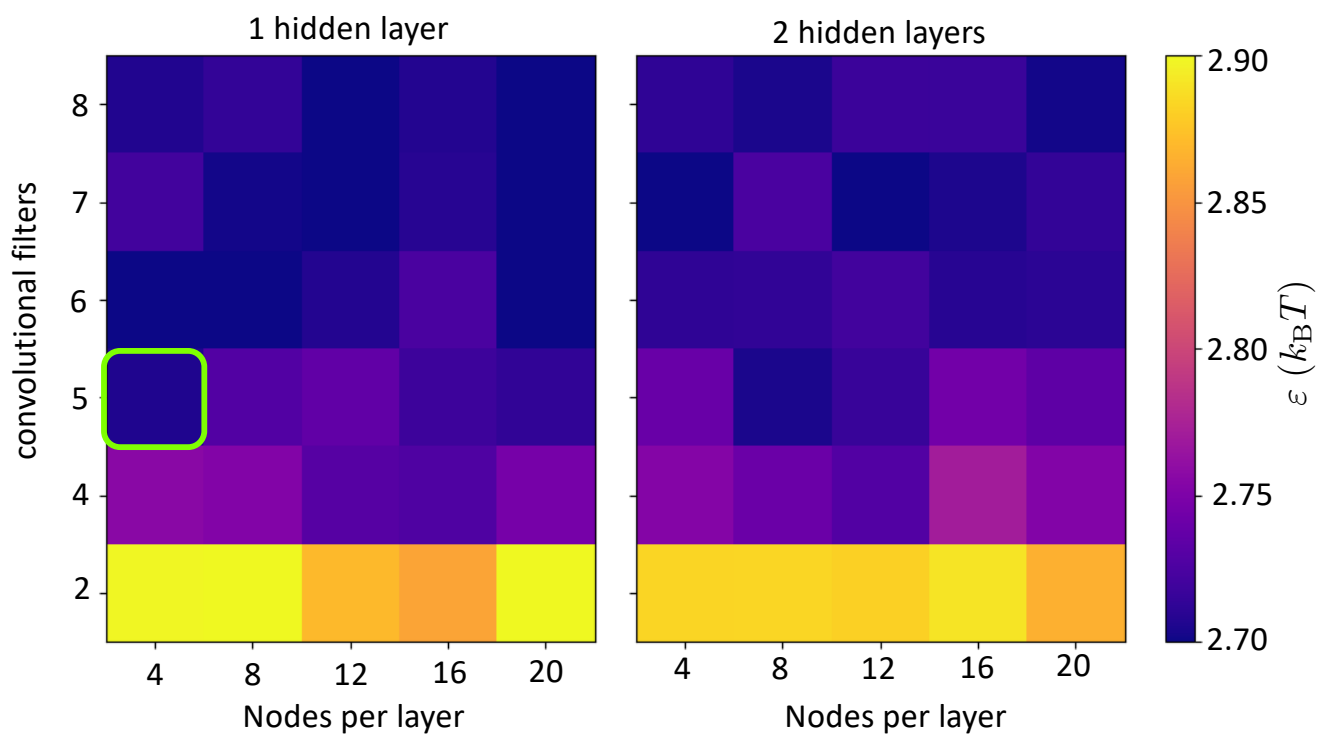
Fig. S8. Determining the optimal CNN architecture for model M1C. A series of models with different numbers of convolutional filters, hidden layers and nodes per hidden layer were systematically trained and the variation of model performance, ε, with the choice of hyper-parameters is shown. The CNN architecture with 5 convolutional filters and 1 hidden ANN layer with 4 hidden nodes (green rectangle) was chosen because it gave rise to the simplest model with high performance.
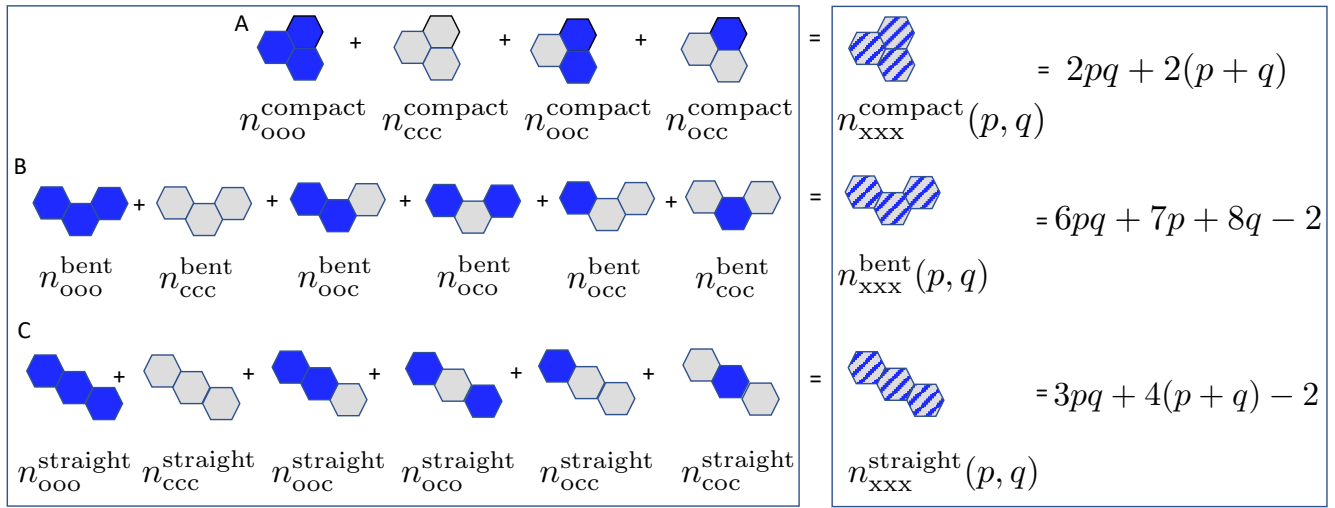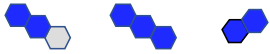
Fig. S9. Enumerating the 16 three-body features (left box). The total number of (A) 'compact', (B) 'bent', or (C) 'straight' three-body features must be constant ($n_{\mathrm{xxx}}^{\mathrm{compact}}$, $n_{\mathrm{xxx}}^{\mathrm{bent}}$, or $n_{\mathrm{xxx}}^{\mathrm{straight}}$, respectively) for patches with specific dimensions, p and q. Using these constraints, we eliminate the following three features: $n_{\mathrm{ccc}}^{\mathrm{compact}}$, $n_{\mathrm{ccc}}^{\mathrm{bent}}$ and $n_{\mathrm{ccc}}^{\mathrm{straight}}$.

Nicholas B. Rego, Andrew L. Ferguson, and Amish J. Patel

(i)

$$n_{\text{ooc}}^{\text{compact}} + 3n_{\text{ooo}}^{\text{compact}} - 2n_{\text{oo}} = 2(p+q) + 2$$

(ii)

$$n_{\text{ooc}}^{\text{bent}} + 2n_{\text{ooo}}^{\text{bent}} - 4n_{\text{oo}} = 6p + 8q$$

(iii)

$$n_{\text{ooc}}^{\text{straight}} + 2n_{\text{ooo}}^{\text{straight}} - 2n_{\text{oo}} = 4(p+q) - 2$$

(iv)

$$n_{\text{occ}}^{\text{compact}} - 3n_{\text{ooo}}^{\text{compact}} + 4n_{\text{oo}} - 6n_{\text{o}} = 2(p+q) - 4$$

(v)

$$n_{\text{occ}}^{\text{bent}} + 2n_{\text{oco}}^{\text{bent}} + 4n_{\text{oo}} - 12n_{\text{o}} = 8(p+q) - 4$$

(vi)

$$n_{\text{occ}}^{\text{straight}} + 2n_{\text{oco}}^{\text{straight}} + 2n_{\text{oo}} - 6n_{\text{o}} = 4(p+q) - 2$$

(vii)

$$n_{\text{coc}}^{\text{bent}} - n_{\text{ooo}}^{\text{bent}} + 4n_{\text{oo}} - 6n_{\text{o}} = p - 2$$

(viii)

$$n_{\text{coc}}^{\text{straight}} - n_{\text{ooo}}^{\text{straight}} + 2n_{\text{oo}} - 3n_{\text{o}} = 0$$

Fig. S10. Stoichiometric and geometric constraints relating three-body features to n_{o} and n_{oo} as well as the patch dimensions, p and q. The eight constraint equations, listed here, enable elimination of eight features in favor of the remaining features, n_{o}, n_{oo}, p and q. For each constraint equation, the first feature on the left-hand side is eliminated.
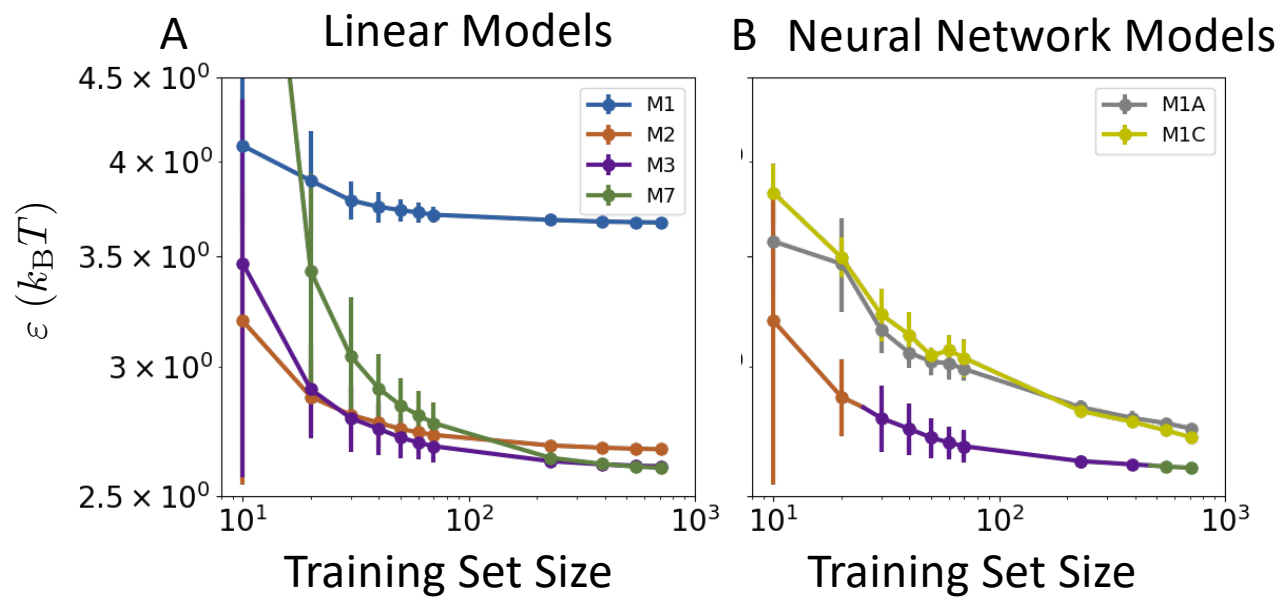
Fig. S11. Dependence of model performance on the size of the training dataset. Learning curves for the (A) linear and (B) neural network models highlight how model error, ε, varies with the size of the training dataset. The performance of the linear models plateaus as the training set size is increased, whereas that of the neural network models continues to improve over the entire range of training set sizes. The orange, purple, and green curve (corresponding to models M2, M3 and M7, respectively) in panel B corresponds to the convex hull of the best-performing models for each training set size.
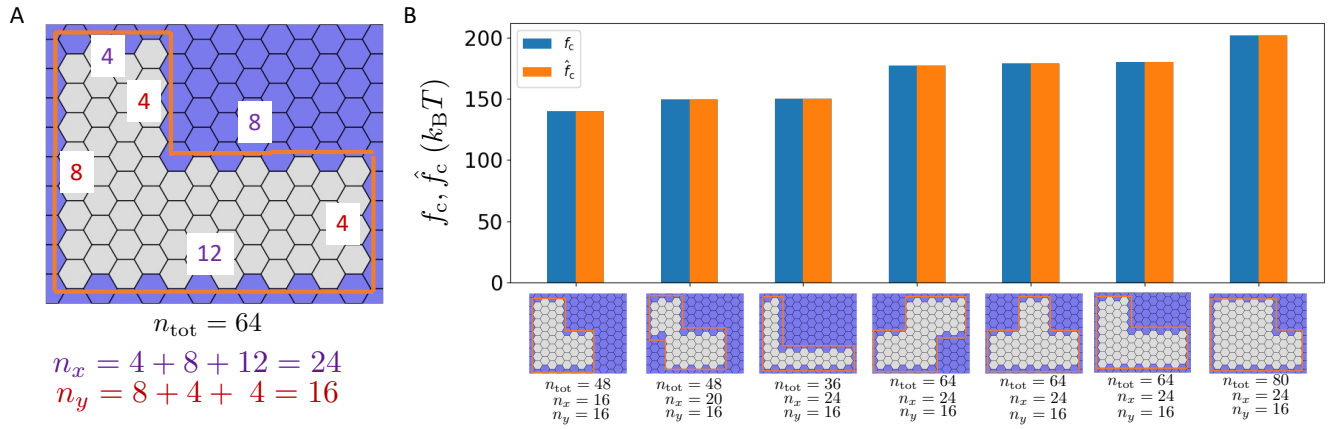
Nicholas B. Rego, Andrew L. Ferguson, and Amish J. Patel

Fig. S12. Predicting the hydrophobicity of uniform nonpolar patches with arbitrary (non-rectangular) shapes. (A) To generalize $\hat{f}_c$ to non-rectangular patches, we recast the model in terms of the total number of patch end groups, n_{tot}, and the number of peripheral end groups, n_x and n_y, in the x and y dimensions, respectively. The values of n_{tot}, n_x (purple segments), and n_y (brown segments) are shown for a sample non-rectangular nonpolar patch (orange). (B) To test the performance of the generalized model, $\hat{f}_c(n_{\text{tot}}, n_x, n_y)$, we estimated the hydrophobicity, f_c, of 7 non-rectangular nonpolar patches (gray, outlined in orange), and found that the model accurately predicts f_c ($\varepsilon = 4.75\ k_{\text{B}}T$); the values of n_{tot}, n_x, and n_y for each patch are also shown.
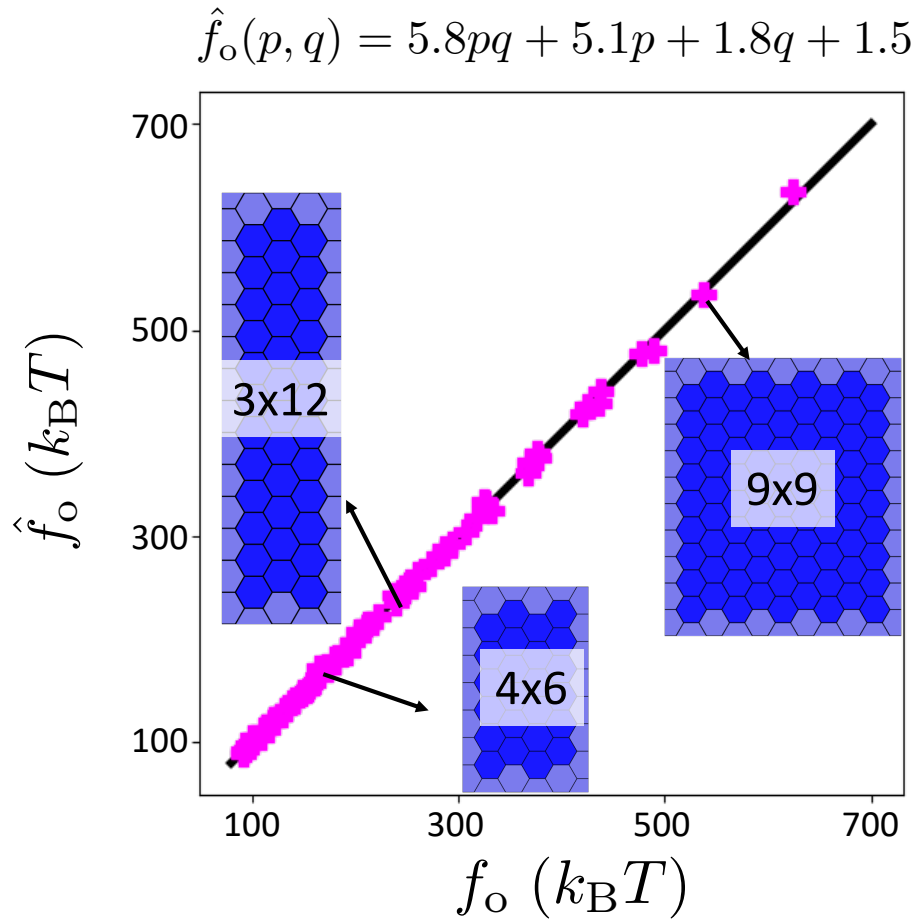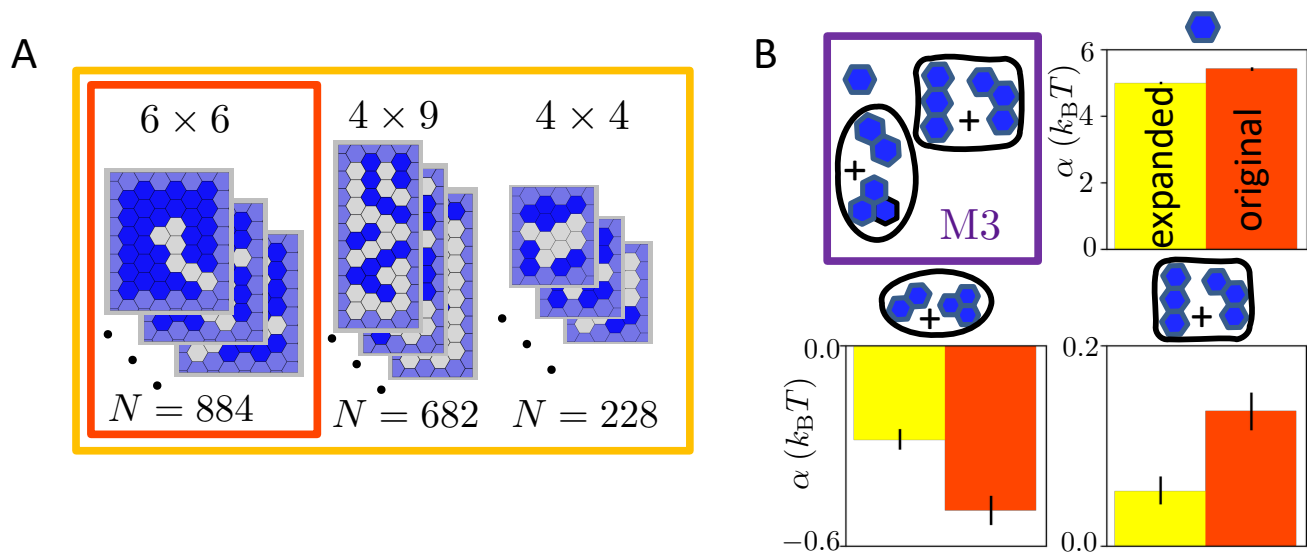
Fig. S13. Predicting the hydrophobicity of uniform polar patches, f_{o}, as a function of their dimensions, p and q. We estimate f_{o} for 62 patches of different sizes and shapes, and use this dataset to train a model $\hat{f}_{\mathrm{o}}(p,q)$ that is linear in pq, p and q. The resulting model predictions, $\hat{f}_{\mathrm{o}}$, plotted against the estimated f_{o}-values, illustrate the performance of the model ($\varepsilon = 4.02\ k_{\mathrm{B}}T$).

Nicholas B. Rego, Andrew L. Ferguson, and Amish J. Patel

Fig. S14. Generalizing model M3 to rectangular patches of different sizes and shapes. (A) In addition to the $N = 884$ patches (6×6) in the original dataset (red box), the expanded dataset (yellow box) also contains $N = 682$ patches of a different shape (4×9) and $N = 228$ patches of a different size (4×4). (B) The coefficients obtained by training model M3 on the original and expanded datasets agree reasonably well.

Fig. S15. The influence of chemical patterning on hydrophobicity for different patch sizes. Two patches with different chemical patterns (clustered and dispersed), but the same number of polar groups (equal to half of the total groups), display substantial differences in their hydrophobicity, f, for a range of patch sizes (from 6×6 to 12×12). Because the dispersed and clustered patterns have the same n_o, the additive model M1 would predict that the difference in their hydrophobicities, $f^{\mathrm{disp}} - f^{\mathrm{clust}}$, should be equal to 0 regardless of patch size. In contrast, the extended models M2 and M3 correctly predict that patches with well-dispersed end groups are more hydrophilic than those with clustered end groups, and they capture the differences in hydrophobicity with chemical patterning reasonably well across the different patch sizes.
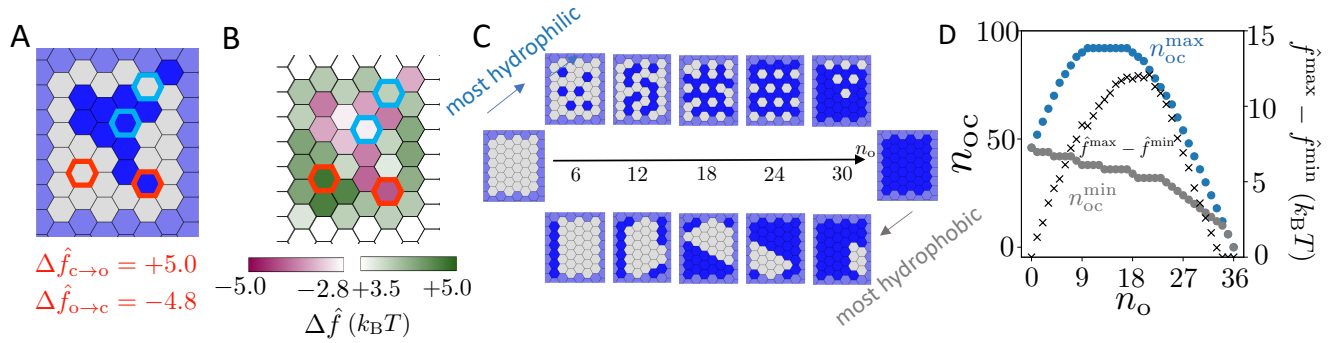
Nicholas B. Rego, Andrew L. Ferguson, and Amish J. Patel

Fig. S16. Using model M3 for the rational design of patterned SAM patches with maximal hydrophobicity/hydrophilicity. (A) For the patch pattern shown, an exhaustive screening of all possible point mutations is performed using model M3. (B) The changes in patch hydrophobicity, $\Delta \hat{f}$, in response to point mutations, enables identification of 'hot-spot' locations (red outline). Mutations that result in the smallest absolute change in patch hydrophobicity are also highlighted (cyan outline). (C) The maximally hydrophilic (top) and hydrophobic (bottom) patches with a particular polar content, n_{o}, were uncovered using an iterative, greedy design protocol; the most hydrophilic patches tend to disperse their polar end groups, whereas the most hydrophobic patches cluster their nonpolar end groups. (D) To quantify the relative dispersion of end groups in the optimal patches, we plot the number of polar – nonpolar neighbors, n_{oc}, for the most hydrophilic ($n_{\mathrm{oc}}^{\max}$, blue) and the most hydrophobic ($n_{\mathrm{oc}}^{\min}$, gray) patches as a function of polar content, n_{o}. We also plot the difference, $\Delta n_{\mathrm{oc}} \equiv n_{\mathrm{oc}}^{\max} - n_{\mathrm{oc}}^{\min}$.
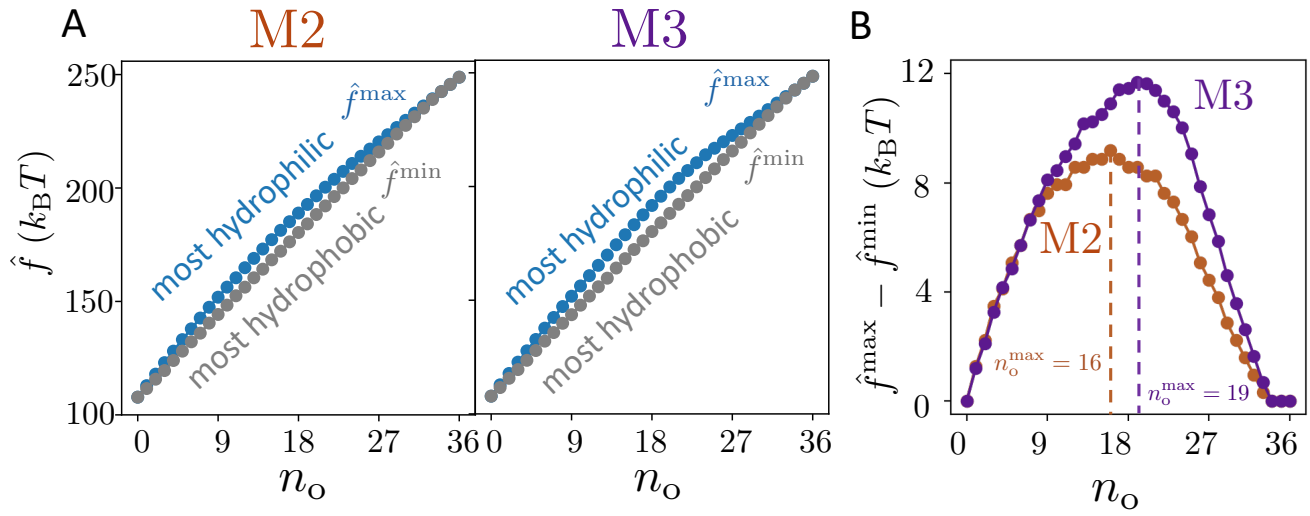
Fig. S17. (A) The hydrophobicities of the most hydrophilic ($\hat{f}^{\mathrm{max}}$, blue) and the most hydrophobic ($\hat{f}^{\mathrm{min}}$) patches, uncovered using the iterative greedy design protocol with either model M2 (left) or model M3 (right), are shown as a function of patch polar content, n_{o}. (B) The difference, $\hat{f}^{\mathrm{max}} - \hat{f}^{\mathrm{min}}$, corresponds to the range of hydrophobicities attainable by patches with given polar content, n_{o}, and is shown for the optimal patches obtained using models M2 (orange) and M3 (purple); the models predict that this difference is greatest for patches with roughly 50% polar content.

Nicholas B. Rego, Andrew L. Ferguson, and Amish J. Patel

References

1. F Wang and DP Landau. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.*, 86(10):2050–2053, 2001.

2. N Shenogina, R Godawat, P Keblinski, and S Garde. How wetting and adhesion affect thermal conductance of a range of hydrophobic to hydrophilic aqueous interfaces. *Phys. Rev. Lett.*, 102(15):156101, 2009.

3. R Godawat, SN Jamadagni, and S Garde. Characterizing hydrophobicity of interfaces by using cavity formation, solute binding, and water correlations. *Proc. Natl. Acad. Sci. U.S.A.*, 106(36):15119–15124, 2009.

4. E Xi, V Venkateshwaran, L Li, N Rego, AJ Patel, and S Garde. Hydrophobicity of proteins and nanostructured solutes is governed by topographical and chemical context. *Proc. Natl. Acad. Sci. U.S.A.*, 114(51):13345–13350, 2017.

5. B Hess, C Kutzner, D van der Spoel, and E Lindahl. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, 4(3):435–447, 2008.

6. RW Hockney, SP Goel, and JW Eastwood. Quiet high-resolution computer models of a plasma. *J. Comput. Phys.*, 14(2):148–158, 1974.

7. G Bussi, D Donadio, and M Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126(1):014101, 2007.

8. HJC Berendsen, JR Grigera, and TP Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, 91(24):6269–6271, 1987.

9. M Mondello, GS Grest, EB Webb, and P Peczak. Dynamics of n-alkanes: Comparison to Rouse model. *J. Chem. Phys.*, 109(2):798–805, 1998.

10. V Hornak, R Abel, A Okur, B Strockbine, A Roitberg, and C Simmerling. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*, 65(3):712–725, 2006.

11. WL Jorgensen, DS Maxwell, and J Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, 118(45):11225–11236, 1996.

12. T Darden, D York, and L Pedersen. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089–10092, 1993.

13. B Hess, H Bekker, HJ Berendsen, and JGEM Fraaije. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.*, 18(12):1463–1472, 1997.

14. S Miyamoto and PA Kollman. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.*, 13(8):952–962, 1992.

15. AJ Patel, P Varilly, and D Chandler. Fluctuations of water near extended hydrophobic and hydrophilic surfaces. *J Phys. Chem. B*, 114(4):1632–1637, 2010.

16. AJ Patel, P Varilly, D Chandler, and S Garde. Quantifying density fluctuations in volumes of all shapes and sizes using indirect umbrella sampling. *J. Stat. Phys.*, 145(2):265–275, 2011.

17. E Xi, SM Marks, S Fialoke, and AJ Patel. Sparse sampling of water density fluctuations near liquid-vapor coexistence. *Mol. Simul.*, 44(13-14):1124–1135, 2018.

18. F Pedregosa, G Varoquaux, A Gramfort, V Michel, B Thirion, O Grisel, M Blondel, P Prettenhofer, R Weiss, V Dubourg, J Vanderplas, Alexandre Passos, D Cournapeau, M Brucher, M Perrot, and É Duchesnay. Scikit-learn: Machine learning in Python. *J. Mach. Learn Res.*, 12(85):2825–2830, 2011.

19. A Paszke, S Gross, F Massa, A Lerer, J Bradbury, G Chanan, T Killeen, Z Lin, N Gimelshein, L Antiga, A Desmaison, A Kopf, E Yang, Z DeVito, M Raison, A Tejani, S Chilamkurthy, B Steiner, L Fang, J Bai, and S Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

20. DP Kingma and J Ba. Adam: A method for stochastic optimization. *arXiv:1412.6980 [cs]*, 2017.

21. C Steppa and TL Holch. HexagDLy—Processing hexagonally sampled data with CNNs in PyTorch. *SoftwareX*, 9:193–198, 2019.

22. MT Ribeiro, S Singh, and C Guestrin. "Why Should I Trust You?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 1135–1144, New York, NY, USA, August 2016. Association for Computing Machinery. ISBN 978-1-4503-4232-2. URL https://doi.org/10.1145/2939672.2939778.