Examining Chemistry-Property Relationships in Lubricating Monolayer Films through Molecular Dynamics Screening

Andrew Z. Summers†,‡, *Justin B. Gilmer*†,‡, *Christopher R. Iacovella*†,‡, *Peter T. Cummings*†,‡, *and Clare M^{c}Cabe**,†,‡,§

†Department of Chemical and Biomolecular Engineering, ‡Multiscale Modeling and Simulation (MuMS) Center, and §Department of Chemistry, Vanderbilt University, Nashville, Tennessee 37235, United States

ABSTRACT

The tribological effectiveness of chemisorbed alkane-based monolayers is closely related to their chemical composition. Molecular simulation is a useful tool for studying these relationships, by providing precise control over system variables and molecular-level resolution, yet the lack of robust initialization tools has limited the ability of previous studies to perform comprehensive structure-property screening of monolayers over a wide chemical parameter space. Herein, we leverage the recently developed Molecular Simulation and Design Framework (MoSDeF) to perform molecular dynamics screening of functionalized monolayer films, in particular focusing on systems where the two films have different terminal group chemistries. This screening reveals that systems where one film is hydrophobic yield favorable adhesion and identifies several film

chemistries that also simultaneously show low coefficients of friction. The breadth of the data obtained from this screening is sufficient to utilize machine learning to develop predictive models for the tribology of functionalized monolayer films and to extract information on terminal group characteristics that most influence tribology. It is observed that molecular shape most strongly influences the coefficient of friction, whereby linear and planar molecules are preferable, while charge distribution is most closely linked to adhesion. The workflows, insights, and models developed in this work lay the groundwork for future screening of more complex films.

INTRODUCTION

The design space of nanoscale devices has been constrained by the friction and wear associated with surfaces in sliding contact, with conventional lubrication schemes proven ineffective. Self-assembled monolayer films have shown promise as a potential solution to these issues, providing a dense layer of surface-bound chains that prevent direct surface-surface contact and reduce both adhesive and frictional forces. A favorable characteristic of monolayer lubricants is their highly-tunable chemistry, which can be altered (1) at the level of the chain (e.g. chain length, internal and terminal functional groups), (2) by mixing different chain chemistries within a single film (i.e. multi-component monolayers), and (3) by functionalizing opposing surfaces with monolayers of different chemistries. This presents a near-infinite chemical parameter space that can be utilized to optimize monolayer chemistry for tribological applications, and considerable efforts have been made in kind to understand how chemical changes influence frictional response. For example, it has been observed that increasing the backbone chain length yields reduced frictional forces[1–4] (although it should be noted that some studies have observed a friction force minimum at intermediate chain lengths),[5] attributed to an increase in attractive inter-chain van der Waals

(VDW) forces with longer backbones, resulting in highly ordered films. Phenyl-terminated monolayer films have been observed to yield higher frictional forces than methyl-terminated films, attributed to the presence of additional energy dissipation modes through the twisting of phenyl groups during shear.[6] Additionally, monolayers terminated with fluorinated moieties have shown both increased[7] and decreased[8] frictional forces compared with standard hydrogenated monolayers. The presence of hydroxyl (OH) and carboxyl (COOH) terminal moieties has been shown to lead to increased frictional and adhesive forces, attributed to the formation of hydrogen bonds between the two contacting interfaces.[3,9–11] Simulations of mixed alkane and perfluoroalkane monolayers demonstrated reduced COF compared to either equivalent pure-component system.[8] The work of Brewer et al., in examining the effects of contacting two chemically-dissimilar monolayer films, observed that a CH_3-functionalized friction force microscope tip in contact with OH- and COOH-terminated monolayers resulted in lower coefficients of friction (COFs) than similar CH_3-terminated monolayers.[3] As such, the literature has clearly demonstrated that the tribological properties are highly related the chemistry of monolayer films; however, despite the considerable attention that monolayer films have received over the past several decades, much of the vast chemical space remains unexplored.

Experimentally, synthesis of monolayer films can be non-trivial; examination of novel chemistries often requires changes to synthetic approaches, which can make it difficult to decouple the effects of chemical changes from the effects of other properties such as monolayer density. Furthermore, comparisons between different experimental studies can be challenging, if different synthesis protocols or techniques for measuring tribological properties are used. For example, two common techniques for experimental analysis of monolayer films, atomic force microscopy (AFM) and tribometry, feature probes with radii of curvature that differ by several orders of

magnitude, and it has been shown that even slight differences in probe shape can enable different mechanisms of energy dissipation in monolayer films.[12] As a result, molecular dynamics (MD) simulation has become a useful tool for probing the tribological properties of monolayers,[5,7,10,12–15] affording atomic-level resolution and direct control over system variables. MD can be utilized for systematic variation of monolayer chemistry, in order to identify both tribologically-favorable chemistries as well as chemistry-property relationships that can provide insight used for the design of more favorable monolayers. However, typical MD workflows often lend themselves better to screening over thermodynamic space (e.g., varying temperature or normal force) as opposed to chemical space (e.g., terminal group chemistry), with the majority of studies examining chemical influences on tribology restricting themselves to a comparison between only a small number of systems/chemistries. This can be associated, in part, with the lack of tools designed to enable screening; i.e., tools that provide the ability to systematically vary chemistry during model setup, to automatically apply force field parameters to these models, and to manage the execution of the simulation workflows for large numbers of systems. Instead, the current state of MD simulation typically involves system setup in an *ad-hoc* manner, often using in-house scripts and human manipulation, which makes large-scale screening studies intractable *and* limits reproducibility of published results. As a means to resolve these issues, the Molecular Simulation and Design Framework (MoSDeF)[16] has been designed as an open-source Python framework for generating molecular systems as objects with exchangeable chemical parts and for automatically applying force field parameters.

Here, we use the MoSDeF toolkit, along with the Signac framework[17] for workflow management, to enable large-scale, non-equilibrium MD screening simulations of lubricating monolayer films, whereby the exact inputs, procedures, and processes can be captured in a way

that make simulations TRUE: Transparent, Reproducible, Usable by others, and Extensible. In this work, focus is specifically placed on the monolayer chain length and terminal group chemistry, where 5 chain lengths and 16 unique terminal group chemistries are compared. In addition to chemically-identical films, 84 unique combinations of chemically-dissimilar films are considered, where the two contacting monolayers feature different terminal groups. In total, 164 unique system chemistries are examined. Both the coefficient of friction (COF) and the force of adhesion are examined for each system to evaluate tribological efficacy. The scope of the monolayer chemical space examined in this work is sufficient such that machine learning is utilized to derive quantitative structure-property relationship (QSPR) models for the coefficient of friction and force of adhesion, to aid in exploring links between chemistry and tribology. These models, which take as an input a SMILES[18] representation of a molecular terminal group, can also aid in the further screening of monolayer chemical space, by identifying chemistries that may have favorable tribological properties, as well as excluding those with potentially unfavorable tribological properties. While previous experimental and simulation studies have examined the effects of terminal group chemistry on friction, the work herein presents, to the best of our knowledge, the most comprehensive single study to date of these relationships. Furthemore, all simulations follow the same exact procedures and methods, allowing for direct cross comparison.

SIMULATION METHODS

Molecular Model

Systems in this work consist of two opposing monolayer films, attached to amorphous silica substrates, consistent with models used in earlier work.[14,15] The use of amorphous surfaces is notable as many studies of monolayer films have used crystalline surfaces,[5,8,19,20] where the in-

plane ordering of attachment sites has been shown to exaggerate chain orientational order.[14] Our procedure for carving the amorphous surfaces builds upon procedures found in the literature[21,22] and used in prior work[12] and is detailed in the Supporting Information. This procedure results in surfaces with an atomic-scale roughness of ~0.11 nm, found to closely match prior work that explicitly considered surface oxidation treatment typical of experiment;[15] most prior studies that have used amorphous silica substrates have considered atomically smooth surfaces, which earlier work suggests influences the behavior and trends.[12,14,15]

The mBuild Python package,[23,24] a component of the MoSDeF toolkit designed for molecule building, was used to construct the dual-monolayer systems. Systems are constructed in a hierarchical manner, whereby (1) a prototype for each chain is constructed, (2) duplicates are attached to assigned sites on a silica surface (where the specific attachment sites can be randomized), and (3) the monolayer is duplicated, rotated by 180 degrees, and shifted to yield the complete system. Chain prototypes are generated such that both the backbone chain length (excluding the terminal group) and the functional group used to terminate the chain are tunable. The approach, shown graphically in Figure 1a, is designed such that both the backbone length and terminal group chemistry can be trivially modified, making these scripts easily extensible to other chemistries not considered here. Furthermore, due to the hierarchical, component-based framework of mBuild, these routines can modified to allow for screening over additional chemical degrees of freedom, such as backbone monomer, multi-component films, etc.

The pool of terminal group chemistries examined in this work is shown in Fig. 1a and was constructed to satisfy two primary criteria. Namely that the functional group should be able to be described using existing parameters within the OPLS force field (described in further detail below) and that the entire functional group pool should span a wide range of chemical characteristics (e.g.

size, shape, polarity) to facilitate structure/property analysis. It is worth noting that several of the chosen functional groups may not be synthesizable, or would readily react following synthesis.[25] However, as reactions are not considered in the simulations, given that classical force fields with fixed bonds are used, such groups remain stable. Furthermore, as our goal is simply to study a chemically-diverse range of terminal group chemistries, such systems still provide valuable data for the development of property-prediction models.

Screening of dual-monolayer systems was performed over two distinct parameter spaces: Chemically-identical systems (Fig. 1b, where the top and bottom monolayer films feature the same chemistry) and chemically-dissimilar systems (Fig. 1c, where the top and bottom monolayer films feature different chemistries). The motivation for including chemically-dissimilar systems in this study was two-fold. First, in the context of performing a structure-property analysis between terminal group chemistry and monolayer tribology, the inclusion of chemically-dissimilar systems provides increased sampling of unique interfacial chemistries without the need to expand the pool of terminal group chemistries. This provides a significant advantage, as it reduces the number of unique force field parameters necessary to perform the study. Second, the motivation for examining chemically-dissimilar systems concerns findings in the literature of favorable tribological properties for such systems, outperforming their chemically-identical counterparts.[3,26–29]

For each chemistry, five systems were generated, each corresponding to a unique arrangement of chains on the silica surface. Structural and tribological properties were evaluated for each of the five replicas and averaged to obtain values for each chemistry that are independent of chain arrangement. For systems featuring chemically-identical films, all 16 terminal group chemistries from Fig. 1a were considered along with five chain lengths (5, 8, 11, 14, and 17 backbone carbons,

excluding the terminal group), equaling 80 unique chemistries and 400 systems in total. Chemically-dissimilar films feature backbone chain lengths of 17 carbons and include all combinations of seven select terminal groups (carboxyl, fluorophenyl, hydroxyl, isopropyl, methyl, nitro, and perfluoromethyl) with the 16 terminal groups in Fig. 1a, leading to an additional 84 unique chemistries (after removing duplicates) and 420 total systems. In sum, 820 systems were therefore examined, with 164 unique monolayer chemistries. For the development of machine learning models, a set of 100 unique terminal group combinations (both chemically identical and dissimilar) with a chain back length of 17 carbons are considered.

Force field

After construction of the monolayer structures, the systems are atom-typed, i.e., the appropriate force field parameters are determined for the bonded and non-bonded interactions of the system. Atom-typing is handled by the Foyer library[30,63] as part of MoSDeF. In Foyer, force field parameters and their usage rules are encoded within a single XML-formatted file that currently builds upon the OpenMM[31] XML force field file format. Usage rules are encoded using SMARTS[32], along with "overrides" statements to set rule precedents, making these definitions both human and machine readable, and contained within a single unambiguous format. In Foyer, each molecular model is treated as a graph, and atom-types are determined by matching chemical environments to the patterns defined by the SMARTS usage rules. This file format also provides digital object identifiers (DOIs) corresponding to the source of the parameters, to aid in reproducibility. Further details on Foyer can be found in reference 63 and online.[33]

The OPLS all-atom force field has been employed in this work for all systems examined,[34–40] and tables containing all parameters used can be found in the Supporting Information and found,

along with usage rules, online in the Foyer compatible XML file format. The specific OPLS version used was that provided with the Gromacs 5.1 distribution,[41] in addition to parameters for silica that were obtained from Lorenz et al.[13] It should be noted that for several systems, certain parameters were not available (most commonly dihedrals including either alpha, beta, or gamma carbons neighboring the terminal group). In these cases, additional parameters were introduced according to sensible conventions; for example, in the case of missing dihedrals including alpha, beta, or gamma carbons, standard C-C-C-C alkane dihedrals were used, as this was found to be the convention within the rest of the OPLS force field.

Molecular Dynamics Simulations

Molecular dynamics (MD) simulations were performed using the GROMACS molecular dynamics engine (version 5.1.0).[41] An initial, short, distance-limited NVE simulation was performed using the LAMMPS molecular dynamics engine[42,43] (as this routine was not available within Gromacs) to remove overlaps between terminal groups from the initial configuration. After converting the final structure from the initial LAMMPS simulation to Gromacs format, simulations were executed in four stages: (1) energy minimization, (2) equilibration, (3) compression, and (4) shear. Energy minimization was performed using a steepest descent algorithm. Following energy minimization, equilibrium molecular dynamics simulations were performed, whereby monolayers and surfaces (excluding the outer 4Å of each surface) were allowed to relax over 1ns to reach a low-energy state, using a timestep of 1fs. Following equilibration, a force of 5nN was applied to the bottom surface in the *z* direction (i.e., towards the top monolayer) to bring the two monolayers into contact, where they were compressed over 0.5ns, allowing the inter-surface distance to reach a steady-state value. Beginning from snapshots at the end of the compression stage, three independent

simulations were performed where monolayers were sheared under normal loads of 5, 15, and 25nN (corresponding to pressures of 200, 600, and 1000MPa, respectively), consistent with normal loads used for such systems in the literature.[7,44,45] To maintain a constant normal load a constant force was applied to the bottom silica surface in the *z* direction. Shear was introduced by coupling the top surface to a ghost particle via a harmonic spring (with a spring constant of 10,000 $\mathrm{kJ/(mol \cdot nm^2)}$) and pulling the ghost particle in the *x* direction at a rate of 10 m/s. Shear was performed for 10ns, whereby over the first 5ns monolayers reach a steady-state configuration where chains tend to align in the direction of shear, and the final 5ns was used for sampling. All MD simulations were performed under the NVT ensemble using a Nose-Hoover thermostat[46,47] to maintain a system temperature of 298.15 K. Previous work using a thermostat in the *x* and *y* dimensions only (i.e., not in the shear dimension) revealed no shear-induced heating,[14] agreeing with reports in the literature for similar systems,[44] thus application of a standard thermostat is justified. Bonds with hydrogen atoms were constrained using the LINCS algorithm[48], removing high-frequency atomic motions, and affording a time step of 2 fs to be used for the compression and shear stages. The particle-mesh Ewald method[49,50] was used for long-range electrostatics, using a force and pressure correction in the *z*-dimension to support slab geometries.

The MD workflow was maintained and executed using the Signac workflow manager (v0.5.4), a part of the Signac framework.[17] With Signac-flow, individual molecular dynamics operations (e.g., equilibration) were wrapped into Python functions. These operations were then performed on each state point (i.e., each system), with Signac-flow keeping track of all operations that had been performed on each state point. All scripts used to execute simulations and analysis are freely available on GitHub.

Tribological and Structural Analysis Methods

Monolayer tribology is evaluated via calculation of the coefficient of friction (COF) and adhesive force as defined by the Derjaguin form of Amontons' Law of Friction[26]:

$$F_f = F_0 + \mu \cdot F_n \quad (1)$$

where F_f, μ, and F_n represent the friction force, COF, and normal force, respectively. F_0 represents the friction force between the two surfaces under zero normal load. This term is often referred to as the adhesion term, offset term, or Derjaguin offset. Here, we refer to this term interchangeably as the adhesion force or force of adhesion. For each system, shear simulations are performed under a series of normal loads (5, 15, and 25nN), and the average friction force is measured for each simulation, allowing COF and the force of adhesion to be determined from Eq. 1 through linear regression. Values reported for each interfacial chemistry (i.e., unique combination of terminal groups) represent the average of the five systems with different chain configurations on the surface.

In addition to tribological metrics, monolayer structure is also examined through use of a nematic order parameter,[12,14,15] which provides a measure of the global orientational order of the monolayer chains. Briefly, a director vector is obtained for each monolayer chain by taking the eigenvector associated with the lowest eigenvalue of the chain's moment of inertia tensor. The nematic order parameter can then be calculated from:

$$S_2 = \langle \frac{3}{2} \cos^2 \theta - \frac{1}{2} \rangle \quad (2)$$

where θ is the angle between a chain's director vector and the monolayer director (i.e., the average of all chain directors) and the angle brackets represent an ensemble average. An S_2 value of 1 indicates perfect ordering (i.e., all chains are aligned), while deviations towards 0 indicate increased disorder.

As with the MD workflow, the management and execution of analysis routines is handled by the Signac (v0.8.5) and Signac-flow (v0.5.4) Python packages within the Signac framework.[17]

QSPR modeling

Monolayers are also examined via the development of topological quantitative structure-property relationship (QSPR) models relating characteristics of the monolayer chemistries to their tribological response. These models are given as an input "fingerprint" for each system, which represents each system as a series of numerical values. These numerical values are referred to as "molecular descriptors" and characterize a variety of molecular aspects such as size (e.g., approximate surface area), shape (e.g., asphericity), and charge distribution (e.g., topological polar surface area). Calculation of molecular descriptors is facilitated through use of the RDKit Python package.[51] While the majority of these descriptors are topological, meaning they are inferred from the molecular graph, several additional descriptors are considered that require a 3D molecular conformation. These conformations are generated by RDKit using the Experimental-Torsion Distance Geometry approach with "basic knowledge" terms (ETKDG), as proposed by Riniker and Landrum.[52] For descriptors requiring information concerning molecular charge, charges are assigned via RDKit to the molecular graph using the approach of Gasteiger and Marsili.[53] Along with the descriptors calculated via RDKit, we have introduced an additional descriptor, which we call the "hydrogen bond factor" that provides a relative estimate of the availability for inter-monolayer hydrogen bonding. The descriptor is calculated using counts of the number of hydrogen bond donor and acceptor atoms in the two terminal groups of a dual monolayer system via:

$$hb_{tb} = \begin{cases} 0, & \min\left(0, \max\left(d_t, a_b\right)\right) = 0 \\ \max\left(d_t, a_b\right), & \min\left(0, \max\left(d_t, a_b\right)\right) > 0 \end{cases}$$

$$hb_{bt} = \begin{cases} 0, & \min\left(0, \max\left(d_b, a_t\right)\right) = 0 \\ \max\left(d_b, a_t\right), & \min\left(0, \max\left(d_b, a_t\right)\right) > 0 \end{cases} \quad (3)$$

$$hbonds = hb_{tb} + hb_{bt}$$

where d and a represent the number of hydrogen donor and acceptor atoms present in terminal groups for the top (t) or bottom (b) monolayers, and hb_{tb} and hb_{bt} are intermediate quantities that estimate the ability of the top and bottom monolayers to act as the hydrogen donor, respectively. A complete list of all descriptors used, including brief descriptions of each, is provided in the Supporting Information.

The strategy used in this work for performing fingerprinting of the monolayer systems studied is shown in Fig. 2, where the primary challenge concerns how to properly design a system fingerprint that can be used for both chemically-identical and chemically-dissimilar systems (i.e., where the contributions of *both* terminal groups must be accounted for). First, as shown in Fig. 2a-d., "molecular fingerprints" (i.e., fingerprints for individual molecules rather than the complete dual monolayer system) are calculated for the terminal groups of each monolayer. In molecular fingerprinting, a decision must be made concerning the molecule used to represent each terminal group. Here, we utilize two separate molecules to represent terminal groups (as shown in Fig. 2b), calculating a subset of molecular descriptors on each. Specifically, molecules featuring terminal groups capped by hydrogen atoms (where the hydrogen atom is attached at the location where the terminal group is attached to the chain backbone in the complete chain) are used in the calculation of descriptors that relate to molecular shape, while molecules featuring terminal groups capped by methyl groups are used in the calculation of the remaining descriptors. The methyl group is chosen as this chemistry is comparable to that of the chain backbone, and thus these molecules should provide a reliable distribution of charge (as again, we use Gasteiger charges for fingerprinting as opposed to OPLS charges to simplify data inputs).

Our approach to molecular fingerprinting relies only upon information of the molecular bond graph as an input. As such, molecules are provided to RDKit in the notation of the simplified

molecular-input line-entry system (SMILES)[18], shown in Fig. 2c. This provides a concise input syntax that allows the QSPR models in this work to be easily extensible to additional terminal group chemistries beyond those considered here. After providing RDKit with SMILES representations of the two molecules representing each terminal group in a given system, molecular fingerprints are generated for each terminal group, shown in Fig. 2d. System fingerprints are then created from the two molecular fingerprints by using the *mean* and *minimum* values of each component (descriptor) of the molecular fingerprint (shown in Fig. 2e). While other combinations of these values (such as the *maximum* and *absolute difference*) were considered, analysis of correlations with the target variables (COF and adhesion) suggested only the mean and minimum values were needed, which also helps reduce complexity of the QSPR models.

A random forest regression algorithm[54,55] is used for QSPR models, as implemented in the scikit-learn Python package.[56] Random forest is an ensemble method that utilizes a forest of decision trees obtained from bootstrap sampling of the training data and produces a prediction based on the average values obtained for each tree. This algorithm is a popular choice in the literature for machine learning models; for example, Ballester et al. used random forest to predict binding affinities of protein-ligand complexes.[57] Predictions from random forest converge for a large number of decision trees[58], thus, a large number of 1000 trees are used in this work. One reason for the choice of the random forest algorithm is that contributions to the model from the various molecular descriptors can be easily extracted, and furthermore, interaction terms are implicitly included by the nature of the algorithm. Evaluation metrics for these models consist of the root-mean-square error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2) between the predicted and expected target values within both the training set and a 20% holdout

test (i.e., the model is developed using 80 randomly chosen systems, and tested with remaining 20). Additional model evaluation is accomplished via out-of-bag (OOB) sampling, whereby predictions are made on subsets of the data held out for each tree. The RMSE, MAE, and R^2 on the OOB estimates of each sample are also provided as a measure of model efficacy.

RESULTS AND DISCUSSION

Chemically-identical monolayer films

As a test of model accuracy and to serve as a baseline for the examination of more complex monolayers, relationships between chemistry and tribology are first examined between contacting monolayers with identical chemistries. The effect of backbone chain length is explored by calculating both tribological properties (COF and adhesion force) and structural properties (nematic order) for systems with chain lengths ranging from 5 to 17 backbone carbons, shown in Figure 3 for each terminal group chemistry. The effects of backbone chain length on friction have been explored extensively in the literature[1–5,59,60] and a comparison to these results helps verify that we find similar trends. It has been observed in several studies, primarily concerning methyl-terminated films, that the addition of carbons to the chain backbone correlates with a reduction in the COF.[1–4] Fig. 3 reveals this expected trend, where, in general, the COF is found to decrease as the backbone chain length is increased for each terminal group chemistry. It should be noted that this trend appears stronger for some systems than others, suggesting this may be coupled with the effects of terminal group chemistry. For example, benzene-terminated monolayers show little change in COF for chain lengths greater than 5, which may be related to the "bulkiness" of the terminal groups.

In the literature, the inverse relationship between COF and chain length has often been attributed to an increased ordering of monolayer chains,[1,61] due to an increase in favorable VDW forces between neighboring chain backbones. Here, monolayer order has been quantified via the nematic order parameter (Eq. 2), also shown in Fig. 3 as a function of chain length (note that in Fig. 3, the nematic order values shown are calculated for systems under an intermediate normal load of 15nN). The expected trend is observed for all systems, whereby monolayer nematic order is found to increase as the backbone chain length is increased. It is also observed that terminal group chemistry has diminishing effects on monolayer structure as backbone chain length increases. The range of nematic order values calculated for systems with a short backbone chain length of 5 carbons differs considerably between the various terminal group chemistries; however, systems with a backbone length of 17 carbons all feature a similarly high nematic order (>0.9). Interestingly, the systems where COF exhibits only a slight dependence on chain length also appear to have higher nematic order values at lower chain lengths. These systems correspond to those with "bulkier" terminal groups (e.g., fluorophenyl, nitrophenyl, benzne) for which steric interactions between the terminal groups likely promote more ordered films.

Fig. 3 also shows the effect of backbone chain length on the adhesive force between monolayers during shear, quantified via the Derjaguin offset, F_0 in Eq. 1. Adhesion is found to be less dependent on chain length than COF, and the range of the data comparing the various terminal group chemistries suggests that terminal group chemistry has a greater influence on adhesion than chain length. However, it does appear that for all systems adhesion either decreases or remains at roughly the same value as chain length is increased from 8 to 17 backbone carbons. However, the

results of adhesion combined with those for COF suggest that longer chain monolayers are optimal for lubrication, agreeing with results from the literature.[1–4]

The range of the data between the various systems shown in Fig. 3 suggests that terminal group chemistry influences both COF and adhesion. To provide a better visual comparison of this relation, Figure 4 displays the COF and adhesion force calculated for systems with a backbone chain length of 17 carbons (as the results of Fig. 3 have suggested this to be the most tribologically favorable chain length). Specific trends between terminal group properties and COF from the results of Fig. 4a are not immediately obvious. COF appears to be independent of the size of the terminal group, as large terminal groups are shown to yield both low (fluorophenyl) and high (cyclopropyl) COFs. Terminal group polarity also does not appear to influence COF, as polar and nonpolar groups are shown to yield both high and low values of COF. Thus, a direct comparison between systems appears insufficient at unearthing the specific terminal group characteristics that influence COF. Instead, a more comprehensive approach is required and addressed in this work via the development of QSPR models that also include dissimilar films in contact, discussed later.

In contrast to COF, trends between terminal group chemistry and adhesion, shown in Fig. 4b, are fairly clear. It appears there is a strong correlation between terminal group polarity and adhesion, as might be expected, as the terminal groups with the largest adhesion forces are all polar molecules and the groups with the smallest forces are all nonpolar. Furthermore, the three terminal groups yielding the highest adhesive forces are all groups that can participate in inter-monolayer hydrogen bonding, which again agrees with expectations from the literature.[10] In fact, the results shown in Fig. 4b further support the link between hydrogen bonding and adhesion, as the order of

the three systems with the highest adhesive forces (carboxyl, hydroxyl, and amino, from highest to lowest) corresponds with expected hydrogen bond strengths.

Chemically-dissimilar monolayer films

Although Fig. 4 provides some insight into the relationships between terminal group chemistry and tribological properties, such analysis still relies upon one's chemical intuition. Furthermore, no apparent trends are present between terminal group chemistry and COF, yet the range of the data between the various systems in Fig. 3, while smaller relative to that of adhesion, suggests that there is indeed a relationship. As a means to collect data for additional interfacial compositions, without the need to introduce new terminal group chemistries, a collection of simulations featuring systems of chemically-*dissimilar* monolayer films has also been performed (shown in Fig. 1c). Specifically, 420 chemically-dissimilar systems, all with a backbone chain length of 17 carbons (as this was shown to be the most favorable chain length for lubrication for chemically-similar monolayer films), are examined. The results of five different monolayer surface configurations are averaged for each interfacial chemistry corresponding to 84 unique interfacial chemistries. This expands the dataset, since for example a "hydroxyl-methyl" system should provide unique tribological properties distinct from either a "hydroxyl-hydroxyl" or "methyl-methyl" system. With the increased size of the data set, QSPR models can be developed to provide both greater insight into relationships between terminal group chemistry and tribology (through extraction of the relative contributions of the various components of the QSPR models), along with predictive capacity for terminal groups not included in this study. The information garnered from these systems can be utilized to aid in further exploration of the monolayer chemical parameter space by identifying regions of interest and regions where tribological properties are predicted to be poor.

In Fig. 5a the COF and adhesion results for all systems, both chemically-identical and chemically-dissimilar, are shown. The ideal monolayer chemistry should feature both a low COF and low adhesive force. As such, systems with favorable chemistries would exist in the lower left-hand corner of the plot. While the chemically-dissimilar films are not found to feature substantially lower values of COF or adhesion compared to the chemically-identical systems, several chemically-dissimilar systems do appear to provide favorable tribological properties, as highlighted in Fig. 5b and c. Interestingly, the majority of these systems feature one monolayer that is polar/hydrophilic and another that is nonpolar/hydrophobic. This helps highlight the potential advantages of using a lubrication scheme where the two contacting surfaces feature different lubricant molecules. For example, two of the systems shown in Fig. 5b and c. feature one monolayer that is terminated by carboxyl groups. Carboxyl-terminated monolayers are found to yield the highest adhesive forces for chemically-identical systems (Fig. 4b), yet when paired with a nonpolar counter monolayer the ability to form inter-monolayer hydrogen bonds is eliminated, and these combinations of terminal groups are able to provide favorable COF and adhesion values. This agrees with results in the literature that have observed reduced COF and adhesion for polar-nonpolar systems,[3,27,28] although we note that the effects on COF here are weaker than observed in the cited experimental studies. Along with carboxyls, several other terminal group chemistries are found to appear in multiple systems in Fig. 5b and c. In fact, the three systems observed to have the lowest adhesive forces, including the system that also features the lowest COF, all feature one monolayer that is terminated by a nitrile group. Additionally, vinyl-terminated monolayers are also present in several of these systems. While these two groups (nitrile and vinyl) have several differences (chemical composition, polarity), they share many features as well. Both groups are

linear and feature a cylindrical shape with a small VDW radius. It appears possible that these aspects, combined with the more rigid nature owed to these groups by the presence of a double or triple bond, lead to more favorable tribological properties, where we also note that films 4,5, and 7 also feature a terminal group with a double bond.

QSPR modeling

As a first attempt at developing a model for systems with mixed chemistry, a simple approach using the arithmetic mean from the uniform systems was considered. This was found to provide a reasonable, albeit rough approximation of COF for most cases; however, adhesion force was poorly accounted for (see SI). Furthermore, such a model is limited to predicting the properties of systems composed of known uniform systems. As such, a more detailed model that allows for both improved accuracy and can be used predictively is needed. Fortuitously, the data set of 100 unique chemistries (for a backbone chain length of 17 carbons) provides adequate sampling to develop QSPR models, for the prediction of COF and adhesion for systems of arbitrary chemistry and can be used to aid in further exploration of the monolayer chemical parameter space in future studies. Furthermore, if systems are properly fingerprinted through the use of molecular descriptors, aspects of terminal group functionality that contribute most strongly to these properties can be extracted from these models and exploited to reduce the design space when looking to identify better-performing monolayers. Using the freely-available RDKit Python package over 50 descriptors can be calculated for each terminal group (a complete list of the descriptors used is available in the SI) and dual-monolayer systems can be fingerprinted using this information following the workflow shown in Fig. 2. To aid in model comprehension (as the interpretation of many of these descriptors is not immediately obvious), we have categorized these descriptors into

those that aid in characterization of molecular shape, size, complexity (such as degree of branching), and charge distribution. These clusters are visualized in Fig. 6.

Models for COF and adhesion were derived using a random forest algorithm. Table 1 details the RMSE, MAE, and the coefficient of determination (R^2) between the predicted and expected target values within both the training set and a 20% holdout test set for QSPR models of COF and adhesion, along with results from out-of-bag samples. In addition to our primary QSPR model (Model 1 in Table 1), four additional models (Models 2-5) were developed to provide further insight into performance dependence on different combinations of groups in the testing and training (i.e., test-train splits) to establish the sensitivity of the QSPR models to the input set. It is observed from Table 1 that the models for both COF and adhesion perform consistently well. Random forest is known to overfit to the training data, so the high R^2 values observed for the training data sets are expected, and do not alone provide a great metric for the model's performance. However, R^2 values above 0.6 are observed for both the test data as well as for the out-of-bag samples for both models, providing more reliable estimates of model performance, and suggesting respectable predictive power. In particular, the out-of-bag results should be emphasized, as for the small data set considered here (100 total samples, only 20 of which are included within the test set) evaluation metrics on the test data set for the five models shown in Table 1 are shown to feature moderate fluctuation. This suggests that for some models the small test set does not provide a good representation of the overall population (or at least, the feature distributions within the training and testing sets differ). However, the evaluation metrics obtained for the out-of-bag samples are found to remain mostly stable between the five models, providing greater confidence in the results and in the models' predictive power.

To provide visual evaluation of the models for COF and adhesion, Fig. 7 shows both the training data and the test data used to generate and evaluate each model (Model 1 from Table 1). It is found that for COF, the majority of the test data sits close to the $y=x$ line, with no significant deviations, further corroborating the positive evaluation from Table 1. Meanwhile, for adhesion, it is shown in Fig. 7b that the model's predictive capacity is best for systems with low values of adhesion. Particularly for systems with adhesion forces greater than 4nN the model appears to under-predict these values, potential due to the fact that far only a few systems with such high adhesion were considered. Systems in this region correspond to those that feature inter-monolayer hydrogen bonds. Fortunately, accuracy of this regime is of least concern, as interfacial chemistries featuring large adhesive forces would result in poor performance as lubricants. Thus, the fact that the model still properly acknowledges the large adhesion values for these systems, despite under-predicting them, allows for these systems to be correctly screened out.

Following establishment from Table 1 and Fig. 7 that the random forest models perform well at predicting COF and adhesion, the relative contributions to the performance of each model, the "feature importances", can be extracted to evaluate which elements of the system fingerprint have the most influence over each variable. Figs. 8 and 9 provides the relative contributions of the eight molecular descriptors with the highest contributions to the prediction of COF and adhesion, respectively, along with values for each terminal group chemistry of the top four contributing descriptors to each property. The "-mean" or "-min" in the descriptor names in these figures indicate whether this corresponds to the mean or minimum value, respectively, of the descriptor for the two terminal groups in the system (refer back to Fig. 2e). For COF, the majority of the top contributing descriptors are those which describe mean values between the two terminal groups, supporting what was observed in the mean-based model found in the SI that COF can be reasonably

predicted from the uniform systems. Furthermore, it is observed that the majority of features with high contributions to the model are those which describe molecular shape. In particular, the molecular descriptor with the largest contribution is the mean Hall-Kier alpha value between the two terminal groups in the system, where the positive correlation indicates that lower Hall-Kier alpha values should result in lower COF. The Hall-Kier alpha is described by[62]:

$$\alpha = \sum_{i=1}^{A} \left(\frac{R_i}{R_{C_{sp^3}}} - 1 \right) \quad (4)$$

where A is the number of (non-hydrogen) atoms in the molecule, $R_{C_{sp^3}}$ is the covalent radius of the sp^3 carbon, and R_i is the covalent radius of atom i. From this definition, the Hall-Kier alpha is shown to provide a measure of hybridization, where contributions to α decrease from $sp^3 > sp^2 >$ sp hybridized atoms. This agrees with the terminal group chemistries observed in Fig. 8b to feature the lowest Hall-Kier alpha values, which are those featuring aromatic rings, and to a lesser extent carbonyl-containing groups and the nitrile group. Fig. 8a also shows model dependence on the "inertial shape factor", or ISF, which (described further in the SI) is a ratio of the systems principal moments of inertia. A negative correlation is observed between ISF and COF, and from Fig. 8b it is observed that the ISF of the nitrile group dominates the values calculated for any other group (whose values can be approximated as zero). This is a result of the perfect linear nature of the nitrile terminal group and reveals how the model accounts of the low COF of systems containing a nitrile group (which was observed in Fig. 5c). From these results, it can be summarized that of shape, size, charge distribution, and complexity, molecular shape is the most important terminal group aspect relating to COF, and that COF will be minimized for terminal group moieties that are planar or linear.

Feature contributions to the QSPR model for the prediction of adhesion are shown in Fig. 9a. Following expectations, nearly all of the top contributing molecular descriptors provide measures of the distribution of charge on the terminal group. Furthermore, nearly all of these descriptors are found to be those that correspond to the *minimum* value between the two terminal groups, rather than the mean value (thus one film will dominate the behavior). The top contributing feature is found to be the minimum topological polar surface area (TPSA) between the two terminal groups. That this feature is a strong predictor of adhesion is sensible, as large adhesive forces will require both terminal groups in the system to feature significant charge imbalances (i.e., large dipole moments). From Fig. 9b, the terminal groups featuring the high TPSA are found to be the nitro and nitrophenyl groups in addition to the three groups featuring hydrogen bond donors and acceptors (carboxyl, hydroxyl, and amino). However, the fact that a positive correlation is observed with the *minimum* TPSA suggests that if one of the monolayers features a terminal group with a low polar surface area (a nonpolar molecule such as a hydrocarbon), the adhesion force between the two monolayers will be low. This agrees with the findings of Fig. 5c. A modest contribution to the QSPR model is also observed from the "hbonds" descriptor, developed in this work and described in further details in the Methods section. This descriptor characterizes a system's ability to form inter-monolayer hydrogen bonds, and the positive correlation indicates that the more combinations of hydrogen bond donors and acceptors the higher the adhesive force should be. While systems featuring inter-monolayer hydrogen bonding are found to not have their adhesive forces predicted with high accuracy, as evidenced by Fig. 7b, the addition of the "hbonds" descriptor does allow the model to achieve a qualitatively correct prediction that such systems will feature high adhesion. Thus, for future exploration of the monolayer chemical parameter space, such systems can be successfully filtered out with this QSPR model.

Discussion

The use of MoSDeF to facilitate screening over chemical space enables relationships to be uncovered that could easily be overlooked or obscured in small-scale studies. As a specific example of the utility of this screening study, consider the nitrile-terminated monolayers which were observed in Fig. 4a to feature the lowest COF for chemically-identical films, yet the adhesive force for these films (2.27nN ± 0.14nN) is prohibitively high due to their polarity. However, when paired with a counter monolayer that is hydrophobic, the adhesion force is decreased while the COF is largely unaffected, and several nitrile-containing systems result in the most optimal systems (Fig. 5c) from the pool of terminal groups examined in this study. Thus, the use of chemically-dissimilar surfaces is found to provide a means to utilize terminal groups that yield low COFs, even if they would normally feature a large adhesive force, provided the other monolayer is hydrophobic.

While the difference in COF between nitrile-terminated films, and other, hydrophobic films (such as vinyl), is found to be slight, there may be polar terminal groups not included within this study that yield lower COFs, and the use of chemically-dissimilar monolayers would provide a route to take advantage of that property. It should be noted that the distribution of points in Fig. 5a, suggests that the lower bound of COF may not be substantially reduced by the mixing of monolayer chemistries (i.e., the COF of the mixed system will likely not be lower than the lowest COF of either of the chemically-identical systems). Thus, if seeking to reduce COF, it appears that additional modifications to the monolayer would be necessary; either using a functional group not included in the pool of this study or through some other means such as the use of multicomponent films (CITE PFA/alkane paper).

In addition to providing insight into aspects of terminal group chemistry that have the greatest influence on COF and adhesion, the QSPR models generated in this work should also have utility in predicting values for these variables for systems with arbitrary terminal group chemistry, which should aid in further exploration of the monolayer chemical parameter space. Although the R^2 values of around 0.6 for these models in Table 1 may not appear exceptional at first, these results are in fact quite remarkable when considering the simplicity of the inputs required. The simulations in this work were performed using the OPLS force field, while molecular descriptors were calculated for terminal group moieties provided to the models in the form of SMILES, where Gasteiger charges were assigned in place of OPLS charges and 3D conformations were generated through a stochastic method. If OPLS charges and 3D structures generated using OPLS had been provided to the models, their ability to predict values of COF and adhesion would have been limited to only those chemical moieties that could be described by OPLS. However, by using generic SMILES as inputs, arbitrary molecules that may not be parameterized for use with OPLS can still have values of COF and adhesion predicted. The QSPR models themselves are hosted on GitHub, along with all of the code related to this work, and should have utility in future studies with the aim of further exploring the monolayer chemical parameter space.

As emphasized in the Introduction, the monolayer chemical parameter space is vast, and brute force exploration is costly. Each system in this work required roughly 16 real hours of computing time to perform all the MD simulations and analysis, despite GPU acceleration. Thus, expanding such a study to 1000s or 10000s of systems could be intractable through a pure brute force approach. This further emphasizes the utility of the QSPR models derived here in filtering areas of this parameter space where exploration is unnecessary (large values of COF or adhesion). The use of these models in such a predictive capacity will be explored in future work.

CONCLUSION

In this work, the screening of functionalized monolayer films, enabled by use of the MoSDeF software suite and Signac framework, has been performed using MD simulations. In agreement with prior literature, increases in the length of the chain backbone are found to reduce COF and increase monolayer order. Adhesion between monolayer films is observed to be relatively insensitive to backbone chain length. The effects of terminal group chemistry on monolayer tribology are examined for both chemically-identical systems (i.e., where both monolayers feature the same chemistry) and chemically-dissimilar systems. It is observed that combinations of polar and nonpolar terminal groups in chemically-dissimilar films yield favorable tribological properties (i.e., low COF and low adhesion), thus the utility of chemically-dissimilar systems appears primarily in the reduction of adhesion through the inclusion of at least one nonpolar terminal group.

The breadth of the data in this study has afforded the ability to generate QSPR models, which require a simple SMILES representation as an input, and are found to yield reasonable predictive capability. Feature extraction from these models reveals that COF is most sensitive to terminal group shape, whereby planar or linear groups result in the lowest COF values. Adhesion is found to be most sensitive to charge distribution on the terminal group, whereby both the polar surface area and ability for the formation of inter-monolayer hydrogen bonding are found to be strong predictors of adhesion. The QSPR models generated in this work have utility in narrowing the scope of the monolayer parameter space for future screening. Furthermore, the workflow utilized in this study should be readily extensible to the examination of more complex monolayer films (e.g., multi-component films) which may provide even more favorable tribological properties.

ASSOCIATED CONTENT

Supporting Information. A complete description of all force field parameters used is provided in tabular format. Short definitions for all molecular descriptors used in QSPR models are also provided. Also included are details on mean-based models for estimating COF and adhesion of chemically-dissimilar systems from the results obtained for chemically-identical films. A short description of the procedure for carving amorphous silica surfaces is also provided.

AUTHOR INFORMATION

Corresponding Author

*E-mail: c.mccabe@vanderbilt.edu

Funding Sources

Funding for this work has been provided by the National Science Foundation (NSF) through Grant ACI-1047827.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

The authors would like to thank Simon Adorf for assistance in onboarding with the Signac framework. We would additionally like to acknowledge Christoph Klein and Janos Sallai for their work on the mBuild and Foyer packages facilitating this work. The National Science Foundation (NSF) is acknowledged for providing funding for this work through Grant ACI-1047827. This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge

National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. Computational time was also provided the U.S. Department of Energy INCITE program, grant NTI112.

ABBREVIATIONS

AFM, atomic force microscopy; COF, coefficient of friction; MAE, mean absolute error; MEMS, microelectromechanical systems; MoSDeF, Molecular Simulation and Design Framework; QSPR, quantitative structure-property relationships; RMSE, root-mean-square error; VDW, van der Waals.

REFERENCES

(1) Xiao, X.; Hu, J.; Charych, D. H.; Salmeron, M. Chain Length Dependence of the Frictional Properties of Alkylsilane Molecules Self-Assembled on Mica Studied by Atomic Force Microscopy. *Langmuir* **1996**, *12* (2), 235–237.

(2) Lio, A.; Charych, D. H.; Salmeron, M. Comparative Atomic Force Microscopy Study of the Chain Length Dependence of Frictional Properties of Alkanethiols on Gold and Alkylsilanes on Mica. *J. Phys. Chem. B* **1997**, *101* (19), 3800–3805.

(3) Brewer, N. J.; Beake, B. D.; Leggett, G. J. Friction Force Microscopy of Self-Assembled Monolayers: Influence of Adsorbate Alkyl Chain Length, Terminal Group Chemistry, and Scan Velocity. *Langmuir* **2001**, *17* (6), 1970–1974.

(4) Huo, L.; Du, P.; Zhou, H.; Zhang, K.; Liu, P. Fabrication and Tribological Properties of Self-Assembled Monolayer of n-Alkyltrimethoxysilane on Silicon: Effect of SAM Alkyl Chain Length. *Appl. Surf. Sci.* **2016**.

(5) Chandross, M.; Grest, G. S.; Stevens, M. J. Friction between Alkylsilane Monolayers: Molecular Simulation of Ordered Monolayers. *Langmuir* **2002**, *18* (11), 8392–8399.

(6) Yu, B.; Qian, L.; Yu, J.; Zhou, Z. Effects of Tail Group and Chain Length on the Tribological Behaviors of Self-Assembled Dual-Layer Films in Atmosphere and in Vacuum. *Tribol. Lett.* **2009**, *34* (1 SPEC. ISS.), 1–10.

(7) Park, B.; Lorenz, C. D.; Chandross, M.; Stevens, M. J.; Grest, G. S.; Borodin, O. A. Frictional Dynamics of Fluorine-Terminated Alkanethiol Self-Assembled Monolayers. *Langmuir* **2004**, *20* (23), 10007–10014.

(8) Lewis, J. Ben; Vilt, S. G.; Rivera, J. L.; Jennings, G. K.; McCabe, C. Frictional Properties of Mixed Fluorocarbon/Hydrocarbon Silane Monolayers: A Simulation Study. *Langmuir* **2012**, *28* (40), 14218–14226.

(9) Booth, B. D.; Vilt, S. G.; McCabe, C.; Jennings, G. K. Tribology of Monolayer Films: Comparison between n-Alkanethiols on Gold and n-Alkyl Trichlorosilanes on Silicon. *Langmuir* **2009**, *25* (17), 9995–10001.

(10) Rivera, J. L.; Jennings, G. K.; McCabe, C. Examining the Frictional Forces between Mixed Hydrophobic-Hydrophilic Alkylsilane Monolayers. *J. Chem. Phys.* **2012**, *136* (24), 244701.

(11) Park, B.; Chandross, M.; Stevens, M. J.; Grest, G. S. Chemical Effects on the Adhesion and Friction between Alkanethiol Monolayers: Molecular Dynamics Simulations. *Langmuir* **2003**, *19* (22), 9239–9245.

(12) Summers, A. Z.; Iacovella, C. R.; Cummings, P. T. Investigating Alkylsilane Monolayer

Tribology at a Single-Asperity Contact with Molecular Dynamics Simulation. **2017**, 11270–11280.

(13) Lorenz, C. D.; Webb, E. B.; Stevens, M. J.; Chandross, M.; Grest, G. S. Frictional Dynamics of Perfluorinated Self-Assembled Monolayers on Amorphous SiO2. *Tribol. Lett.* **2005**, *19* (2), 93–98.

(14) Summers, A. Z.; Iacovella, C. R.; Billingsley, M. R.; Arnold, S. T.; Cummings, P. T.; McCabe, C. Influence of Surface Morphology on the Shear-Induced Wear of Alkylsilane Monolayers: Molecular Dynamics Study. *Langmuir* **2016**.

(15) Black, J. E.; Iacovella, C. R.; Cummings, P. T.; McCabe, C. Molecular Dynamics Study of Alkylsilane Monolayers on Realistic Amorphous Silica Surfaces. *Langmuir* **2015**, *31*, 3086–3093.

(16) MoSDeF Github https://github.com/mosdef-hub (accessed Jul 10, 2017).

(17) Adorf, C. S.; Dodd, P. M.; Ramasubramani, V.; Glotzer, S. C. Simple Data and Workflow Management with the Signac Framework. *Comput. Mater. Sci.* **2018**, *146*, 220–229.

(18) Weininger, D. SMILES, a Chemical Language and Information System: 1: Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28* (1), 31–36.

(19) Zhang, L.; Li, L.; Chen, S.; Jiang, S. Measurements of Friction and Adhesion for Alkyl Monolayers on Si (111) by Scanning Force Microscopy. **2002**, No. 111, 5448–5456.

(20) Zhang, L.; Jiang, S. Molecular Simulation Study of Nanoscale Friction between Alkyl Monolayers on Si(111) Immersed in Solvents. *J. Chem. Phys.* **2003**, *119* (2), 765–770.

(21) Zhuravlev, L. T. The Surface Chemistry of Amorphous Silica. Zhuravlev Model. *Colloids Surfaces A Physicochem. Eng. Asp.* **2000**, *173* (1–3), 1–38.

(22) Hartkamp, R.; Siboulet, B.; Dufrêche, J.-F.; Coasne, B. Ion-Specific Adsorption and Electroosmosis in Charged Amorphous Porous Silica. *Phys. Chem. Chem. Phys.* **2015**, *17* (38), 24683–24695.

(23) Klein, C.; Sallai, J.; Jones, T. J.; Iacovella, C. R.; McCabe, C.; Cummings, P. T. A Hierarchical, Component Based Approach to Screening Properties of Soft Matter. In *Foundations of Molecular Modeling and Simulation*; 2016; pp 79–92.

(24) mBuild: a component-based molecule builder tool that relies on equivalence relations for component composition. http://mosdef-hub.github.io/mbuild/ (accessed Jul 10, 2017).

(25) Labukas, J. P.; Drake, T. J. H.; Ferguson, G. S. Compatibility of ω -Functionality in the Electrochemically Directed Self-Assembly of Monolayers on Gold from Alkyl Thiosulfates. **2010**, *26* (12), 9497–9505.

(26) Clear, S. C.; Nealey, P. F. Chemical Force Microscopy Study of Adhesion and Friction between Surfaces Functionalized with Self-Assembled Monolayers and Immersed in Solvents. **1999**, *250*, 238–250.

(27) Frisbie, C. D.; Rozsnyai, L. F.; Noy, A.; Wrighton, M. S.; Lieber, C. M. Functional Group Imaging by Chemical Force Microscopy. *Science (80-.).* **1994**, *265* (5181), 2071–2074.

(28) Noy, A.; Frisbie, C. D.; Rozsnyai, L. F.; Wrighton, M. S.; Lieber, C. M. Chemical Force Microscopy: Exploiting Chemically-Modified Tips To Quantify Adhesion, Friction, and

Functional Group Distributions in Molecular Assemblies. *J. Am. Chem. Soc.* **1995**, *117* (30), 7943–7951.

(29) Beake, B. D.; Leggett, G. J. Friction and Adhesion of Mixed Self-Assembled Monolayers Studied by Chemical Force Microscopy. **1999**, 3345–3350.

(30) Iacovella, C. R.; Sallai, J.; Klein, C.; Ma, T. Idea Paper : Development of a Software Framework for Formalizing Forcefield Atom-Typing for Molecular Simulation. In *4th Workshop on Sustainable Software for Science: Practice and Experiences (WSSSPE4)*; 2016.

(31) Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.; Beauchamp, K. A.; Lane, T. J.; Wang, L. P.; Shukla, D.; et al. OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *J. Chem. Theory Comput.* **2013**, *9* (1), 461–469.

(32) SMARTS - A Language for Describing Molecular Patterns.

(33) Foyer Github https://github.com/mosdef-hub/foyer (accessed Jul 10, 2017).

(34) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118* (15), 11225–11236.

(35) McDonald, N. A.; Jorgensen, W. L. Development of an All-Atom Force Field for Heterocycles. Properties of Liquid Pyrrole, Furan, Diazoles, and Oxazoles. *J. Phys. Chem. B* **1998**, *102* (41), 8049–8059.

(36) Price, D. J.; Roberts, J. D.; Jorgensen, W. L. Conformational Complexity of Succinic Acid and Its Monoanion in the Gas Phase and in Solution: Ab Initio Calculations and Monte Carlo Simulations. *J. Am. Chem. Soc.* **1998**, *120* (37), 9672–9679.

(37) Rizzo, R. C.; Jorgensen, W. L. OPLS All-Atom Model for Amines: Resolution of the Amine Hydration Problem. *J. Am. Chem. Soc.* **1999**, *121* (20), 4827–4836.

(38) Price, M. L. P.; Ostrovsky, D.; Jorgensen, W. L. Gas-Phase and Liquid-State Properties of Esters, Nitriles, and Nitro Compounds with the OPLS-AA Force Field. *J. Comput. Chem.* **2001**, *22* (13), 1340–1352.

(39) Watkins, E. K.; Jorgensen, W. L. Perfluoroalkanes: Conformational Analysis and Liquid-State Properties from Ab Initio and Monte Carlo Calculations. **2001**, 4118–4125.

(40) Jorgensen, W. L.; Ulmschneider, J. P.; Tirado-Rives, J. Free Energies of Hydration from a Generalized Born Model and an All-Atom Force Field. *J. Phys. Chem. B* **2004**, *108* (41), 16264–16270.

(41) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, 1–7.

(42) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19.

(43) LAMMPS web page: http://lammps.sandia.gov.

(44) Chandross, M.; Webb III, E. B.; Stevens, M. J.; Grest, G. S.; Garofalini, S. H. Systematic

Study of the Effect of Disorder on Nanotribology of Self-Assembled Monolayers. *Phys. Rev. Lett.* **2004**, *93* (16), 166103.

(45) Devaprakasam, D.; Biswas, S. K. Molecular Damping: Mechanical Response of Self-Assembled Monomolecular Layer to Compression. *Phys. Rev. B - Condens. Matter Mater. Phys.* **2005**, *72* (May), 1–8.

(46) Nosé, S. A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *J. Chem. Phys.* **1984**, *81* (1), 511.

(47) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31* (3), 1695–1697.

(48) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. {LINCS}: {A} {L}inear {C}onstraint {S}olver for {M}olecular {S}imulations. *{J}.~{C}omput. {C}hem.* **1997**, *18* (12), 1463–1472.

(49) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N ·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092.

(50) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577–8593.

(51) RDKit: Open-source cheminformatics.

(52) Riniker, S.; Landrum, G. A. Better Informed Distance Geometry: Using What We Know to Improve Conformation Generation. *J. Chem. Inf. Model.* **2015**, *55* (12), 2562–2574.

(53) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity—a

Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36* (22), 3219–3228.

(54) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45* (1), 5–32.

(55) Geurts, P.; Ernst, D.; Wehenkel, L. Extremely Randomized Trees. *Mach. Learn.* **2006**, *63* (1), 3–42.

(56) Pedregosa, F.; Weiss, R.; Brucher, M. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

(57) Ballester, P. J.; Mitchell, J. B. O. A Machine Learning Approach to Predicting Protein-Ligand Binding Affinity with Applications to Molecular Docking. *Bioinformatics* **2010**, *26* (9), 1169–1175.

(58) Svetnik, V.; Liaw, A.; Tong, C.; Christopher Culberson, J.; Sheridan, R. P.; Feuston, B. P. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (6), 1947–1958.

(59) Lee, D. H.; Oh, T.; Cho, K. Combined Effect of Chain Length and Phase State on Adhesion/Friction Behavior of Self-Assembled Monolayers. *J. Phys. Chem. B* **2005**, *109* (22), 11301–11306.

(60) Booth, B. D.; Vilt, S. G.; Lewis, J. Ben; Rivera, J. L.; Buehler, E. A.; McCabe, C.; Jennings, G. K. Tribological Durability of Silane Monolayers on Silicon. *Langmuir* **2011**, *27* (10), 5909–5917.

(61) Cheng, H.; Hu, Y. Influence of Chain Ordering on Frictional Properties of Self-Assembled Monolayers (SAMs) in Nano-Lubrication. *Adv. Colloid Interface Sci.* **2012**, *171–172*, 53–

65.

(62) Hall, L. H.; Kier, L. B. The Molecular Connectivity Chi Indexes and Kappa Shape Indexes in Structure-Property Modeling. In *Reviews in Computational Chemistry, Volume 2*; 1991; pp 367–422.

(63) Klein, C.; Summers, A.Z.; Thompson, M.W.; Gilmer, J.B.; McCabe, C.; Cummings, P.T.; Sallai, J.; Iacovella, C.R. Fomralizing Atom-typing and the Dissemination of Force Fields with Foyer. *Computational Materials Science*; **2019**, 167, 215-227.

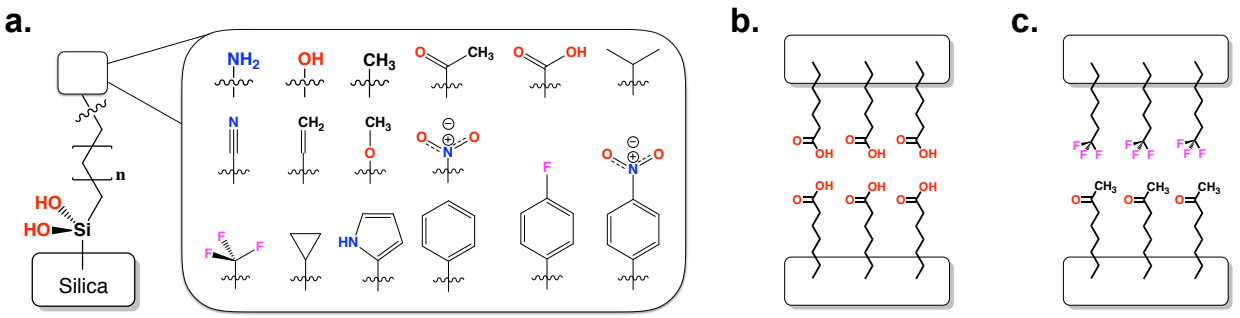

Figure 1. a.) Overview of the chemical parameter space examined in this work. The pool of terminal group chemistries consists of (row 1) amino, hydroxyl, methyl, acetyl, carboxyl, isopropyl; (row 2) nitrile, vinyl, methoxy, nitro; (row 3) perfluoromethyl, cyclopropyl, 2-pyrrole, phenyl, fluorophenyl, and nitrophenyl. Dual-monolayer systems include both b.) chemically-identical and c.) chemically-dissimilar compositions.

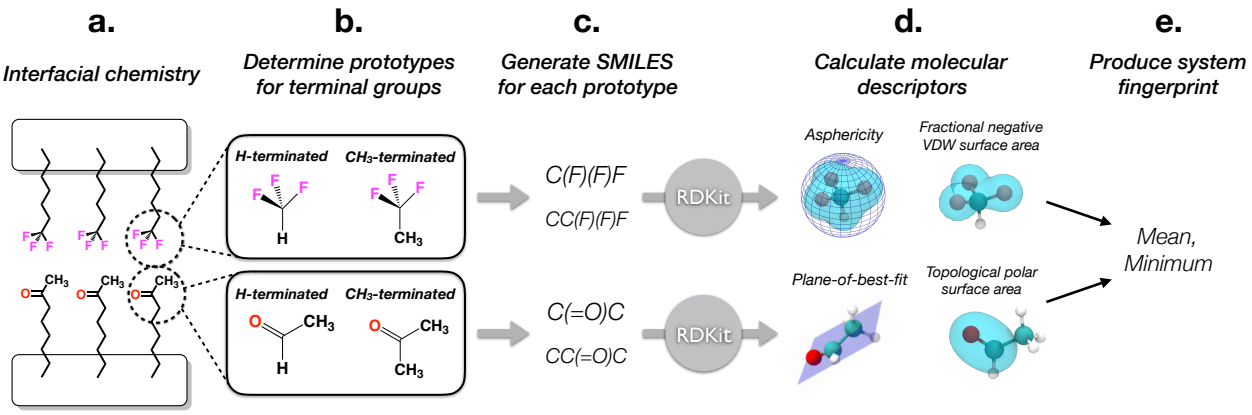

Figure 2. Workflow of the process used to fingerprint each interfacial chemistry. a.) For a given interfacial chemistry (a perfluoromethyl-acetyl system is shown as an example), b.) prototypes are determined for each terminal group, c.) these prototypes are then converted to SMILES representations, d). and fed to the RDKit Python package used to calculate a variety of molecular descriptors, e.) the fingerprint for the interfacial chemistry is then described by the mean and minimum values for all molecular descriptors calculated for the two terminal groups.

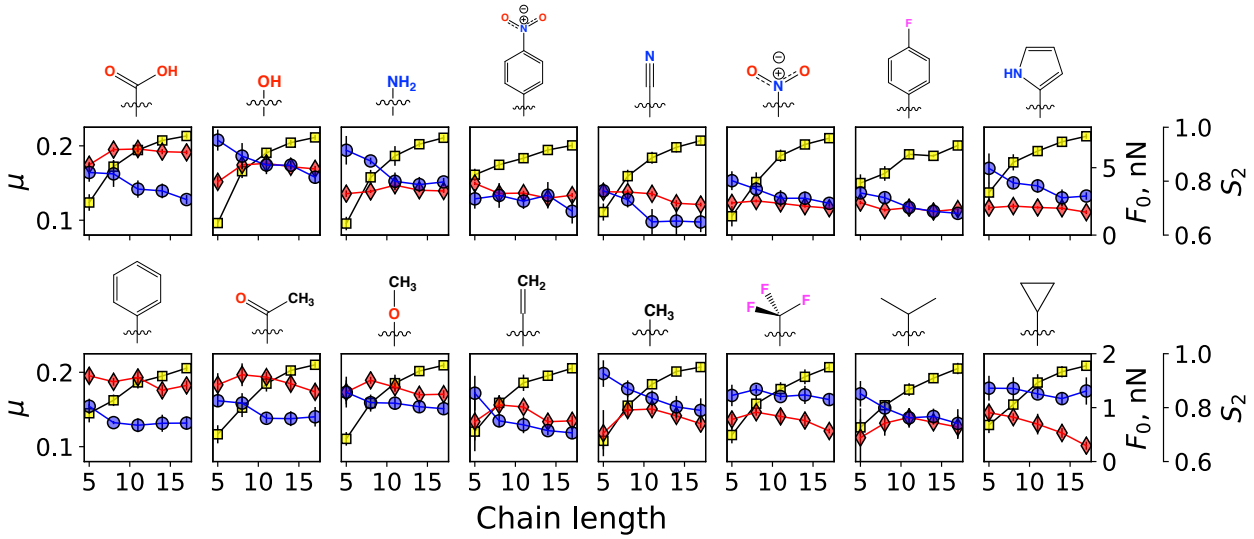

Figure 3. Coefficient of friction (μ, blue circles), adhesive force (F_0, red diamonds), and nematic order (S_2, yellow squares) for chemically-identical systems shown for each terminal group chemistry as a function of chain length. Note the scale for F_0 differs between the two rows in order to clearly display trends for both polar and nonpolar systems. Error bars represent a single standard deviation calculated from the average of the five chain attachment configurations for each system.

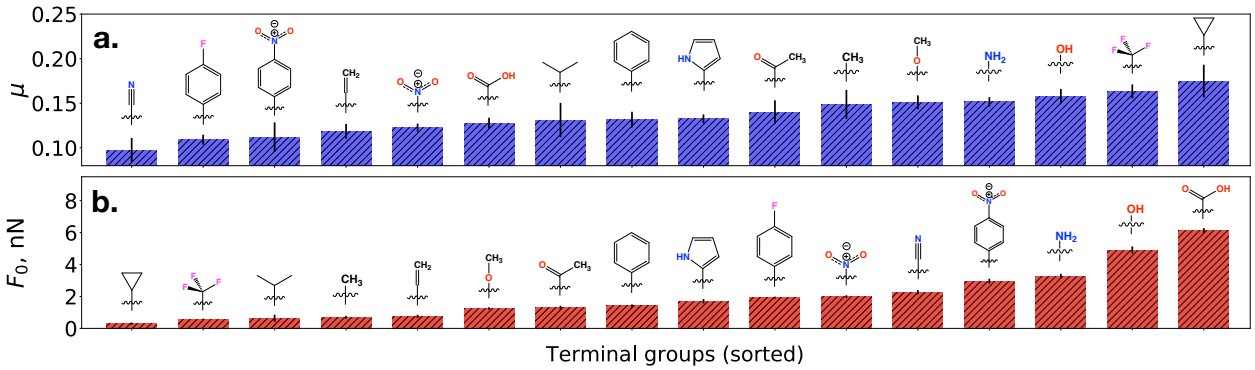

Figure 4. Bar graphs displaying a.) coefficient of friction and b.) adhesive force as a function of terminal group chemistry for chemically-identical monolayer systems with a backbone chain length of 17 carbons.

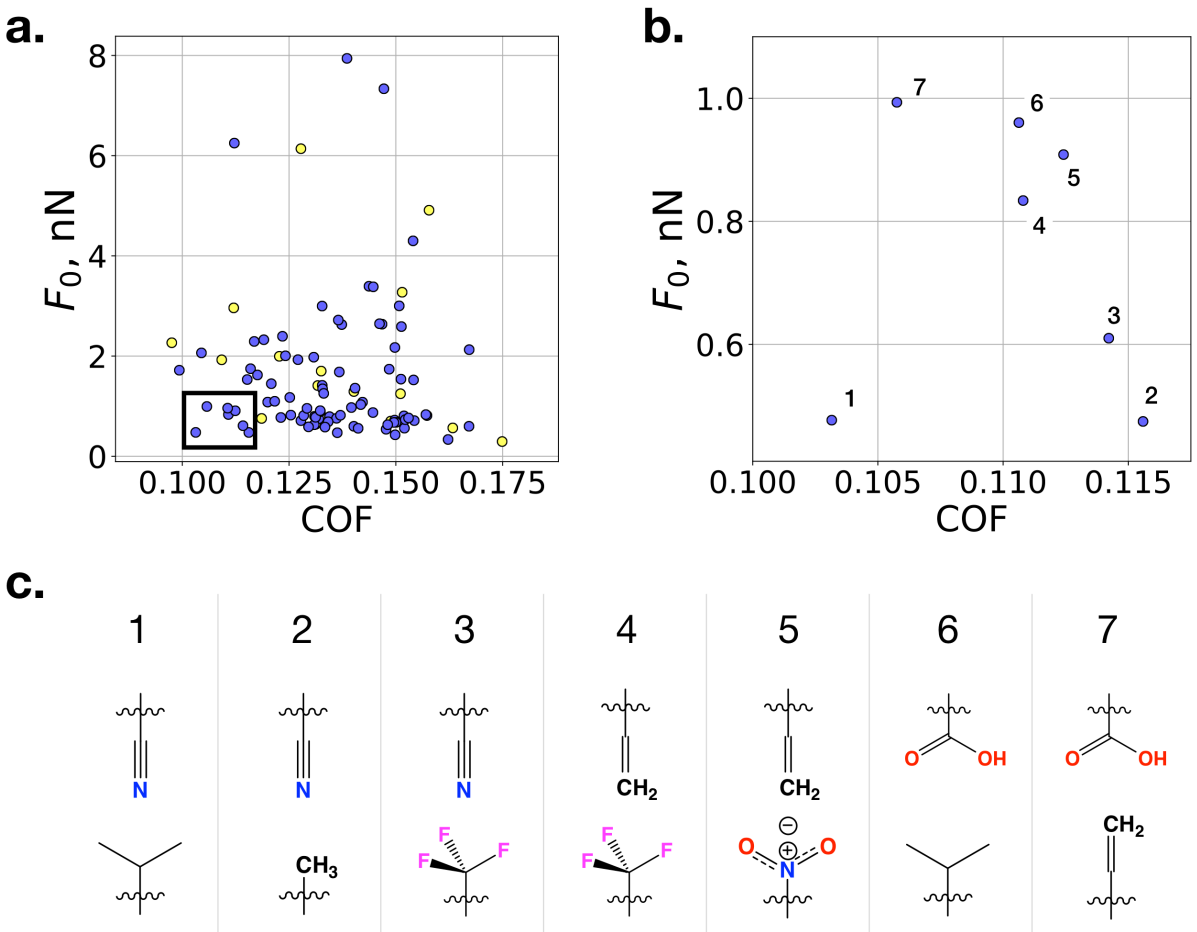

Figure 5. a.) Scatter plot of COF and adhesion force data for chemically-identical systems (yellow) and chemically-dissimilar systems (blue). The highlighted region indicates the area featuring the most tribologically favorable systems and is shown in larger in b.), where the seven systems in this region are annotated and the corresponding terminal group chemistries are shown in c).

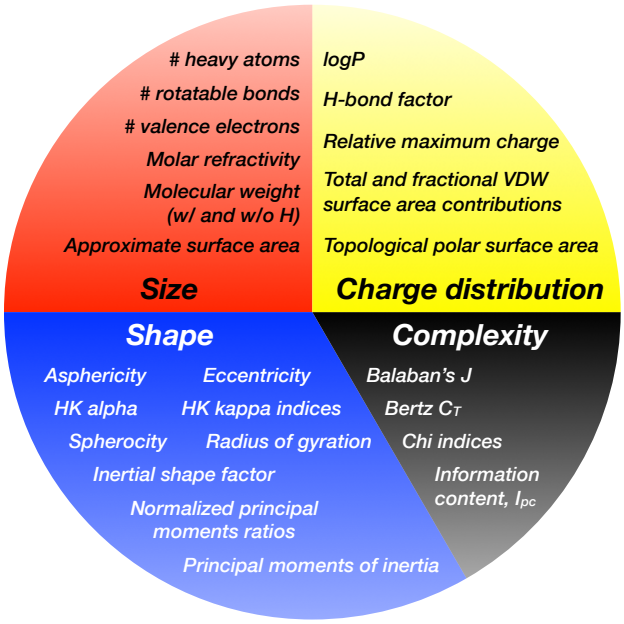

Figure 6. Features considered for the QSPR models, grouped into clusters describing charge distribution, complexity (i.e., degree of connectivity/branching), shape, and size. Additional descriptions of these features can be found in the Supporting Information.

Table 1. Evaluation of random forest regression models for COF and F_0[†]

Model Number	Target Variable	Training set			Cross-validation (OOB)			Test set		
		R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
1	COF	0.9496	0.0037	0.0029	0.6306	0.0100	0.0079	0.6524	0.0090	0.0068
	F_0 (nN)	0.9600	0.2644	0.1383	0.6985	0.7259	0.3782	0.6905	0.9530	0.4237
2	COF	0.9473	0.0036	0.0027	0.6067	0.0099	0.0074	0.6049	0.0110	0.0086
	F_0 (nN)	0.9539	0.3067	0.1473	0.6586	0.8351	0.4046	0.9422	0.3252	0.2439
3	COF	0.9418	0.0037	0.0028	0.5594	0.0101	0.0078	0.6571	0.0114	0.0087
	F_0 (nN)	0.9747	0.2283	0.1233	0.7978	0.6452	0.3410	0.4066	1.0031	0.5579
4	COF	0.9471	0.0038	0.0029	0.6029	0.0104	0.0081	0.8204	0.0062	0.0048
	F_0 (nN)	0.9435	0.2974	0.1406	0.5740	0.8165	0.3844	0.7787	0.8453	0.6076
5	COF	0.9484	0.0038	0.0030	0.6043	0.0106	0.0081	0.5199	0.0096	0.0069
	F_0 (nN)	0.9613	0.2642	0.1388	0.7119	0.7213	0.3797	0.6367	1.0031	0.4699
Avg.	COF	0.9468 ± 0.0027	0.0037 ± 0.0001	0.0029 ± 0.0001	0.6008 ± 0.0230	0.0102 ± 0.0003	0.0079 ± 0.0003	0.6509 ± 0.0980	0.0094 ± 0.0018	0.0072 ± 0.0014
	F_0 (nN)	0.9587 ± 0.0102	0.2722 ± 0.0278	0.1377 ± 0.0079	0.6882 ± 0.0729	0.7488 ± 0.0693	0.3776 ± 0.0206	0.6909 ± 0.1759	0.8259 ± 0.2569	0.4606 ± 0.1261

[†] The following metrics are used for evaluation; R^2: coefficient of determination between predicted and actual values, RMSE: root-mean-square error, MAE: mean absolute error

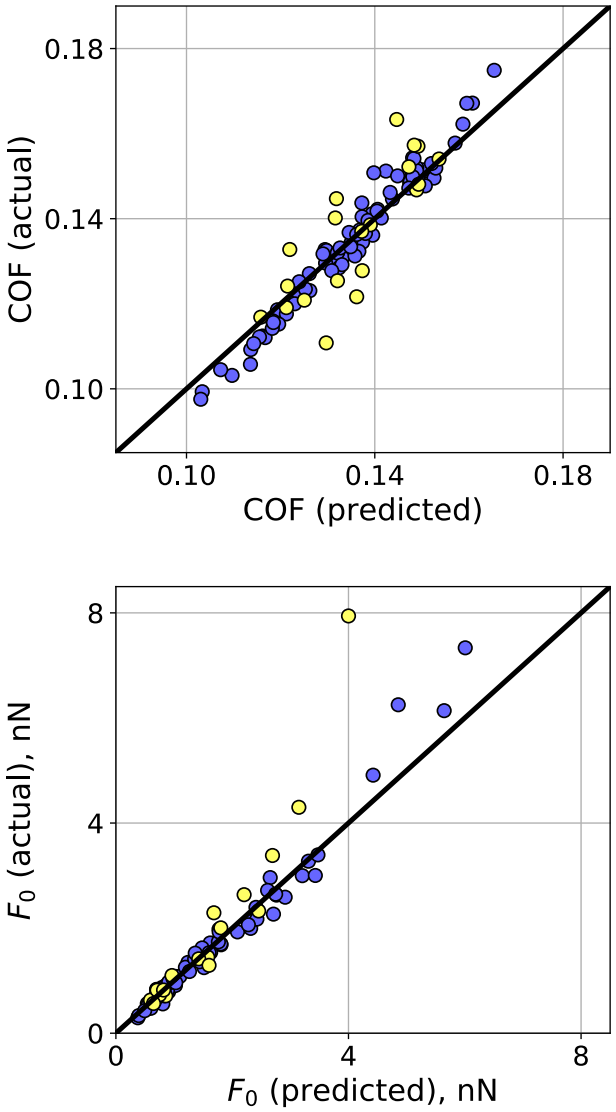

Figure 7. Values predicted by QSPR models for a.) COF and b.) adhesive force compared to the actual values. The $y=x$ line is drawn in black for reference. Points in blue denote data used as part of the training set, while points in yellow denote data that was part of the test set.

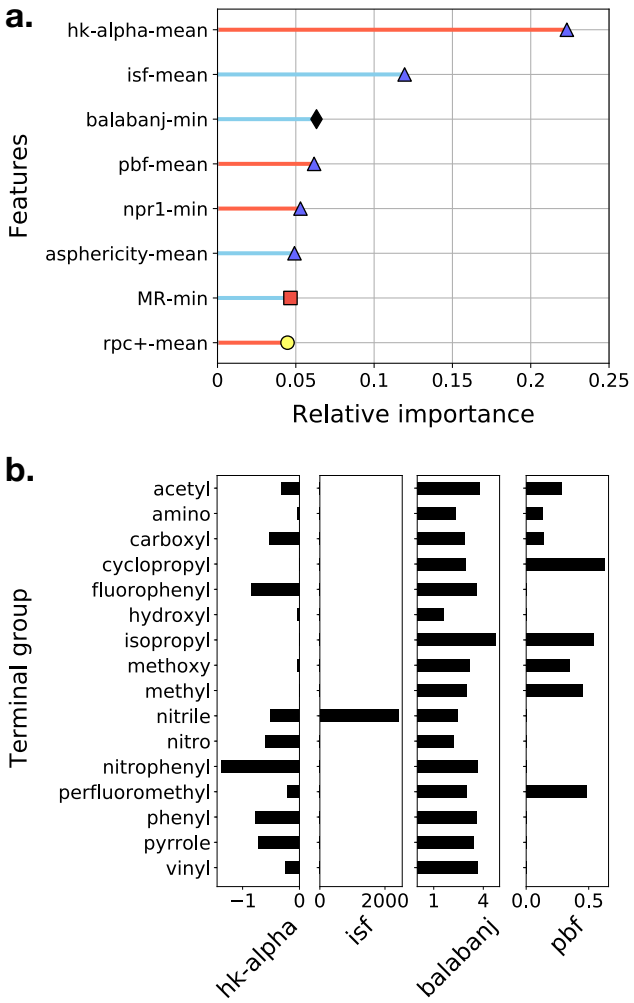

Figure 8. a.) Relative feature importances extracted from the random forest regression model (Model 1 in Table 1) for COF, and b.) values of the top four contributing features for each terminal group chemistry. Symbols in a.) represent the molecular aspect characterized by each feature (yellow circles: charge distribution, blue triangles: shape, red squares: size, black diamonds: complexity). Stick colors represent whether each feature correlates positively (red) or negatively (light blue) with COF.

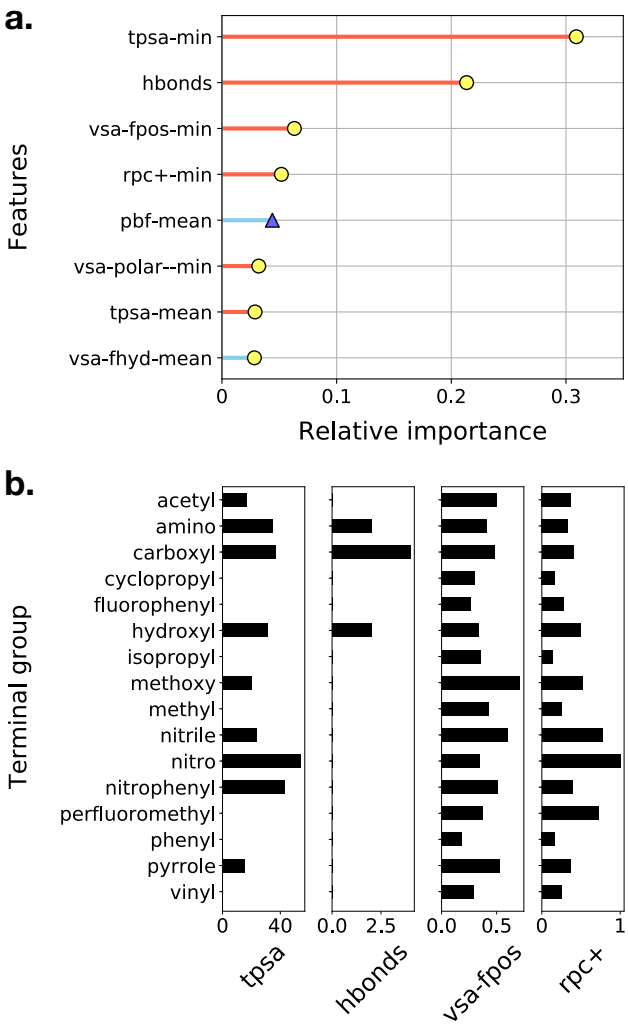

Figure 9. a.) Relative feature importances extracted from the random forest regression model (Model 1 in Table 1) for adhesion, and b.) values of the top four contributing features for each terminal group chemistry. Symbols in a.) represent the molecular aspect characterized by each feature (yellow circles: charge distribution, blue triangles: shape, red squares: size, black diamonds: complexity). Stick colors represent whether each feature correlates positively (red) or negatively (light blue) with adhesion.