JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

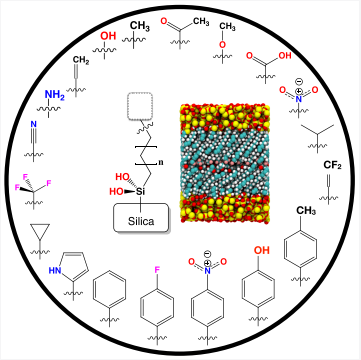
MoSDeF, a Python Framework Enabling Large-Scale Computational Screening of Soft Matter: Application to Chemistry-Property Relationships in Lubricating Monolayer Films

Andrew Z. Summers, Justin B. Gilmer, Christopher R. Iacovella, Peter T. Cummings, and Clare McCabe*

Cite This: J. Chem. Theory Comput. 2020, 16, 1779−1793 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We demonstrate how the recently developed Python-based Molecular Simulation and Design Framework (MoSDeF) can be used to perform molecular dynamics screening of functionalized monolayer films, focusing on tribological effectiveness. MoSDeF is an open-source package that allows for the programmatic construction and parametrization of soft matter systems and enables TRUE (transferable, reproducible, usable by others, and extensible) simulations. The MoSDeF-enabled screening identifies several film chemistries that simultaneously show low coefficients of friction and adhesion. We additionally develop a Python library that utilizes the RDKit cheminformatics library and the scikit-learn machine learning library that allows for the development of predictive models for the tribology of functionalized monolayer films and use this model to extract information on terminal group characteristics that most influence tribology, based on the screening data.



1. INTRODUCTION

Computational screening is an emerging approach to materials design and optimization. Utilization of this approach has enabled, for example, the discovery of metal organic frameworks with vastly improved gas adsorption[1] and candidates for improved organic photovoltaics.[2] More recently, machine learning has been leveraged as a means of improving the efficiency of screening, aiding in the identification of systems with desirable target properties.[3−6] However, performing computational screening of soft matter systems, such as monolayer thin films, is a challenging task; unlike hard matter systems where zero temperature, energy minimized states are often sufficient to robustly calculate relevant properties,[3] soft matter simulations require comparably longer computational times in order to relax and equilibrate the underlying constituents before ensemble properties can be measured from simulation trajectories. To achieve these longer time scales, soft matter simulations typically rely on classical force fields and use methods such as molecular dynamics (MD) to equilibrate and sample the ensemble space. This reliance on classical force fields results in additional challenges in terms of screening, as any changes to the chemistry requires not only reinitialization of the system configuration but also reparametrization of the system interactions as dictated by the force field. As a result, typical MD workflows lend themselves better to screening over thermodynamic space (e.g., varying temperature or pressure) where the system chemistry and parametrization is not changed. In practice, system setup and parametrization of MD simulations is typically performed in an *ad hoc* manner, often using human manipulation of data/workflows, thus making large-scale screening studies intractable. Furthermore, the use of in-house software and human manipulation as part of the standard practice can significantly limit reproducibility of published results.[7] Reproducibility in science is an ever-growing concern;[8,9] capturing simulation workflows in a reproducible manner is a challenge for small scale-studies, and such challenges will only be compounded as the design space is expanded by orders of magnitude for large-scale screening. In order to efficiently perform screening over chemical space, tools are required that provide the ability to systematically vary chemistry during model setup, to automatically apply force field parameters to these models, to manage the execution of the simulation workflows for large numbers of systems, and to capture all these many inputs and processes in a reproducible manner. The lack of such tools has been a limiting factor in terms of the widespread adoption of screening in soft matter, although recent efforts by numerous groups have been aimed at developing new software and approaches related to system construction/initialization[10−21] and workflow management[17,22−24] (see ref 24 for a detailed discussion of workflow management software).

Received: November 26, 2019
Published: January 31, 2020

As a means to resolve these issues, the Molecular Simulation and Design Framework (MoSDeF)[25] has been architected as an open-source Python framework that includes tools for generating molecular systems as objects with exchangeable chemical parts,[26−28] for automatically applying force field parameters,[29−32] and for generating syntactically correct input files for common molecular simulation software packages. As described in refs 26, 29, and 30, the tools underlying MoSDeF allow for all aspects of the system construction and parametrization to be performed in a programmatic, scriptable fashion, eliminating the need for intermediate human manipulation of the data or workflows, essential for reproducibility.[7] In conjunction with the Signac framework[33] for workflow management, large-scale screening simulations can therefore be conducted over chemical space using MoSDeF, whereby the exact inputs, procedures, and processes can be captured and preserved, thus ensuring that simulations have the characteristics of being transparent, reproducible, usable by others, and extensible (i.e., TRUE). The concept of TRUE simulations (specifically, the acronym TRUE) was first proposed by the authors in ref 34 as a goal for reproducible molecular simulations that could be enabled through use of the MoSDeF tools.

Here, we present the use of MoSDeF to screen the chain length and terminal group chemistry of functionalized monolayer films, examining 194 unique chemistries (2730 total simulations) with the goal of elucidating the tribological effectiveness of these films. The use of monolayer films is a promising approach to protecting and lubricating nanoscale contacts for use in applications such as nano- and micro-electromechanical systems (NEMS and MEMS).[35−38] In this work, large-scale, nonequilibrium MD screening simulations of monolayer films under shear are performed, wherein all simulations can be considered TRUE by using the MoSDeF toolkit and the Signac framework.[33] In conjunction with the screening studies, a Python library that builds upon the RDKit[39] and scikit-learn[40] Python libraries has been designed to facilitate the generation of predictive machine learning models of monolayer tribology; this library is tested with the data generated from computational screening and used to identify correlations between monolayer chemistry and tribology.

The paper is organized as follows. In Section 2, we provide background on monolayer-based nanoscale tribology and an overview of the MoSDeF library. In Section 3 we provide specific details into the usage of MoSDeF for system setup and initialization of monolayer films and an overview of the simulation and analysis methodologies. Section 4 presents results of the screening and the development of predictive models of tribology using machine learning methods. Section 5 presents concluding remarks.

2. BACKGROUND

2.1. Monolayer Lubrication. 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.[41,42] 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.[35−38] A favorable characteristic of monolayer lubricants is their highly tunable chemistry, which can be altered in numerous ways, such as (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., multicomponent monolayers), and (3) by functionalizing opposing surfaces with monolayers of different chemistries. This presents a rich chemical parameter space that can be investigated to optimize monolayer chemistry for tribological applications, and considerable efforts have been made to understand how chemical changes influence frictional response. For example, it has been observed that increasing the backbone chain length yields reduced frictional forces[43−46] attributed to an increase in attractive interchain van der Waals (VDW) forces with longer backbones.[47] 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.[48] Monolayers terminated with fluorinated moieties have shown both increased[49] and decreased[50] 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.[45,51−53] A CH_3-functionalized friction force microscope tip in contact with either an OH- or COOH-terminated monolayer resulted in a lower COF than the same tip in contact with a CH_3-terminated monolayer.[45] 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 nontrivial; 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, atomic force microscopy (AFM) and tribometry, two common techniques, feature probes with radii of curvature that differ by several orders of magnitude; it has been shown that even slight differences in probe shape can enable different mechanisms of energy dissipation in monolayer films.[54] As a result, MD simulation has become a useful tool for probing the tribological properties of monolayers,[47,49,52,54−57] affording atomic-level resolution and direct control over system variables. MD can be utilized for systematic, self-consistent variation of monolayer chemistry in order to identify both tribologically favorable chemistries, as well as chemistry-property relationships that can provide useful insight for the design of more favorable monolayers, as will be the focus of this work.

2.2. Molecular Simulation and Design Framework (MoSDeF). MoSDeF[25] is an open-source Python library made up of a set of composable/modular tools, where each "module" is designed such that it can be used within MoSDeF, or as a standalone package. This modular approach allows MoSDeF to be more easily modified, tested, extended, and have fewer bugs[58] than a monolithic approach. MoSDeF is built using concepts from the computer science field of model integrated computing (MIC).[59,60] MIC is a systems engineering approach that focuses on the creation of domain specific modeling languages to capture the essential features of the individual components of a given process, at the level of abstraction that

https://dx.doi.org/10.1021/acs.jctc.9b01183
J. Chem. Theory Comput. 2020, 16, 1779−1793

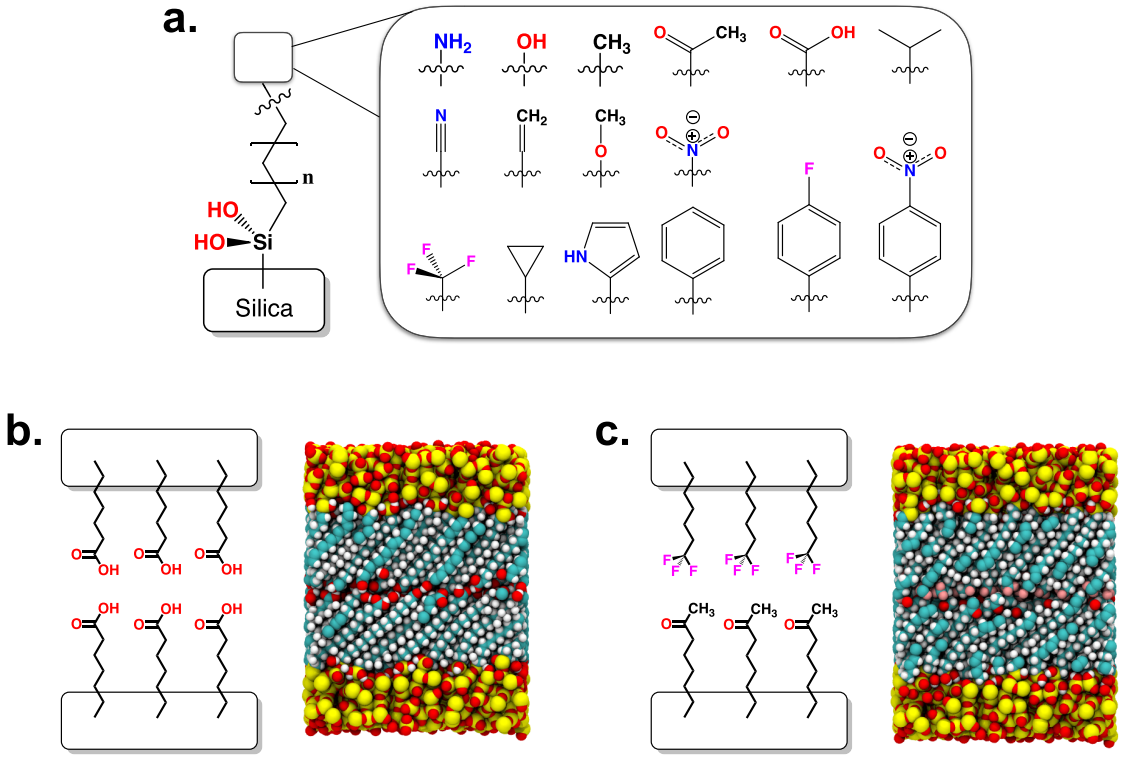


Figure 1. a.) Overview of the chemical parameter space examined in this work. The pool of terminal group chemistries consists of (row 1) amino, hydroxyl, methyl, acetyl, carboxyl, isopropyl; (row 2) nitrile, vinyl, methoxy, nitro; and (row 3) perfluoromethyl, cyclopropyl, 2-pyrrole, phenyl, fluorophenyl, and nitrophenyl. Representative dual-monolayer systems include both b.) chemically identical and c.) chemically dissimilar compositions, shown as both schematics and simulation renderings, where oxygen is colored red, silicon is colored yellow, carbon is colored cyan, hydrogen is colored white, and fluorine is colored green.

is appropriate for the end users. Two key tools are part of the MoSDeF library, namely mBuild and Foyer; MoSDeF interfaces readily with the Signac workflow management system developed by the Glotzer group,[33] to handle job submission and data workspace management. The combination of these tools allows all simulation initialization and parametrization procedures, including variation in chemical space, to be captured in a single Python script, that can be embedded into a Signac project, allowing the exact inputs and procedures to be encapsulated and preserved. While details of the underlying tools have been presented elsewhere,[26,29,30] they have seen continued development and improvement since their introduction, in particular to facilitate the interoperability between mBuild, Foyer, and Signac, as well as with other community tools and simulation engines; this development has allowed for the novel combination of tools to facilitate large-scale screening of monolayer films and development of machine learning models relating tribology and chemistry, as will be the focus of this work.

Briefly, the mBuild Python library,[27,28] originally presented in ref 26, is a general, customizable tool for constructing arbitrarily complex system configurations in a programmatic fashion (i.e., scriptable), designed such that complex systems are built-up from smaller, interchangeable pieces, following the concepts from generative, or procedural, modeling. The power of the mBuild software lies in its underlying Compound data structure; a Compound is a general "container" that can describe effectively anything: an atom, a collection of atoms, a molecule, a collection of Compounds, etc.; Compounds can also include functions that modify the underlying data (e.g., a function to perform polymerization). The same base operations (e.g., translation, rotation, duplication, etc.), can be applied to any Compound using the same syntax, regardless of what the Compound represents. To allow individual Compounds to be joined, Ports are used to define both location and orientation of a possible connection; for atomistic systems, the number of Ports and their locations/orientation typically represent the underlying chemistry. A user (or algorithm) simply states which Ports on two Compounds should connect, and mBuild automatically performs translations and orientations, creating a new (composite) Compound via an equivalence transform (see Klein et al.[26] for more details). No information is lost when joining two Compounds (i.e., the underlying Compounds and Ports are preserved), allowing such connections to also be removed in a reversible fashion. This design approach allows for declaratively expressing repetitive structures, such as polymer chains and planar tilings, essential for the creation of monolayer films. Because Compounds are general containers, they can be easily swapped, making it trivial to change characteristics such as the repeat unit of a polymer or a terminal functional group. Thus, mBuild allows for significant reuse of individual Compounds and data generation routines, enabling significant modifications to system chemistry to be performed with minimal or even no changes to the initialization routines. Thus, one can programmatically vary system parameters, an essential functionality required to perform screening studies as a function of chemistry, the focus of the work presented here. Detailed, up-to-date documentation and source code of mBuild can be found online.[27,28]

The Foyer Python library[31,32] is a general tool for applying force fields to molecular systems (i.e., atom-typing), that

https://dx.doi.org/10.1021/acs.jctc.9b01183
J. Chem. Theory Comput. 2020, 16, 1779−1793

provides a standardized approach to defining chemical context and atom-typing rules; a detailed explanation of the underlying concepts of the software is presented in refs 29 and 30. Foyer does not hard-code rules into the source code and does not rely on a complex hierarchy of nested conditional statements to determine the appropriate atom-type. In Foyer, force field parameters and their usage rules are encoded within a single XML-formatted file that extends the OpenMM[61] XML force field file format. Usage rules are encoded using SMARTS[62] to define the chemical environment for which an atom type applies, along with "overrides" statements to set rule precedence, 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 appropriate atom-types are determined by matching chemical environments to the patterns defined by the SMARTS usage rules. This file format also supports digital object identifiers (DOIs) corresponding to the source of the parameters, to aid in reproducibility. By encoding parameters and usage rules in a separate XML file, the Foyer software does not need to change when the rules themselves change or are extended, making it easier for the rules to be archived, shared, updated, customized, and disseminated in a publication, enhancing reproducibility, and enabling TRUE simulations. Detailed, up-to-date documentation and the source code of Foyer can be found online.[31,32]

3. MODEL AND METHODS

3.1. Initialization of System Configurations Using the MoSDeF mBuild Library. Systems in this work consist of two opposing monolayer films attached to amorphous silica substrates, consistent with models used in earlier work;[56,57] crystalline surfaces have been used in other work;[47,50,63,64] however, the in-plane ordering of attachment sites has been shown to exaggerate chain orientational order.[56] The surfaces used here have an atomic-scale roughness of ∼0.11 nm, found to closely match prior work that explicitly considered a surface oxidation treatment typical of experiment;[57] most prior studies that have used amorphous silica substrates have considered atomically smooth surfaces, which earlier work suggests influences the behavior and trends.[54,56,57] The procedure for carving the amorphous surfaces builds upon procedures found in the literature[65,66] and used in prior work[54] (detailed in the Supporting Information, including links to the associated mBuild source code). Systems are constructed using the mBuild library 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), (3) any external facing oxygens that are a part of the silica interface and *not* consumed by one of the duplicated chains are then hydrogenated, mimicking surface oxidation, and (4) the monolayer is duplicated, rotated by 180 deg, and shifted to yield the complete system. Chain prototypes are generated such that both the backbone chain length and the functional group used to terminate the chain are tunable, thus enabling screening to be easily performed and abstracted for end user consumption. Listing 1 includes a snippet of the script used to generate the monolayers in this work, and the approach is shown graphically in Figure 1. Using concepts from the field of MIC, mBuild is designed such that the underlying code can be presented at a level that captures the appropriate details/language to describe the given system and processes; specifically, in Listing 1, SilicaInterface provides a clear way to define the silica substrate, Alkylsilane provides a way to define the molecule to be grafted, and SurfaceMonolayer provides a clear way to define how these the surface and chain are associated and modified to create a monolayer. Due to the hierarchical, component-based framework of mBuild, these routines can be trivially modified to allow screening to be extended to additional chemical degrees of freedom, such as backbone monomer, multicomponent films, etc., which can be considered in future studies. Scripts for the initialization of films in mBuild (i.e., the specific routines underling listing 1) are available online,[67,68] ensuring the procedure is transparent and reproducible thus meeting the definition of a TRUE simulation. In all cases, monolayers consist of 100 chains attached to an amorphous silica substrate of dimensions 5 nm × 5 nm, with a surface density of 4.0 chains/nm^2, consistent with prior studies.[69,70]

The pool of terminal group chemistries examined in this work is shown in Figure 1a and was chosen to satisfy two primary criteria: (1) functional group should be able to be described using existing parameters within the OPLS force field (described in further detail below) and (2) the entire functional group pool should span a wide range of chemical characteristics (e.g., size, shape, polarity) to facilitate chemistry/property analysis. Several of the chosen functional groups may not be synthesizable or may readily react following synthesis.[71] However, as reactions are not considered in the simulations, since classical force fields with fixed bonds are used, such groups remain stable. Furthermore, as the goal is to study a chemically diverse range of terminal group chemistries, such systems still provide valuable data in terms of chemistry-property relationships.

Screening of dual-monolayer systems was performed over two distinct parameter spaces: (1) chemically identical systems (Figure 1b, where the top and bottom monolayer films feature the same chemistry) as a function of terminal group and chain length and (2) chemically dissimilar systems (Figure 1c, where the top and bottom monolayer films feature different chemistries) for a single fixed chain length of 17 carbons. For each chemistry, five systems were generated, each corresponding to a unique arrangement of chains on the available binding sites on the silica surface, by changing the "seed" provided to SurfaceMonolayer (see Listing 1). 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 Figure 1a were considered along with five chain lengths (5, 8, 11, 14, and 17 backbone carbons, excluding the terminal group), and three different normal loads (5, 15, 25 nN) resulting in 80 unique chemistries and 1200 simulations in total. Chemically dissimilar films feature backbone chain lengths of 17 carbons and include select combinations of seven terminal groups (carboxyl, fluorophenyl, hydroxyl, isopropyl, methyl, nitro, and perfluoromethyl) with the 16 terminal groups in Figure 1a (ignoring duplicates already considered for chemically identical films), for an additional 84 unique chemistries and 1260 total simulations. For the development of machine learning models, the set of 100 unique terminal group combinations (both chemically identical and dissimilar) with a chain back length of 17 carbons is considered. Signac workflows that define this parameter space and all the MoSDeF routines are available online for chemically identical[72,73] and chemically dissim-

https://dx.doi.org/10.1021/acs.jctc.9b01183
J. Chem. Theory Comput. 2020, 16, 1779−1793

ilar[74,75] monolayer systems. To evaluate the predictive nature of the models, toluene and phenol terminal groups are considered and placed in contact with opposing monolayers composed of the original 16 terminal groups (see Figure 1a) with a chain length of 17 carbons. This provides 30 additional unique chemistries (phenyl-phenol and pyrrole-phenol are not considered due to instability of the simulations under high normal loads); 3 normal loads (5, 15, 25 nN) are considered and 3 replicates for each of these systems, for a total of 270 additional simulations; Signac workflows that define this parameter space for these systems can be found online.[76,77] In sum, 2730 simulations are therefore performed, with 194 unique dual monolayer chemistries with the aid of MoSDeF.

```
#extract screening variables from Signac
chainlength = job.statepoint()['chainlength']
n_chains = job.statepoint()['n']
seed = job.statepoint()['seed']
terminal_group = job.statepoint()['terminal_group']

#Generate amorphous silica interface
surface = SilicaInterface(thickness=1.2, seed=seed)

#Generate prototype of functionalized alkylsilane chain
chain_prototype = Alkylsilane(chain_length=chainlength,
                              terminal_group=terminal_group)

#Create monolayer on surface, backfilled with hydrogen caps
monolayer = SurfaceMonolayer(surface=surface, chains=chain_prototype,
                             n_chains=n_chains, seed=seed,
                             backfill=H(), rotate=False)

#Duplicate to yield two opposing monolayers
dual_monolayer = DualSurface(monolayer, separation=2.0)

#Make sure box is elongated in z to be pseudo-2D periodic
box = dual_monolayer.boundingbox
dual_monolayer.periodicity += np.array([0, 0, 5. * box.lengths[2]])

#define location of the Foyer force field file
forcefield_dir = resource_filename('atools', 'forcefields')

#atom-type and save to GROMACS input
dual_monolayer.save('init.gro', overwrite=True)
dual_monolayer.save('init.top',
    forcefield_files=os.path.join(forcefield_dir, 'oplsaa.xml'),
    combining_rule='geometric', overwrite=True)

#atom-type and save to LAMMPS input
dual_monolayer.save('init.lammps',
    forcefield_files=os.path.join(forcefield_dir, 'oplsaa.xml'),
    combining_rule='geometric', overwrite=True)
```

Listing 1. Code snippet of the Python code used by Signac, that calls various mBuild and Foyer functions, to initialize the systems for screening.

3.2. Application of Force Fields Using the MoSDeF Foyer Library.

After creation of the monolayer structures, the systems are atom-typed, i.e., the appropriate force field parameters are determined for the bonded and nonbonded interactions of the system using the Foyer library[29,30] in MoSDeF. The OPLS all-atom force field is used.[78−84] The specific OPLS version used was that provided with the Gromacs 5.1 distribution,[85] in addition to parameters for silica that were obtained from Lorenz et al.[55] 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. All parameters are provided in the Supporting Information and in Foyer compatible XML file format in the associated GitHub repositories.[72−77] As such, the parametrization of the models is transparent and reproducible, again meeting the definition of a TRUE simulation.

3.3. Simulation Methods.

MD simulations are performed using the GROMACS molecular dynamics engine (version 5.1.0 for the initial screening; model validation for toluene and phenol use version 2018, as this provides increased performance).[85] An initial, short, distance-limited (i.e., particles were not allowed to move more than 0.1 Å per time step) NVE simulation is performed using the LAMMPS molecular dynamics engine[86,87] (March 2018 version) to remove overlaps between terminal groups from the initial configuration. After converting the final structure from the LAMMPS to GROMACS format, simulations are executed using GROMACS in four stages: (1) energy minimization, (2) equilibration, (3) compression, and (4) shear. Energy minimization is performed using a steepest descent algorithm, where minimization is stopped once the maximum force on any atom is less than 10 kJ/mol/nm. Following energy minimization, equilibrium molecular dynamics simulations are performed, whereby monolayers and surfaces (excluding the outer 4 Å of each surface which is kept rigid) are allowed to relax over 1 ns to reach a low-energy state, using a time step of 1 fs. Following relaxation, a force of 5 nN is applied to the bottom surface in the z direction (i.e., toward the top monolayer) to bring the two monolayers into contact, where they are compressed over 0.5 ns, allowing the intersurface distance to reach a steady-state value. Beginning from snapshots at the end of the compression stage, three independent simulations are performed where monolayers are sheared under normal loads of 5, 15, and 25 nN (corresponding to pressures of 200, 600, and 1000 MPa, respectively), consistent with normal loads used for such systems in the literature.[49,88,89] To maintain a constant normal load a constant force is applied to the bottom silica surface in the z direction. Shear is introduced by coupling the top surface to a ghost particle via a harmonic spring (with a spring constant of 10,000 kJ/(mol·nm^2)) and pulling the ghost particle in the x direction at a rate of 10 m/s. Shear is performed for 10 ns, where the final 5 ns is used for sampling and analysis. All MD simulations are performed under the NVT ensemble using a Nose-Hoover thermostat[90,91] with damping coefficient (tau-t) set to 1.0 ps 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,[56] agreeing with reports in the literature for similar systems,[88] thus application of a standard thermostat is justified. Bonds to hydrogen atoms are constrained using the LINCS algorithm,[92] 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[93,94] is used for long-range electrostatics, using a force and pressure correction in the z-dimension to support slab geometries; systems are periodic in the monolayer plane.

The MD workflow is maintained and executed using the Signac workflow manager (v0.5.4), a part of the Signac framework.[33] With Signac-flow, individual molecular dynamics operations (e.g., equilibration) are wrapped into Python functions. These operations are 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, and Signac managing the dataspace on the filesystem. All scripts used to execute simulations and analysis, along with initialization scripts and force fields, are freely available online,[67,68,72−77] ensuring that the full and exact procedures

https://dx.doi.org/10.1021/acs.jctc.9b01183
J. Chem. Theory Comput. 2020, 16, 1779−1793

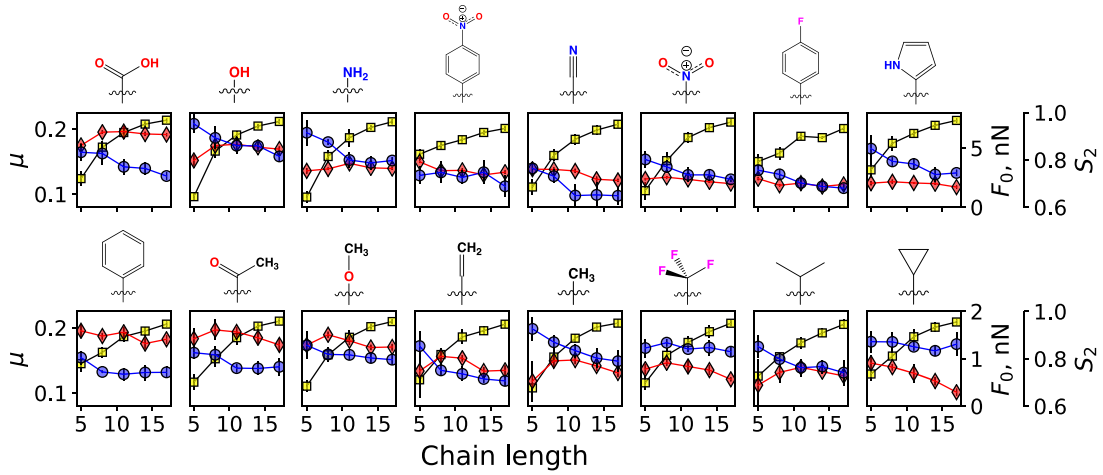


Figure 2. Coefficient of friction (μ, blue circles), adhesive force (F_0, red diamonds), and nematic order (S_2, yellow squares) for chemically identical systems shown for each terminal group chemistry as a function of chain length. Note the scale for F_0 differs between the two rows in order to clearly 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.

and system inputs are transparent, and thus the simulations can be reproduced, used by others, and extended (i.e., the simulations are TRUE).

3.4. 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[95]

$$F_f = F_0 + \mu \cdot F_n \tag{1}$$

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, 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 25 nN), 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 unique chemistry (i.e., unique combination of terminal groups and/or chain length) represent the average of the five systems with different chain configurations on the surface. Monolayer structure is examined via the nematic order parameter,[54,56,57] which provides a measure of the global orientational order of the monolayer backbones in each film (note, the two monolayers in each system are considered separately and the average is reported). 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

$$S_2 = \left\langle \frac{3}{2}\cos^2\theta - \frac{1}{2} \right\rangle \tag{2}$$

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 toward 0 indicate increased disorder, with a value of ~0.8 visually correlating with a disordered film. Reported values represent the average of the top and bottom films. All analysis routines can be found online;[67,68] the management and execution of the analysis routines are also handled by the Signac (v0.8.5) and Signac-flow (v0.5.4) Python packages within the Signac framework.[33]

4. RESULTS AND DISCUSSION

4.1. Screening Chemically Identical Monolayers as a Function of Backbone Length and Terminal Group Chemistry. The effect of backbone chain length is explored by calculating COF, adhesion force, and nematic order for chemically identical monolayer film systems with chain lengths of 5, 8, 11, 14, and 17 backbone carbons; this is presented in Figure 2 for each terminal group chemistry (numerical values are included in the Supporting Information in Tables S1−S5). In general, the COF is found to decrease as the backbone chain length is increased for each terminal group chemistry. Similar trends, primarily concerning methyl-terminated films, are reported in the literature[43−47,96,97] demonstrating that the addition of carbons to the chain backbone correlates with a reduction in the frictional forces.[43−46] Experiments of methyl terminated alkylsilanes, examined a similar range to the simulations in this work, e.g., Xiao et al.[98] considered backbone lengths of 2, 5, 7, and 17, and Lio et al.[44] considered 5, 7, 11, and 17. Figure 2 shows that this trend is stronger for some terminal group chemistries than others, where for example benzene-terminated monolayers show little reduction in COF over the chain lengths considered. The nematic order parameter for the intermediate normal load of 15 nN is reported in Figure 2 as a function of chain length, whereby monolayer nematic order is found to increase as the backbone chain length is increased, consistent with prior work.[57] 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 correspond to those systems with "bulkier" terminal groups (e.g., fluorophenyl, nitrophenyl, benzene) for which steric interactions between the terminal groups likely promote more ordered films. The adhesion force is also plotted in Figure 2 and is found to be less dependent on chain length than COF and instead is primarily dependent on terminal group chemistry. For all systems examined, adhesion either decreases or remains at roughly the same value as chain length

https://dx.doi.org/10.1021/acs.jctc.9b01183
J. Chem. Theory Comput. 2020, 16, 1779−1793

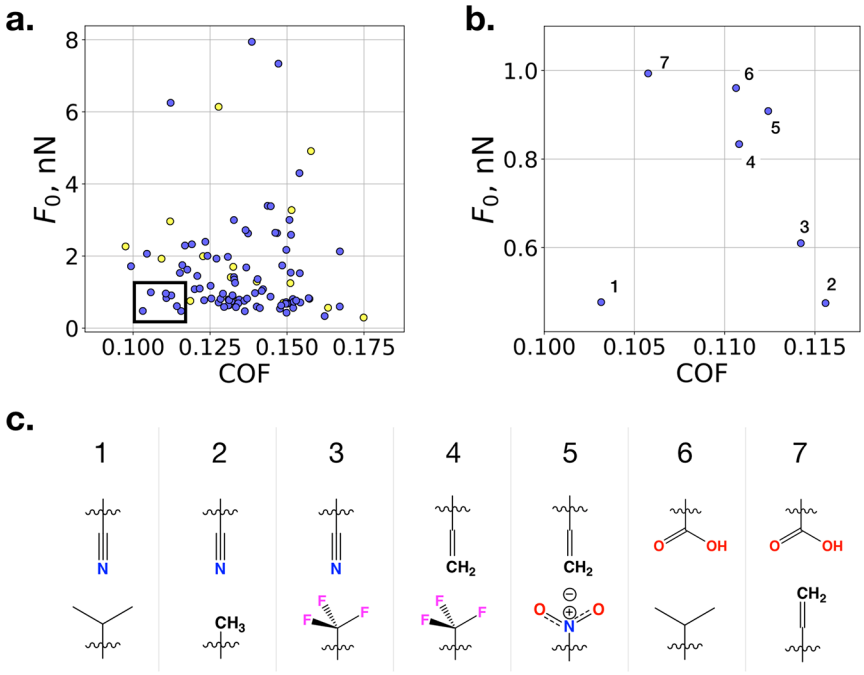


Figure 3. (a) Scatter plot of COF and adhesion force data for chemically identical systems (yellow) and chemically dissimilar systems (blue). The highlighted region indicates the area featuring the most tribologically favorable systems. (b) Plots the data from the highlighted region, where the seven systems are annotated and (c) provides the corresponding terminal group chemistries in the favorable region. These include (1) nitrile-isopropyl, (2) nitrile-methyl, (3) nitrile-perfluoromethyl, (4) vinyl-perfluoromethyl, (5) vinyl-nitro, (6) carboxyl-isopropyl, and (7) carboxyl-vinyl.

is increased from 8 to 17 backbone carbons (i.e., excluding the shortest chain length that tends to show low film order). While there does not appear to be an obvious relationship between the magnitude of the COF and terminal group chemistry, a strong correlation between terminal group polarity and adhesion is observed. Systems with the largest adhesion forces are all polar molecules, and the groups with the smallest adhesion forces are all nonpolar (see Figure S1 in the Supporting Information). This is consistent with the literature,[45,52] where, for example, experiments showed that the adhesion was higher for carboxylic-terminated monolayers with a backbone length of 10, as compared to equivalent length methyl terminated monolayers.[45]

4.2. Screening of Chemically Dissimilar Monolayer Films as a Function of Terminal Group Chemistry. In Figure 3a the COF and adhesion results for systems of both chemically identical (16 in total) and chemically dissimilar (84 in total) films with a backbone length of 17 carbons are shown; numerical values for chemically dissimilar films are included in the Supporting Information in Table S6. From a tribological standpoint, the ideal monolayer chemistry should feature both a low COF and low adhesive force, and thus favorable chemistries would exist in the lower left-hand corner of Figure 3a. Several chemically dissimilar systems appear to provide favorable tribological properties as compared to chemically identical systems, as highlighted in Figure 3b and c. The majority of these systems features one monolayer that is polar/hydrophilic and another that is nonpolar/hydrophobic. As an example, while carboxyl-terminated monolayers are found to yield the highest adhesive forces for chemically identical systems (see Figure 2), when paired with a nonpolar counter monolayer the ability to form intermonolayer hydrogen bonds is eliminated, and these combinations of terminal groups are able to provide favorable COF and adhesion values, consistent with prior studies in the literature.[45,99,100] In particular, favorable tribological properties are found for monolayers featuring nitrile and vinyl terminal groups, as shown in Figure 3c. While nitrile and vinyl have several differences (chemical composition, polarity), both groups are linear and feature a cylindrical shape with a small VDW radius, and a more rigid nature owed to the presence of a double or triple bonds.

4.3. Implementation of Machine Learning and QSPR Modeling for Monolayer Tribology. To further examine the relationships between chemistry and tribology, a Python library is developed that provides a framework for the creation of topological quantitative structure−property relationship (QSPR) models using the random forest machine learning method for monolayer films. The routines and procedures used to perform this analysis and model development are available online.[101−104] Here, the approach taken for the generation of these models is described. These models are given a "fingerprint" for each system as input, 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), complexity (e.g., Chi indices), and charge distribution (e.g., topological polar surface area), as categorized in Figure 4.

As shown in Figure 5a−d, "molecular fingerprints" are first calculated for the terminal groups of each monolayer (i.e., we do not consider the complete monolayer system or any properties measured from it). Here, we utilize two separate molecules to represent terminal groups (as shown in Figure 5b): (1) a hydrogen atom caps the attachment site where the terminal group is attached to the chain backbone and is used in the calculation of descriptors that relate to molecular shape, and (2) a methyl group caps this attachment site and is 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 should provide a reasonable estimate of the

https://dx.doi.org/10.1021/acs.jctc.9b01183
J. Chem. Theory Comput. 2020, 16, 1779−1793

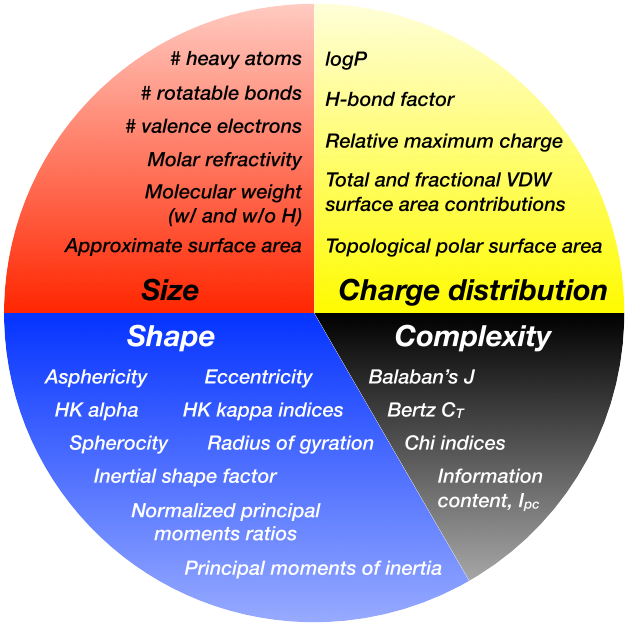


Figure 4. Features considered for the QSPR models, grouped into clusters describing charge distribution, complexity (i.e., degree of connectivity/branching), shape, and size. Additional descriptions of these features can be found in the Supporting Information in Table S7.

distribution of charge. Note, chain backbone length is not considered in the model presented herein, as the data set we are focusing on only includes systems with backbone lengths of 17 carbons.

The approach developed in this work for molecular fingerprinting relies only upon information on the molecular bond graph as an input. Molecules are provided to RDKit in the notation of the simplified molecular-input line-entry system (SMILES),[105] shown in Figure 5c. 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; that is molecular fingerprints do not depend upon simulation data. After providing RDKit[39] 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 Figure 5d. While the majority of these descriptors is 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.[106] For descriptors requiring information concerning molecular charge, charges are assigned via RDKit to the molecular graph using the approach of Gasteiger and Marsili.[107] We have additionally implemented an additional descriptor, termed the "hydrogen bond factor" that provides a relative estimate of the maximum availability for intermonolayer 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

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

$$hb_{tb} = \begin{cases} 0, & \min(d_b, a_t) = 0 \\ \max(d_b, a_t), & \min(d_b, a_t) > 0 \end{cases}$$

$$hbonds = hb_{tb} + hb_{tb} \tag{3}$$

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 (i.e., those presented in Figure 4), including brief descriptions of each, is provided in the Supporting Information in Table S7. 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 Figure 5e). 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,[108] which also helps reduce complexity of the QSPR models.

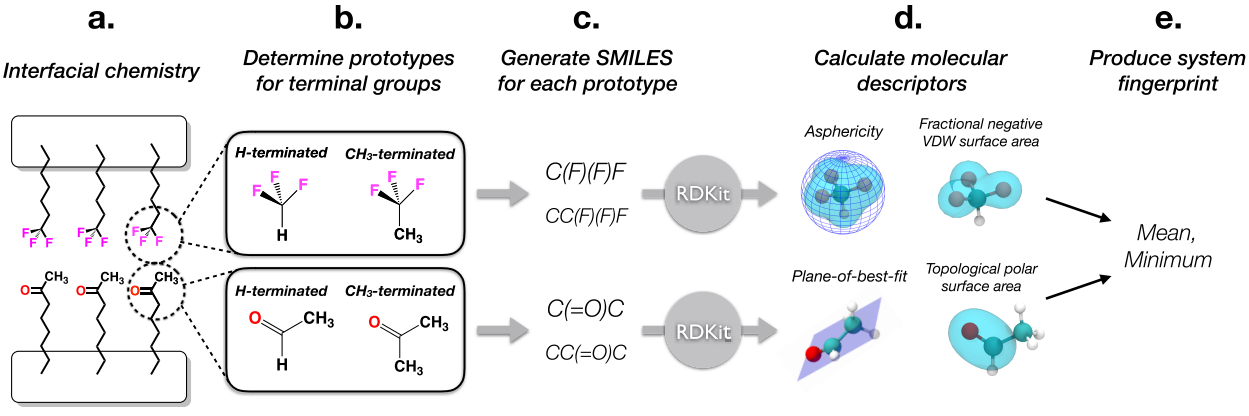


Figure 5. Workflow of the process used to fingerprint each interfacial chemistry. a.) For a given interfacial chemistry (a perfluoromethyl-acetyl system is shown as an example), b.) prototypes are determined for each terminal group. c.) These prototypes are then converted to SMILES representations d). and fed to the RDKit Python package used to calculate a variety of molecular descriptors (see Figure 4). 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.

https://dx.doi.org/10.1021/acs.jctc.9b01183
J. Chem. Theory Comput. 2020, 16, 1779−1793

A random forest regression algorithm[109,110] is used for QSPR models, as implemented in the scikit-learn Python package.[40] 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. Predictions from random forest converge for a large number of decision trees,[111] thus, 1000 trees are used in this work. The random forest algorithm also allows contributions to the model from the various molecular descriptors to 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 chemistries and tested with the remaining 20). In this work, 5 different models are constructed using different randomly chosen sets where the mean and standard deviation of the models are used for prediction of the tribological properties (i.e., test-train splits); note, COF and force of adhesion are each trained independently for each model but use the same descriptors and subset of data for training. 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.

4.4. Model Evaluation. 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 OOB samples, for systems with a backbone length of 17 carbons. It is observed from Table 1 that the models for both COF and adhesion perform consistently well and do not depend heavily on the different random subset of the data used (see Table S8 for more details). The random forest algorithm is known to overfit to the training data, so the high R^2 values observed for the training data sets are expected and alone do not provide a great metric for the performance of the models. R^2 values above 0.6 are observed for both the test data as well as for the OOB samples, providing more reliable estimates of model performance and suggesting respectable predictive power. 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 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); this is most obvious for the force of adhesion, likely related to the fact only a few data points exist in the higher adhesion regime. However, the evaluation metrics obtained for the OOB samples are found to remain mostly stable between the five models for COF and force of an adhesion, providing greater confidence in the results and in the models' predictive power.

To provide visual evaluation of the models for COF and adhesion, Figure 6a,b shows both the training data and the test data used to generate and evaluate Model 1. For COF, the majority of the test data sits close to the $y = x$ line, with no significant deviations with most data falling within the error of the values calculated from the simulations. For the adhesion

Table 1. Evaluation of Random Forest Regression Models for COF and F_0[a] for Systems with a Carbon Backbone Length of 17 for the Training, Out of Bag (OOB) Cross Validation and Test Sets

		training set			cross-validation (OOB)			test set		
model no.	target variable	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
av	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

[a]Evaluation metrics reported are the coefficient of determination between predicted values and those calculated from the simulations (R2), root-mean-square error (RMSE) and mean absolute error (MAE).

https://dx.doi.org/10.1021/acs.jctc.9b01183
J. Chem. Theory Comput. 2020, 16, 1779−1793

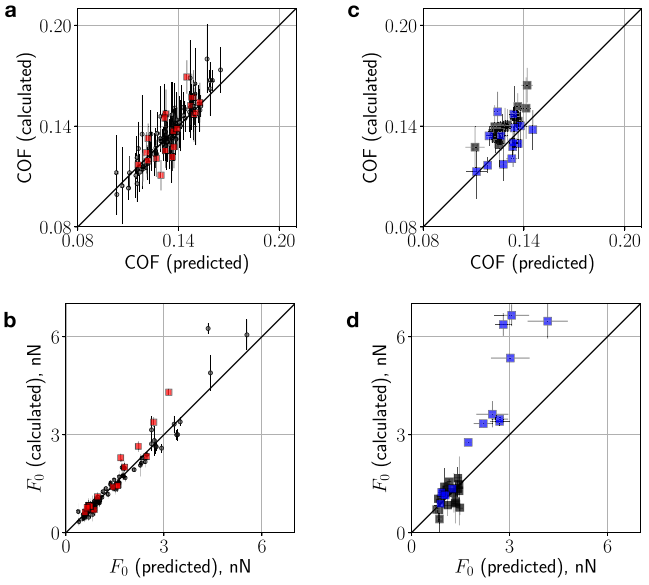


Figure 6. Values predicted by QSPR Model 1 for a.) COF and b.) adhesive force compared to the values calculated from simulation for systems with 17 carbon backbones. The $y = x$ line is drawn in black for reference. Black circles denote data used as part of the training set, while red squares denote data that was part of the test set. Comparison of predicted values and those calculated from simulation of c.) COF and d.) adhesive force for phenol (black) and toluene (blue).

force shown in Figure 6b, the model's predictive capacity is best for systems with lower values of adhesion, e.g., less than 4 nN, again, possibly due to the fact that only a few systems with such high adhesion are considered. Systems in this higher adhesion region correspond to those that feature intermonolayer hydrogen bonds. Fortunately, accuracy in this regime is of least concern, as interfacial chemistries featuring large adhesive forces would be considered poor candidates for lubricating films due to their higher likelihood to exhibit stiction, and thus the model still properly differentiates between low and high adhesion values, allowing these unfavorable systems to be correctly screened out. Similar plots evaluating Models 2−5 are included in Figures S2−S5, and numerical values of the mean predictions of the 5 Models for each system are included in Table S6. We note that the models presented here are only trained with data for a backbone length of 17 and thus are not designed to predict the tribological properties for other chain lengths. The framework could be adapted to take into account chain length, but this would require collection of additional data as a function of chain length in order to generate a sufficiently large training set. However, as seen in Figure 2, adhesion is generally not strongly dependent on chain length, and thus we expect that the values of the model would provide reasonable estimates even for shorter chain lengths. Similarly, Figure 2 shows that the COF tends to increase as chain length is reduced; while we would not expect the current models to quantitatively capture the COF values for shorter chain lengths, we would reasonably expect qualitative trends to hold with regards to which terminal group combinations provide favorable tribological properties.

To further evaluate the utility of the models in Table 1, and in particular their ability to prescreen parameter space, simulations are performed for toluene and phenol terminal groups (not part of the original training set) in contact with the initial 16 terminal groups (see Figure 1a). Figure 6c,d plots predicted vs simulated values of COF and adhesion, where the predicted values represent the average of the predictions of the 5 models in Table 1 and error bars correspond to the standard deviation of these 5 predictions (numerical values of both the simulated and predicted values are tabulated in Table S9). The models predict the COF with reasonable accuracy, where most of the predictions are within the standard error of the simulated values and data points visually track the line $y = x$. The ability to predict the force of adhesion values is similar to that observed in Figure 6b, in that values are well predicted in the lower adhesion regime but underpredicted for higher values; specifically, quantitative agreement is not observed for phenol under conditions where significant hydrogen bonding can occur between the two layers, although, as before, the model does capture the appropriate trends. The QSPR models themselves are hosted on GitHub[103] and archived on Zenodo,[104] along with all of the code related to this work, allowing for transparency and the extension/use in future studies that explore the monolayer chemical parameter space.

This validation provides evidence as to the predictive nature of the models and that correlations between chemistry and tribology used by the machine learning model are appropriate. As such, the "feature importances" are extracted from the models to evaluate which elements of the system fingerprint have the most influence over each variable. Figures 7 and 8 provide 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 to Figure 5e). For COF, the majority of the top contributing descriptors is those which describe mean values between the two terminal groups (for additional details on mean-based modeling, see Figure S6). Furthermore, it is observed that the majority of features with high contributions to the model are those that 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[112]

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

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 Figure 7b to have low COF, where those terminal groups that feature the lowest Hall-Kier alpha values are those with aromatic rings and to a lesser extent carbonyl-containing groups and the nitrile group. Dependence of the model on the "inertial shape factor" (ISF) is also observed, where ISF is a ratio of the systems principal moments of inertia (described further in Table S7). A negative

https://dx.doi.org/10.1021/acs.jctc.9b01183
J. Chem. Theory Comput. 2020, 16, 1779−1793

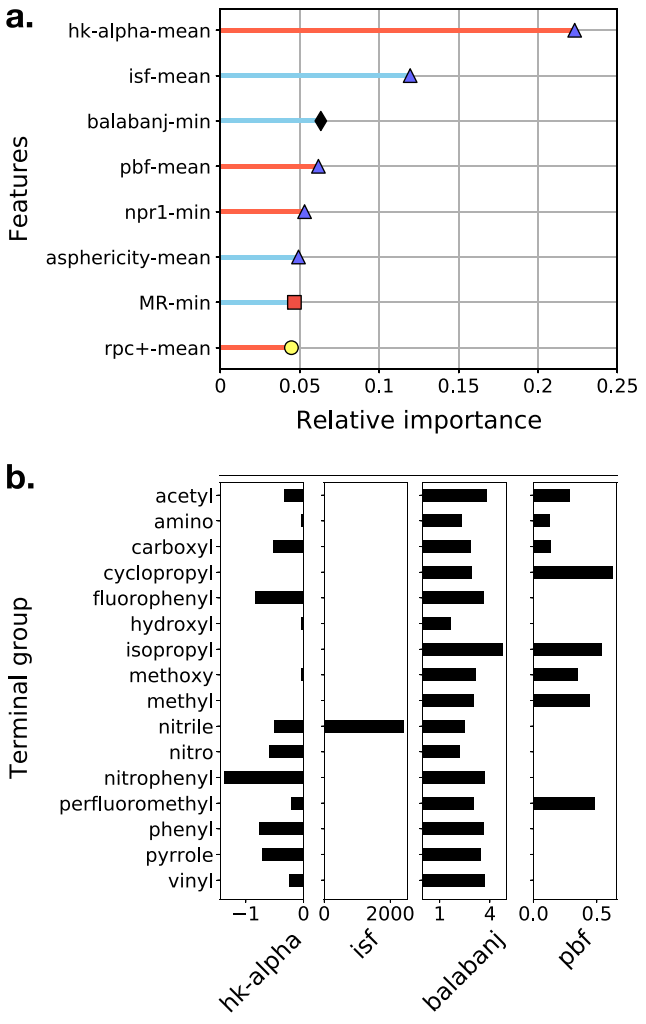


Figure 7. a.) Relative feature importances extracted from the random forest regression model (Model 1 in Table 1) for COF and b.) values of the top four contributing features for each terminal group chemistry. Symbols in (a.) represent the molecular aspect characterized by each feature (yellow circles: 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.

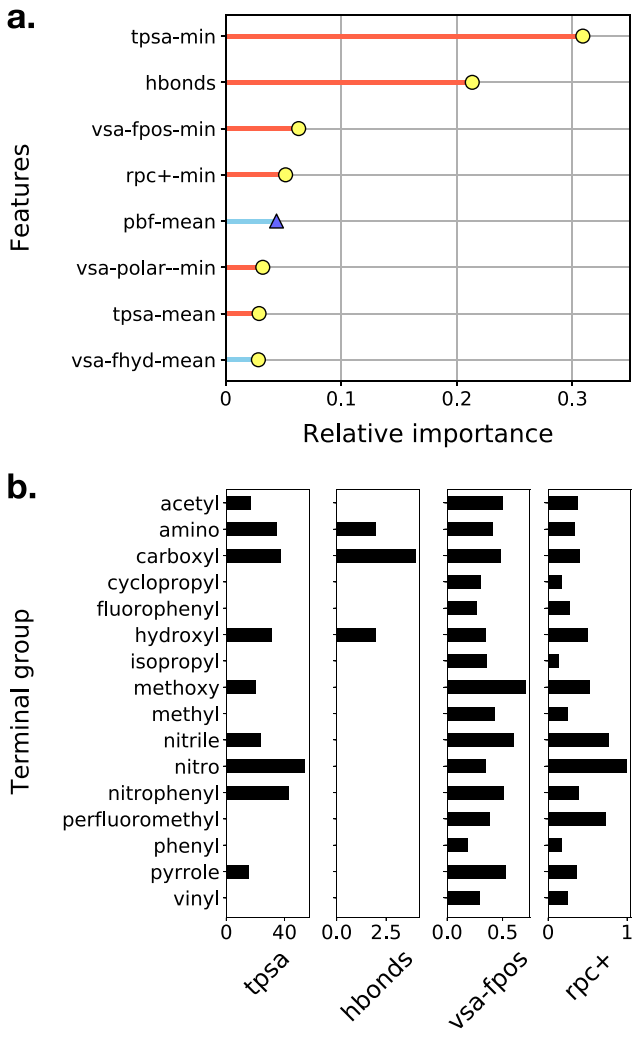


Figure 8. a.) Relative feature importances extracted from the random forest regression model (Model 1 in Table 1) for adhesion and b.) values of the top four contributing features for each terminal group chemistry. Symbols in (a.) represent the molecular aspect characterized by each feature (yellow circles: 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.

correlation is observed between ISF and COF, and 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 for the low COF of systems containing a nitrile group (which was observed in Figure 3c). 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 Figure 8a. Following expectations, nearly all of the top contributing molecular descriptors provide measures of the distribution of charge on the terminal group. 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 and 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 Figure 8b, 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 Figure 3c. A modest contribution to the QSPR model is also observed from the "hbonds" descriptor. This descriptor characterizes a system's ability to form intermonolayer hydrogen bonds, and

https://dx.doi.org/10.1021/acs.jctc.9b01183
J. Chem. Theory Comput. 2020, 16, 1779−1793

the positive correlation indicates that the more combinations of hydrogen bond donors and acceptors the higher the adhesive force should be. While systems featuring intermonolayer hydrogen bonding are found to not have their adhesive forces predicted with high accuracy, as evidenced by Figure 6, the addition of the "hbonds" descriptor does allow the model to achieve a qualitatively correct prediction that such systems will feature high adhesion.

5. CONCLUSION

In this work, the screening of functionalized monolayer films is enabled by use of the MoSDeF software suite and Signac framework. 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 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.[113] The breadth of the data in this study has afforded the ability to use machine learning to generate QSPR models, which require only a simple SMILES representation of the terminal group as an input and are found to yield reasonable predictive capability. Feature extraction from these models reveals that the 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 intermonolayer 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., multicomponent films) which may provide even more favorable tribological properties. Finally, through the SI and codes deposited in open-source repositories[67,68,72−74,76,77,101−104] and enabled by the use of MoSDeF, all of the simulations underpin this paper as TRUE[34] simulations and thus are able to be transferred, reproduced, used, and extended by other researchers.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.9b01183.

Description of all force field parameters in tabular format; short definitions for all molecular descriptors used in QSPR models as well as tables of which specific systems were used in training the different models; details on mean-based models for estimating COF and adhesion of chemically dissimilar systems from results obtained for chemically identical films; description of procedure for carving amorphous silica surfaces; and numerical values of simulated and predicted tribological values presented in figures (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Clare McCabe − *Department of Chemical and Biomolecular Engineering, Multiscale Modeling and Simulation (MuMS) Center, and Department of Chemistry, Vanderbilt University, Nashville, Tennessee 37235, United States;* orcid.org/0000-0002-8552-9135; Phone: 615 322 6853; Email: c.mccabe@vanderbilt.edu

Authors

Andrew Z. Summers − *Department of Chemical and Biomolecular Engineering and Multiscale Modeling and Simulation (MuMS) Center, Vanderbilt University, Nashville, Tennessee 37235, United States;* orcid.org/0000-0001-8477-3059

Justin B. Gilmer − *Multiscale Modeling and Simulation (MuMS) Center and Interdisciplinary Program for Materials Science, Vanderbilt University, Nashville, Tennessee 37235, United States*

Christopher R. Iacovella − *Department of Chemical and Biomolecular Engineering and Multiscale Modeling and Simulation (MuMS) Center, Vanderbilt University, Nashville, Tennessee 37235, United States;* orcid.org/0000-0003-0557-0427

Peter T. Cummings − *Department of Chemical and Biomolecular Engineering and Multiscale Modeling and Simulation (MuMS) Center, Vanderbilt University, Nashville, Tennessee 37235, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.9b01183

Funding

Funding for this work has been provided by the National Science Foundation (NSF) through Grants ACI-1047827, OAC-1835874, and OAC-1535150.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

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. An award of computer time was provided by the INCITE program. This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725. All simulations were performed on the Titan supercomputer at the Oak Ridge Leadership Computing Facility.

■ 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; OOB, out-of-bag

REFERENCES

(1) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. Large-Scale Screening of Hypothetical Metal-Organic Frameworks. *Nat. Chem.* **2012**, *4* (2), 83−89.

(2) Hachmann, J.; Olivares-Amaya, R.; Atahan-Evrenk, S.; Amador-Bedolla, C.; Sánchez-Carrera, R. S.; Gold-Parker, A.; Vogt, L.; Brockway, A. M.; Aspuru-Guzik, A. The Harvard Clean Energy Project: Large-Scale Computational Screening and Design of Organic Photovoltaics on the World Community Grid. *J. Phys. Chem. Lett.* **2011**, *2* (17), 2241−2251.

(3) Schmidt, J.; Marques, M. R. G.; Botti, S.; Marques, M. A. L. Recent Advances and Applications of Machine Learning in Solid-State Materials Science. *npj Comput. Mater.* **2019**, *5* (1), 83.

(4) Ahmad, Z.; Xie, T.; Maheshwari, C.; Grossman, J. C.; Viswanathan, V. Machine Learning Enabled Computational Screening of Inorganic Solid Electrolytes for Suppression of Dendrite Formation in Lithium Metal Anodes. *ACS Cent. Sci.* **2018**, *4* (8), 996−1006.

(5) Afzal, M. A. F.; Haghighatlari, M.; Ganesh, S. P.; Cheng, C.; Hachmann, J. Accelerated Discovery of High-Refractive-Index Polyimides via First-Principles Molecular Modeling, Virtual High-Throughput Screening, and Data Mining. *J. Phys. Chem. C* **2019**, *123* (23), 14610−14618.

(6) Haghighatlari, M.; Hachmann, J. Advances of Machine Learning in Molecular Modeling and Simulation. *Curr. Opin. Chem. Eng.* **2019**, *23*, 51−57.

(7) Sandve, G. K.; Nekrutenko, A.; Taylor, J.; Hovig, E.; Vilo, J. Ten Simple Rules for Reproducible Computational Research. *PLoS Comput. Biol.* **2013**, *9* (10), No. e1003285.

(8) Baker, M. 1,500 Scientists Lift the Lid on Reproducibility. *Nature* **2016**, *533* (7604), 452−454.

(9) *Reproducibility and Replicability in Science*; National Academies Press: Washington, DC, 2019; DOI: 10.17226/25303.

(10) Hjorth Larsen, A.; Jørgen Mortensen, J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dułak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C.; et al. The Atomic Simulation Environment—a Python Library for Working with Atoms. *J. Phys.: Condens. Matter* **2017**, *29* (27), 273002.

(11) Fortunato, M. E.; Colina, C. M. *pysimm*. https://github.com/polysimtools/pysimm (accessed Jan 26, 2020).

(12) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminf.* **2011**, *3* (10), 33.

(13) *Open Babel Github repository*. https://github.com/openbabel (accessed Jan 26, 2020).

(14) Fortunato, M. E.; Colina, C. M. Pysimm: A Python Package for Simulation of Molecular Systems. *SoftwareX* **2017**, *6*, 7−12.

(15) Girard, M.; Ehlen, A.; Shakya, A.; Bereau, T.; de la Cruz, M. O. Hoobas: A Highly Object-Oriented Builder for Molecular Dynamics. *Comput. Mater. Sci.* **2019**, *167* (May), 25−33.

(16) *Hoobas repository*. https://bitbucket.org/NUaztec/hoobas/src/master/ (accessed Dec 9, 2019).

(17) Hedges, L. O.; Mey, A. S. J. S.; Laughton, C. A.; Gervasio, F. L.; Mulholland, A. J.; Woods, C. J.; Michel, J. BioSimSpace: An Interoperable Python Framework for Biomolecular Simulation. *J. Open Source Softw.* **2019**, *4* (43), 1831.

(18) ParmEd *ParmEd documentation*. http://parmed.github.io/ParmEd/html/index.html (accessed Jan 26, 2020).

(19) *indigox Github repository*. https://github.com/allison-group/indigox (accessed Jan 26, 2020).

(20) Jo, S.; Kim, T.; Im, W. Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLoS One* **2007**, *2* (9), No. e880.

(21) Zanette, C.; Bannan, C. C.; Bayly, C. I.; Fass, J.; Michael, K.; Shirts, M. R.; Chodera, J. D.; Mobley, D. L. Toward Learned Chemical Perception of Force Fi Eld Typing Rules. *J. Chem. Theory Comput.* **2019**, *15*, 402−423.

(22) *WebFF Github repository*. https://github.com/usnistgov/WebFF-Documentation (accessed Jan 26, 2020).

(23) Jain, A.; Ong, S. P.; Chen, W.; Medasani, B.; Qu, X.; Kocher, M.; Brafman, M.; Petretto, G.; Rignanese, G.-M.; Hautier, G.; et al. FireWorks: A Dynamic Workflow System Designed for High-Throughput Applications. *Concurr. Comput. Pract. Exp.* **2015**, *27* (17), 5037−5059.

(24) 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.

(25) MoSDeF Github. https://github.com/mosdef-hub (accessed Jan 26, 2020).

(26) 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, DOI: 10.1007/978-981-10-1128-3_5.

(27) *mBuild documentation*. http://mbuild.mosdef.org/ (accessed Jan 26, 2020).

(28) *mBuild Github*. https://github.com/mosdef-hub/mbuild (accessed Jan 26, 2020).

(29) Klein, C.; Summers, A. Z.; Thompson, M. W.; Gilmer, J. B.; McCabe, C.; Cummings, P. T.; Sallai, J.; Iacovella, C. R. Formalizing Atom-Typing and the Dissemination of Force Fields with Foyer. *Comput. Mater. Sci.* **2019**, *167* (February), 215−227.

(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) *Foyer Github*. https://github.com/mosdef-hub/foyer (accessed Jan 26, 2020).

(32) *Foyer documentation*. http://foyer.mosdef.org (accessed Jan 26, 2020).

(33) 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.

(34) Cummings, P. T. Computational Screening of Soft Materials Systems with Application to Nano-Lubrication Systems. In *Telluride Science Research Center Workshop "Molecular engineering of soft matter: Spanning small molecules to macromolecules," June 19−24, 2017*; 2017.

(35) Bhushan, B.; Sundararajan, S. Micro/Nanoscale Friction and Wear Mechanisms of Thin Films Using Atomic Force and Friction Force Microscopy. *Acta Mater.* **1998**, *46* (11), 3793−3804.

(36) Tambe, N. S.; Bhushan, B. Nanotribological Characterization of Self-Assembled Monolayers Deposited on Silicon and Aluminium Substrates. *Nanotechnology* **2005**, *16* (9), 1549−1558.

(37) Vilt, S. G.; Leng, Z.; Booth, B. D.; McCabe, C.; Jennings, G. K. Surface and Frictional Properties of Two-Component Alkylsilane Monolayers and Hydroxyl-Terminated Monolayers on Silicon. *J. Phys. Chem. C* **2009**, *113* (33), 14972−14977.

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

(39) *RDKit: Open-source cheminformatics*. https://www.rdkit.org/ (accessed Jan 26, 2019).

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

(41) Bhushan, B. Nanotribology and Nanomechanics in Nano/Biotechnology. *Philos. Trans. R. Soc., A* **2008**, *366* (1870), 1499−1537.

(42) Cummings, P. T.; Docherty, H.; Iacovella, C. R. Phase Transitions in Nanoconfined Fluids: The Evidence from Simulation and Theory. *AIChE J.* **2010**, *56* (4), 842−848.

(43) 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.

(44) Lio, A.; Charych, D. H.; Salmeron, M. Comparative Atomic Force Microscopy Study of the Chain Length Dependence of

https://dx.doi.org/10.1021/acs.jctc.9b01183
J. Chem. Theory Comput. 2020, 16, 1779−1793

Frictional Properties of Alkanethiols on Gold and Alkylsilanes on Mica. *J. Phys. Chem. B* **1997**, *101* (19), 3800−3805.
(45) 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.
(46) 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.* **2017**, *396*, 865−869.
(47) Chandross, M.; Grest, G. S.; Stevens, M. J. Friction between Alkylsilane Monolayers: Molecular Simulation of Ordered Monolayers. *Langmuir* **2002**, *18* (22), 8392−8399.
(48) 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.
(49) 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.
(50) Lewis, J. B.; 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.
(51) 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.
(52) 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.
(53) 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.
(54) Summers, A. Z.; Iacovella, C. R.; Cummings, P. T.; MCabe, C. Investigating Alkylsilane Monolayer Tribology at a Single-Asperity Contact with Molecular Dynamics Simulation. *Langmuir* **2017**, *33* (42), 11270−11280.
(55) 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.
(56) 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**, *32* (10), 2348−2359.
(57) 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* (10), 3086−3093.
(58) Schappals, M.; Mecklenfeld, A.; Kröger, L.; Botan, V.; Köster, A.; Stephan, S.; García, E. J.; Rutkai, G.; Raabe, G.; Klein, P.; et al. Round Robin Study: Molecular Simulation of Thermodynamic Properties from Models with Internal Degrees of Freedom. *J. Chem. Theory Comput.* **2017**, *13* (9), 4270−4280.
(59) Sztipanovits, J.; Karsai, G. Model-Integrated Computing. *Computer (Long. Beach. Calif).* **1997**, *30* (4), 110−111.
(60) Iacovella, C. R.; Varga, G.; Sallai, J.; Mukherjee, S.; Ledeczi, A.; Cummings, P. T. A Model-Integrated Computing Approach to Nanomaterials Simulation. *Theor. Chem. Acc.* **2013**, *132* (1), 1315.
(61) 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.; 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.
(62) *SMARTS - A Language for Describing Molecular Patterns*. https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html (accessed Jan 26, 2019).
(63) Zhang, L.; Li, L.; Chen, S.; Jiang, S. Measurements of Friction and Adhesion for Alkyl Monolayers on Si (111) by Scanning Force Microscopy. *Langmuir* **2002**, *18*, 5448−5456.
(64) 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.
(65) Zhuravlev, L. T. The Surface Chemistry of Amorphous Silica. Zhuravlev Model. *Colloids Surf., A* **2000**, *173* (1−3), 1−38.
(66) 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.
(67) *PTC-CMC/atools*. https://zenodo.org/record/3387608#.XkWCc2hKiUk (accessed Jan 26, 2020).
(68) *atools -- A collection of initialization and analysis routines primarily focused on monolayer systems*. https://github.com/PTC-CMC/atools/releases/tag/v0.1 (accessed Jan 26, 2020).
(69) 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**, *32* (10), 2348.
(70) 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.
(71) 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. *Langmuir* **2010**, *26* (12), 9497−9505.
(72) *Github repository of scripts to enable the screening over monolayer terminal groups for uniform systems*. https://github.com/PTC-CMC/terminal_group_screening/releases/tag/v0.1 (accessed Jan 26, 2020).
(73) *PTC-CMC/terminal_group_screening*. https://zenodo.org/record/3401645#.XkWDn2hKiUk (accessed Jan 26, 2020).
(74) *Github repository of scripts to enable the screening over monolayer terminal groups for non-uniform systems*. https://github.com/PTC-CMC/terminal_groups_mixed/releases/tag/v0.1 (accessed Jan 26, 2020).
(75) *PTC-CMC/terminal_groups_mixed*. https://zenodo.org/record/3401648#.XkWER2hKiUk (accessed Jan 26, 2020).
(76) *Github repository of scripts to enable the screening of monolayer terminal groups with additional terminal groups for validation*. https://github.com/PTC-CMC/terminal_group_mixed_original_16_new_3/releases/tag/v0.1 (accessed Jan 26, 2020).
(77) *PTC-CMC/terminal_group_mixed_original_16_new_3*. https://zenodo.org/record/3401655#.XkWEtGhKiUk (accessed Jan 26, 2020).
(78) 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* (45), 11225−11236.
(79) 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.
(80) 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.
(81) 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.
(82) 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.
(83) Watkins, E. K.; Jorgensen, W. L. Perfluoroalkanes: Conformational Analysis and Liquid-State Properties from Ab Initio and Monte Carlo Calculations. *J. Phys. Chem. A* **2001**, *105*, 4118−4125.

https://dx.doi.org/10.1021/acs.jctc.9b01183
J. Chem. Theory Comput. 2020, 16, 1779−1793

(84) 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.
(85) 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**, *1-2*, 19−25.
(86) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1−19.
(87) *LAMMPS*; Web page. http://lammps.sandia.gov (accessed Jan 26, 2020).
(88) Chandross, M.; Webb, E.; Stevens, M.; Grest, G.; Garofalini, S. Systematic Study of the Effect of Disorder on Nanotribology of Self-Assembled Monolayers. *Phys. Rev. Lett.* **2004**, *93* (16), 166103.
(89) 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), 125434.
(90) Nosé, S. A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *J. Chem. Phys.* **1984**, *81* (1), 511.
(91) Hoover, W. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A: At., Mol., Opt. Phys.* **1985**, *31* (3), 1695−1697.
(92) 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. Comput. Chem.* **1997**, *18* (12), 1463−1472.
(93) 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.
(94) 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.
(95) 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. *J. Colloid Interface Sci.* **1999**, *213* (1), 238−250.
(96) 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.
(97) Booth, B. D.; Vilt, S. G.; Lewis, J. B.; Rivera, J. L.; Buehler, E. A.; McCabe, C.; Jennings, G. K. Tribological Durability of Silane Monolayers on Silicon. *Langmuir* **2011**, *27* (10), 5909−5917.
(98) 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.
(99) Frisbie, C. D.; Rozsnyai, L. F.; Noy, A.; Wrighton, M. S.; Lieber, C. M. Functional Group Imaging by Chemical Force Microscopy. *Science (Washington, DC, U. S.)* **1994**, *265* (5181), 2071−2074.
(100) 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.
(101) *PTC-CMC/atools_ml*. https://zenodo.org/record/3401581#.XkWFRmhKiUk (accessed Jan 26, 2020).
(102) *atools_ml -- A collection of data science and machine learning-related tools for analyzing molecular simulation data*. https://github.com/PTC-CMC/atools_ml/releases/tag/v0.1 (accessed Jan 26, 2020).
(103) *random_forest_tg -- Random forest models for predicting the coefficient of friction and adhesion for systems of two contacting functionalized monolayers*. https://github.com/PTC-CMC/random_forest_tg/releases/tag/v0.1 (accessed Jan 26, 2020).
(104) *PTC-CMC/random_forest_tg*. https://zenodo.org/record/3401662#.XkWFsWhKiUk (accessed Jan 26, 2020).
(105) Weininger, D. SMILES, a Chemical Language and Information System: 1: Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Model.* **1988**, *28* (1), 31−36.
(106) 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.
(107) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity—a Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36* (22), 3219−3228.
(108) Summers, A. Z.; Iacovella, C. R.; Cummings, P. T.; McCabe, C. Examining Structure/Property Relationships in Lubricating Monolayer Films through Molecular Dynamics Screening. In *Seventh International Conference on the Foundations of Molecular Modelling and Simulation (FOMMS 2018)*, Lake Lawn Resort, Delavan, WI, July 15−20, 2018. https://figshare.com/articles/Examining_Structure_Property_Relationships_in_Lubricating_Monolayer_Films_through_Molecular_Dynamics_Screening/6654935 (accessed Feb 13, 2020).
(109) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45* (1), 5−32.
(110) Geurts, P.; Ernst, D.; Wehenkel, L. Extremely Randomized Trees. *Mach. Learn.* **2006**, *63* (1), 3−42.
(111) 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.
(112) Hall, L. H.; Kier, L. B. The Molecular Connectivity Chi Indexes and Kappa Shape Indexes in Structure-Property Modeling. In *Reviews in Computational Chemistry*; 1991; Vol. *2*, pp 367−422, DOI: 10.1002/9780470125793.ch9.
(113) Thompson, M. W.; Matsumoto, R.; Sacci, R. L.; Sanders, N. C.; Cummings, P. T. Scalable Screening of Soft Matter: A Case Study of Mixtures of Ionic Liquids and Organic Solvents. *J. Phys. Chem. B* **2019**, *123* (6), 1340−1347.

https://dx.doi.org/10.1021/acs.jctc.9b01183
J. Chem. Theory Comput. 2020, 16, 1779−1793