Membranes, motors, and molecular modeling

The interface between physics and biology is an active, growing area of science with many unanswered questions. My research program will investigate the physical principles that underlie biological and chemical phenomena, providing new insight into the complex behavior of cell membranes, the operation of molecular motors, and more rigorous methods of developing force fields. In particular, I will focus on three areas that bridge theory, modeling, and simulations: (1) the formation of ion-mediated phospholipid domains, (2) the out-of-equilibrium statistical mechanics of synthetic rotary motors, and (3) open source development of new force fields for drug design incorporating binding thermodynamics.

Method development is vital to my research interests because it allows theory and simulation to gain new perspectives on challenging problems. I will extend my previous work on biological membranes to elucidate the driving forces behind clusters featuring phosphoinositides. I will use nonequilibrium kinetic modeling to suggest new design principles for molecular motors. Finally, I will participate in a systematic overhaul of force field science, changing the way force fields are built and how they are distributed, by focusing on transparent and rigorous optimization criteria, in an open setting.

Insert university-specific paragraph here.

1 Use simulations to understand the formation of domains in realistic membrane models

The goal of this aim is to construct quantitative models of physiological membranes and understand the physical mechanisms that drive the lateral segregation of lipid species.

The primary function of a membrane is to separate what is outside from what is inside. In simple organisms, the plasma membrane is a partition between the complex, and sometimes delicate, molecules required for life from the harsh environment. Mammalian plasma membranes are a rich mixture of hundreds of types of phospholipids intermingled with other types of lipids, sterols, and proteins. The fact that only the inner leaflet of eukaryotic cells contains the acidic phospholipids under physiological conditions results in a significant negative surface charge density. Mobile counterions and proteins in the cytoplasm are drawn to the inner leaflet of the plasma membrane, forming both specific and non-specific interactions with membrane constituents, leading to potentially significant alterations in membrane structure including the formation of membrane curvature and surface patterning.

**See and cite https://pubs.acs.org/doi/10.1021/ja507832e **

Despite advances in super-resolution microscopy, such as PALM and STORM, *in vivo* characterization of membrane heterogeneity remains challenging due to the high spatiotemporal resolution required to study fluctuating nanoscale assemblies of lipids and proteins in living cells. However, a variety of indirect experimental techniques, often using fluorescently labeled lipid analogs, have characterized the existence of sub-micron sized domains in living cells and plasma membrane extracts [@GGtK2c0N; @1Eg1Hzju1; @oBaB5Z87; @BzP79Vj9; @aiu6Tmil].
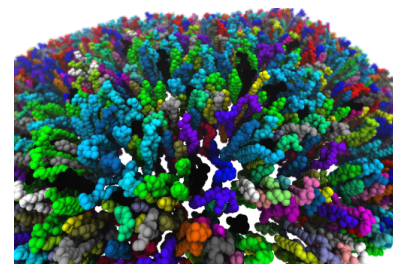

Figure 1: An example of a curved membrane with many different lipid species, colored separately, representative of the systems I will simulate.

My approach is to understand what conditions promote cluster formation and how the behavior of lipids in clusters differs from bulk. I will initially focus on delineating the nucleation of cation-induced clusters of anionic phospholipids. This work will involve extending the methods I developed to investigate single molecule QM/MM and all-atom nanoscale simulations. Extending beyond the all-atom simulations, I will use the [methods](https://github.com/biophyscode) we created to construct and equi-

librate multi-component membrane systems to build models containing rich mixtures of phospholipids (Figure @fig:membrane-setup), going past typical two or three-component systems that are frequently used.

The finite and well-defined size of the domains suggests there is a balance between the mutual electrostatic and steric repulsion of negatively charged lipids and attraction mediated by counterions. I will monitor the extent of phase separation and demixing in simulations containing combinations of physiological divalent and monovalent ions with varying fractions of charged and uncharged lipids. The calculated diffusion coefficient inside the clusters can be compared with values obtained using fluorescence correlation spectroscopy (FCS) and spot variant FCS [@oBaB5Z87]. Simple, physical models based primarily on electrostatics [@10CqL9t0a] predict continual growth of clusters in the presence of excess counterions, yet *in vivo* clusters plateau at a stable size. By using coarse-grained simulations and GPU-accelerated molecular dynamics, I will be able to study the growth and merging of clusters on the microsecond to millisecond timescale. The size and shape of the simulated clusters will be contrasted with those seen with AFM imaging and electron microscopy [@LhOwGz4k].

Experiments pulling thin membrane tethers have shown that phase-separated phospholipid domains form precisely at the junction of highly curved regions [@XIltXoGI], implying a relationship between membrane morphology an lipid sorting. I will compare the bending modulus and mean curvature of clusters with data from pipette aspiration of membrane tubules [@XIltXoGI; @oBaB5Z87; @aiu6Tmil]. The order parameter of lipids in and on the boundary of the cluster can be compared to results complementary non-fluorescent methods, such as EPR, NMR, and neutron scattering.
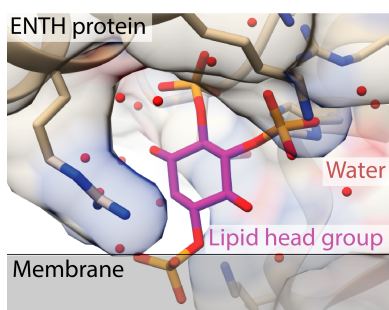

Figure 2: An illustration of stereospecific recognition of phospholipids by proteins.

Finally, I will study whether the act of being in a cluster alters the ability of certain lipids to perform their biological roles. Many of the lipids implicated in cluster formation, such as PtdIns(4,5)P_2, have been shown to bind hundreds of proteins [@uyKE7bWV], and several integral membrane proteins have been shown to partition into lipid raft-like domains [@2TfZ4zWV]. PtdIns(4,5)P_2 binds (some) proteins in a highly stereospecific manner (Figure @fig:stereospecific) requiring a fully extended and accessible head group. In order to evaluate the biological consequences of clustering, I will run simulations containing proteins that bind through either non-specific, electrostatic attraction (e.g., N-WASP and gelsolin) or through tight coordination (e.g., the PH domain of PLCδ. I will test whether these proteins bind "free" lipid molecules as well as they bind clustered ones by using steered molecular dynamics.

I expect the outcome of studying domain formation to be especially helpful for building better models of plasma membranes that reflect the complex cellular environment. I anticipate these results could provide a quantitative insights for pore-formation in cell membranes by toxins and cationic surfactants, the activation of integral ion channels and GPCRs by PtdIns(4,5)P_2, and the dynamics of membrane fusion, fission and other buckling geometries

2   Develop a physical theory of molecular motors

The proposed research the mechanical properties of existing synthetic molecular motors and suggest new design routes for future nanomachines.

The ability of light-driven molecular motors to switch between two or more states makes them suitable for new

forms of optical data storage[1], as "wheels" on nano-scale cars[2], trains[3], worms[4], and walkers[5], or in new forms of responsive materials[6]. Unlike switches, whose work is reversed after every full cycle, molecular motors can be used to progressively move systems away from thermodynamic equilibrium[7]. The synthetic molecular motors of Ben Fergina operate by converting light and heat into directional rotary motion. These motors belong to a class of molecules called overcrowded alkenes, with two stable enantiomers, and adopt a helical shape due to steric strain around the central double bond. The two sets of conjugated rings rotate relative to each other, with the central bond acting as an axle, and for simplicity, one set of rings is designated the "stator" while the other set is called the "rotor." The directional motion of these molecules can be analyzed in terms of two degrees of freedom: E to Z for the isomerization of the double bond (analogous to *cis* and *trans*) and P and M for the overall twist or helicity of the molecule (Figure **??**).

2.1 Calculate the speed, torque, and efficiency of motors

In 2006, it was shown that light driven molecular motors can rotate a glass rod that is more than 10,000 times their size upon irradiation with light and when included as a dopant in a liquid crystal film[8]. While it is clear that artificial motors, when aligned appropriately so their individual effects are magnified, can produce macroscopic effects, it is not known how much force an individual molecular motor can generate. I will take the nonequilibrium model I developed to quantify directional motion in biological motors to these artificial molecular motors, in a new direction, focusing on the "second generation" class of motors from Ben Feringa and colleagues, which possess chemically symmetric stators and a lower energy for the thermal step. I will study the four ground states that comprise the 360° cycle depicted in Figure 4. On the left, the four structural states of the motor demonstrate movement of the upper rotor relative to the fixed stator through two photochemical ($1 \rightarrow 2, 3 \rightarrow 4$) and two thermal steps ($2 \rightarrow 3, 4 \rightarrow 1$). On the right, I've sketched illustrative free energy profiles along a given surface E or Z; these represent the pattern of energy troughs and barriers for changing the helicity of the structure while maintaining the same orientation of the double bond, essentially sliding the bulky biaryl rings connected to the rotor past the stator.
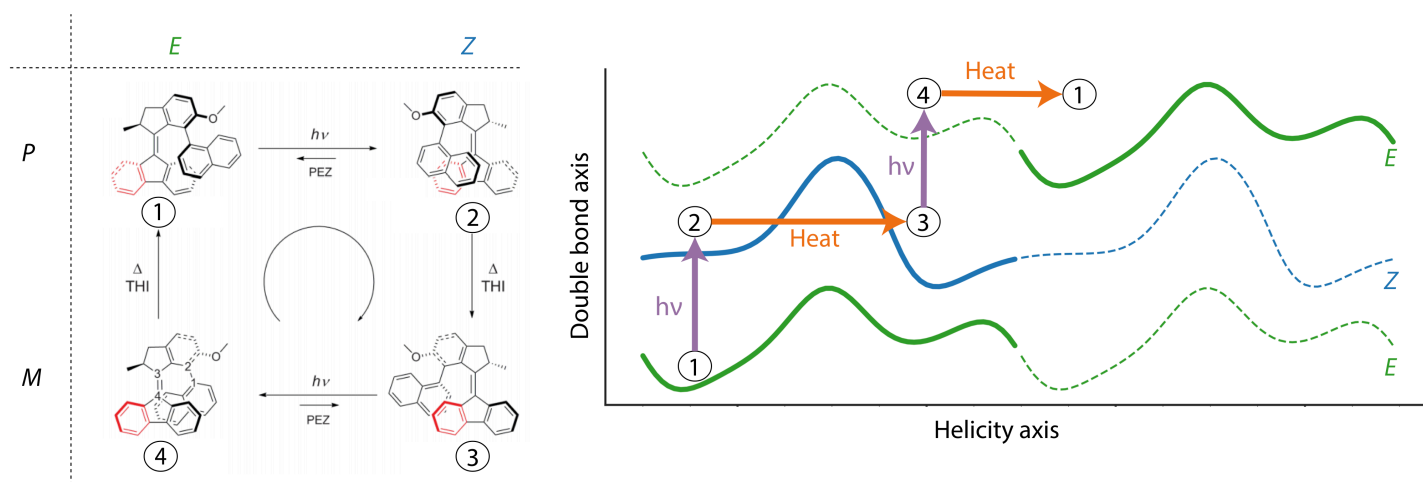

Figure 4: On the left, the four ground state conformations of a second generation motor, adapted from Štacko et al[9]. On the right, the same four states placed on free energy profiles. The energy profiles are periodic, with two cycles shown for both isomerization state. A clone of the lower E energy surface is shown above, for clarity, to demonstrate the progression from state 4 to state 1 requires energy.

Previously we used microsecond-scale molecular dynamics simulations to deter-
mine the energy landscapes based on equilibrium population distributions. How-
ever, because the barriers here are expected to be much larger, I will directly
probe the potential surfaces by evaluating the energy of motor conformations for
different values of the central dihedral angle. I will perform this dihedral scan
using a quantum chemistry package, such as Psi4 with density functional theory
and semi-empirical methods. I will begin with the semi-empirical methods PM3 or
AM1 and test the sensitivity of the results to different quantum approaches (e.g.,
other types of DFT and Hartree-Fock) and basis sets (e.g., 6-31G*). From the
discrete free energy profiles, the rate of bin-to-bin transitions can be deduced, and these are gathered, together
with the surface-to-surface transition rates, in a Markov matrix. Solving for the eigenvectors of the matrix yields
the nonequilibrium steady state (NESS) population in each photochemical state of the motor. From the NESS,
I will calculate the net probability flux for the complete cycle ($1 \rightarrow 1$), as illustrated by the series of purple and
orange arrows in Figure 4, for a given input level of light irradiation.

2.2   Optimize the design of artificial molecular motors
As demonstrated by Richard Feynman in a lecture on Brownian ratchets, the chal-
lenge of designing molecular motors is not how to create movement, but how to
control the directionality of the omnipresent – and random – motion on the molec-
ular scale[10]. This has been referred to as the "gating" of stochastic motion[11]. In
my previous work[12], I showed that the gating can arise purely from the comple-
mentary shape of two potential energy surfaces, whereby an unpassable barrier
on one surface can be circumnavigated by motion on the other surface (see Fig-
ure S6, in particular.) In this section, I will explore the relationship between the
shape of the potential energy surfaces and the directionality of artificial molecular
motors. This sub-aim will test the current understanding of motor mechanisms, most of which has been inferred
through static structural snapshots and chemical intuition, using design proofs and will open the field of rational
design to molecular rotors, pumps, transporters, and other nanorobotics.

It has been empirically determined that adding electron donating and electron
withdrawing moieties has the potential to affect the rotation rate of the motor[13].
Polarizing the molecule by placing electron withdrawing and electron donating
groups on either end of the central double bond, has the potential to lengthen
the effective axle, significantly reducing the barrier for rotation. I will numerically
optimize the potential energy surfaces that I determine in the previous sub-aim,
to yield the maximum directional flux and maximum torque (see Figure 5 for an
example). The surfaces will be modeled as splines with six to eight control points
evenly spaced along the periodic degree of freedom and the spline points will be
allowed to move as the optimization procedure runs. Based on preliminary work, this procedure affords ample
room to significantly change the properties of the surfaces. I will couple the understanding that arises from this
study with the knowledge of how chemical constituents affect overall rotation rates to suggest new avenues for
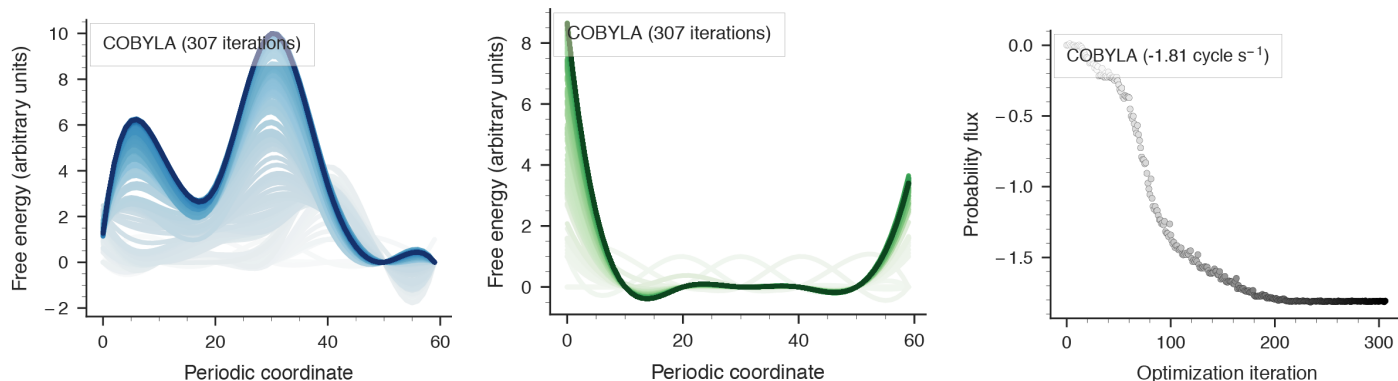synthetic motor design.

Figure 5: Example optimization (using the constrained optimization by linear approximation algorithm, COBYLA) of motor surfaces for maximum flux. Later iterations are darker colors. On the left, the Z surface; in the middle, the E surface; on the right, the net probability flux corresponding to each optimization iteration.

3   Building better force fields for drug discovery using open data and automated optimization

Designing ligands that bind their target with high affinity and specificity is the principal objective of small-molecule drug discovery, yet hit-to-lead and lead-optimization can take upwards of three years owing to the fact it is often necessary to synthesize hundreds or thousands of new compounds. In short, drug discovery is very expensive and often fails. In recent years, the pharmaceutical industry has begun to use absolute and relative binding free energy calculations to help narrow the number of compounds that must be synthesized[14,15]. In particular, the advent of computational calorimetry has enabled quantitative and high-precision comparisons of binding free energies, enthalpies, and entropies (by subtraction) with experimental values determined by isothermal titration calorimetry or NMR[16]. The root mean squared error (RMSE) associated with such calculations is in the range of 1-4 kcal/mol[17], yet even a modest improvement of the prediction accuracy of ~1 kcal/mol would lead to a substantial decrease in the number of compounds that must be manually screened by nearly an order of magnitude[18]. Although it has been suggested by some that changes in the force field functional form are required, it is clear that there is ample room to improve the accuracy of existing parameters through careful analysis[19]. The first goal of this aim is to create better tools to guide early-stage drug discovery by reducing the number of compounds that must be synthesized to find a promising hit that can be carried over into clinical trials.

Host-guest systems are noncovalent complexes between a cavity-like host molecule and a small molecule guest. These systems retain many of the same, strong functional group interactions of protein-ligand systems while being computationally tractable and reaching convergence on a shorter time scale. In the attach-pull-release (APR) method, the reversible work of transferring the guest from the binding site to solution, via a physical pathway, is computed using a series of umbrella sampling windows (Figure **??**). In the "attach" phase, restraints are connected to the guest (and optionally, to the host for better conformational sampling) through a parameter $\lambda \in [0, 1]$ that controls the strength of

the restraints. During the "pull" phase, the equilibrium length of a distance restraint joining the guest and host is increased until the guest is no longer interacting with the host molecule. The "release" phase reverses the work of attaching the restraints and also corrects the concentration of the guest molecule to standard state. Simulating each window and integrating over the partial derivative of the restraint energy with respect to the restraint target, we can generate a potential of mean force along the pulling coordinate that is used to compute the binding free energy at standard state, ΔG°.

3.1 Incorporate host-guest binding data into force field development

APR has consistently been ranked among the most accurate methods for predicting binding thermodynamics in blind challenges[20,21]. I will independently continue the development of APR, through its Python interface, pAPRika, and transform it into a generalized tool that is capable of computing free energies along any physical pathway, using either AMBER or OpenMM as simulation packages. The core functionality of pAPRika is already being used by research groups across the country and in at least one group in China. Extending pAPRika would allow the same rigorous thermodynamic framework to be used in applications such as protein-protein interactions, dimerization of small molecules, or nucleotide flipping. As part of the development, I will also investigate the accuracy of non-pathway, end-point methods, such as the direct calculation of interaction entropy[22].

Most established small molecule force fields (e.g., the general AMBER force field, GAFF) have been tuned using pure liquid state data, such as the average density, enthalpy of vaporization, or the self-diffusion coefficient. Over the past decade, it has become clear that reproducing those properties well does not guarantee binding thermodynamics at a level acceptable for guiding experiments. The open force field group (OpenFF) aims to develop new, collaborative force fields using open access methods, open source software, and high-quality open data sets. One central difference with the OpenFF group is the use of direct chemical perception to apply force field parameters based on a molecular graph instead of atom types[23], and the desire to avoid over-fitting parameters by making Monte Carlo moves in parameter space to find the fewest number of parameters required to describe a set of chemistries[24]. The first prototype force field from this effort was release in June 2018, called SMIRNOFF99Frosst v0.1, based on a predecessor of GAFF. SMIRNOFF99Frosst is able to perform as well as GAFF, and in some cases, alleviate conformational problems caused by GAFF's atom types, in only 335 parameter lines compared to the 6794 lines in GAFF version 2.1. Beyond this initial release, multiple generations of SMIRNOFF99Frosst are planned, incorporating new forms of Lennard-Jones interactions, the addition of atom polarizabilities, and Bayesian estimates to quantify systematic errors in the force field.

As part of the OpenFF consortium, I will benchmark new iterations of SMIRNOFF99Frosst using host-guest thermodynamics computed by pAPRika. Figure **??** shows the very first comparison of SMIRNOFF99Frosst v0.1 and GAFF v1.7 on host-guest systems. GAFF v1.7 is known to overestimate binding affinities by >1 kcal/mol[25], and it appears that SMIRNOFF99Frosst reduces this tendency. To begin, I will use the "benchmark set" of Mobley et al.[26]. There, experimental data are available for rigid curcurbituril and highly flexible cyclodextrin hosts with a variety of drug-like guest molecules. When available, I will also incorporate the binding data resulting from selective derivatization of cyclodextrins designed to evaluate specific functional group interactions[27].

Enthalpies are poorly correlated with free energies, and thus, they represent a nearly independent set of data to compare. I will match the simulations to the experimental conditions, including host and guest concentration, stoichiometry, ion concentration, and buffer conditions. I will report metrics such as RMSE, R^2, Kendall's τ, and the slope of the correlation between candidate parameter sets and (a) a reference force field such as GAFF, CGenFF, or OPLS3 and (b) the experimental thermodynamic data determined using ITC and NMR. The results of my analysis will be reported through continuous integration, whereby each candidate update to the force field is automatically tested through a clear and well-defined set of tests.

3.2 Integrate automated optimization techniques, such as ForceBalance, with host-guest binding data

As part of the open force field effort, we need to be able to rapidly and efficiently optimize force field parameters given functional forms and input data. ForceBalance is a tool, written by Dr. Lee-Ping Wang, that uses thermodynamic fluctuation formulas and reference data (e.g., *ab initio* quantum mechanical calculations and experimentally known molecular or bulk properties) to optimize an objective function, such as the sum of squared differences between the calculated and reference values ForceBalance has already been used to optimize water models[28] and protein force fields[29]. As a member of the OpenFF consortium, Dr. Wang has begun to add support for the SMIRNOFF force field format. I will extend my collaboration with Dr. Wang to include pAPRika in the optimization loop of ForceBalance. I have experience working with the analytic derivatives of the binding free energy with respect to Lennard-Jones parameters[30], and using these derivatives to tune parameters was recently used by a colleague to optimize the TIP3P water model for binding calculations by post-processing existing MD trajectories for small parameter perturbation[31].

A proximate goal is to create a method for passing a candidate set of parameters from ForceBalance to pAPRika and for pAPRika

to return an estimate of the quality of the parameters in machine-readable format, such as JSON or XML. The experimental data for ForceBalance will be taken from the free and open ThermoML database curated by NIST. I will focus on including data that represent intermolecular interactions, such as enthalpies of mixing and partition coefficients, that will broaden the data used for force field development beyond single species properties. Host-guest calculations are computationally expensive, sometimes requiring >1 GPU-day to reach the desired level of convergence, and this has the potential to significantly slow down ForceBalance. I will add the ability to automatically allocate resources in the APR calculation, by calculating the return on investment (ROI) in each window. The ROI is the partial derivative of the overall standard error of the mean binding affinity with respect to the number of frames in window i, $\partial G^\circ_{\text{SEM}}/\partial N^i_{\text{frames}}$. This will allow us to simulate each window long enough such that the overall estimate of the binding free energy reaches below a certain threshold.

Together, these methodological improvements will help create a transparent and robust set of metrics to evaluate the performance of candidate force fields on an equal footing. By incorporating host-guest binding data, the degeneracy in parameter space will be broken, avoiding force fields that agree excellently with experiment for liquid properties and yet agree poorly on binding. I believe the ideas in this aim could be turned into an NIH proposal by demonstrating the need to systematically evaluate force field accuracy for protein-ligand binding, perhaps the most important application of this work for human health.

References

(1) Feringa, B.; Koumura, N.; van Delden, R.; ter Wiel, M. Light-Driven Molecular Switches and Motors. *Appl Phys A* **2002**, *75* (2), 301–308.

(2) Kudernac, T.; Ruangsupapichat, N.; Parschau, M.; Maciá, B.; Katsonis, N.; Harutyunyan, S. R.; Ernst, K.-H.; Feringa, B. L. Electrically Driven Directional Motion of a Four-Wheeled Molecule on a Metal Surface. *Nature* **2011**, *479* (7372), 208–211.

(3) Sasaki, T.; Guerrero, J. M.; Leonard, A. D.; Tour, J. M. Nanotrains and Self-Assembled Two-Dimensional Arrays Built from Carboranes Linked by Hydrogen Bonding of Dipyridones. *Nano Res.* **2008**, *1* (5), 412–419.

(4) Sasaki, T.; Tour, J. M. Synthesis of a New Photoactive Nanovehicle: A Nanoworm. *Org. Lett.* **2008**, *10* (5), 897–900.

(5) von Delius, M.; Leigh, D. A. Walking Molecules. *Chem. Soc. Rev.* **2011**, *40* (7), 3656.

(6) Lucas, L. N.; van Esch, J.; Feringa, B. L.; Kellogg, R. M. Photocontrolled Self-Assembly of Molecular Switches. *Chem. Commun.* **2001**, No. 8, 759–760.

(7) Kassem, S.; van Leeuwen, T.; Lubbe, A. S.; Wilson, M. R.; Feringa, B. L.; Leigh, D. A. Artificial Molecular Motors. *Chem. Soc. Rev.* **2017**, *46* (9), 2592–2621.

(8) Eelkema, R.; Pollard, M. M.; Vicario, J.; Katsonis, N.; Ramon, B. S.; Bastiaansen, C. W. M.; Broer, D. J.; Feringa, B. L. Nanomotor Rotates Microscale Objects. *Nature* **2006**, *440* (7081), 163–163.

(9) Štacko, P.; Kistemaker, J. C. M.; van Leeuwen, T.; Chang, M.-C.; Otten, E.; Feringa, B. L. Locked Synchronous Rotor Motion in a Molecular Motor. *Science* **2017**, *356* (6341), 964–968.

(10) Browne, W. R.; Feringa, B. L. Making Molecular Machines Work. *Nature Nanotech* **2006**, *1* (1), 25–35.

(11) Astumian, R. D. Stochastic Pumping of Non-Equilibrium Steady-States: How Molecules Adapt to a Fluctuating Environment. *Chem. Commun.* **2018**, *54* (5), 427–444.

(12) Slochower, D. R.; Gilson, M. K. Motor-Like Properties of Nonmotor Enzymes. *Biophysical Journal* **2018**, *114* (9), 2174–2179.

(13) Oruganti, B.; Durbeej, B. On the Possibility to Accelerate the Thermal Isomerizations of Overcrowded Alkene-Based Rotary Molecular Motors with Electron-Donating or Electron-Withdrawing Substituents. *J Mol Model* **2016**, *22* (9).

(14) van Vlijmen, H.; Desjarlais, R. L.; Mirzadegan, T. Computational Chemistry at Janssen. *J Comput Aided Mol Des* **2016**, *31* (3), 267–273.

(15) Christ, C. D.; Fox, T. Accuracy Assessment and Automation of Free Energy Calculations for Drug Design. *J. Chem. Inf. Model.* **2013**, *54* (1), 108–120.

(16) Henriksen, N. M.; Fenley, A. T.; Gilson, M. K. Computational Calorimetry: High-Precision Calculation of Host–Guest Binding

Thermodynamics. *J. Chem. Theory Comput.* **2015**, *11* (9), 4377–4394.

(17) Gaieb, Z.; Liu, S.; Gathiaka, S.; Chiu, M.; Yang, H.; Shao, C.; Feher, V. A.; Walters, W. P.; Kuhn, B.; Rudolph, M. G.; et al. D3R Grand Challenge 2: Blind Prediction of Protein–Ligand Poses, Affinity Rankings, and Relative Binding Free Energies. *J Comput Aided Mol Des* **2017**, *32* (1), 1–20.

(18) Shirts, M. R.; Mobley, D. L.; Brown, S. P. Free-Energy Calculations in Structure-Based Drug Design. In *Drug Design: Structure- and Ligand-Based Approaches*; Merz Jr., K. M., Ringe, D., Reynolds, C. H., Series Eds.; Cambridge University Press, 2010.

(19) Fennell, C. J.; Wymer, K. L.; Mobley, D. L. A Fixed-Charge Model for Alcohol Polarization in the Condensed Phase, and Its Role in Small Molecule Hydration. *J. Phys. Chem. B* **2014**, *118* (24), 6438–6446.

(20) Yin, J.; Henriksen, N. M.; Slochower, D. R.; Shirts, M. R.; Chiu, M. W.; Mobley, D. L.; Gilson, M. K. Overview of the SAMPL5 Host–Guest Challenge: Are We Doing Better? *J Comput Aided Mol Des* **2016**, *31* (1), 1–19.

(21) Muddana, H. S.; Fenley, A. T.; Mobley, D. L.; Gilson, M. K. The SAMPL4 Host–Guest Blind Prediction Challenge: An Overview. *J Comput Aided Mol Des* **2014**, *28* (4), 305–317.

(22) Duan, L.; Liu, X.; Zhang, J. Z. Interaction Entropy: A New Paradigm for Highly Efficient and Reliable Computation of Protein–Ligand Binding Free Energy. *J. Am. Chem. Soc.* **2016**, *138* (17), 5722–5728.

(23) Mobley, D.; Bannan, C. C.; Rizzi, A.; Bayly, C. I.; Chodera, J. D.; Lim, V. T.; Lim, N. M.; Beauchamp, K. A.; Shirts, M. R.; Gilson, M. K.; et al. Open Force Field Consortium: Escaping Atom Types Using Direct Chemical Perception with SMIRNOFF V0.1, 2018.

(24) Zanette, C.; Bannan, C. C.; Bayly, C. I.; Fass, J.; Gilson, M.; Shirts, M. R.; Chodera, J.; Mobley, D. Toward Learned Chemical Perception of Force Field Typing Rules, 2018.

(25) Henriksen, N. M.; Gilson, M. K. Evaluating Force Field Performance in Thermodynamic Calculations of Cyclodextrin Host–Guest Binding: Water Models, Partial Charges, and Host Force Field Parameters. *J. Chem. Theory Comput.* **2017**, *13* (9), 4253–4269.

(26) Mobley, D. L.; Gilson, M. K. Predicting Binding Free Energies: Frontiers and Benchmarks. *Annu. Rev. Biophys.* **2017**, *46* (1), 531–558.

(27) Kellett, K.; Kantonen, S. A.; Duggan, B. M.; Gilson, M. K. Toward Expanded Diversity of Host-Guest Interactions via Synthesis and Characterization of Cyclodextrin Derivatives, 2018.

(28) Wang, L.-P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S. Systematic Improvement of a Classical Molecular Model of Water.

J. Phys. Chem. B **2013**, *117* (34), 9956–9972.

(29) Wang, L.-P.; McKiernan, K. A.; Gomes, J.; Beauchamp, K. A.; Head-Gordon, T.; Rice, J. E.; Swope, W. C.; Martínez, T. J.; Pande, V. S. Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. *J. Phys. Chem. B* **2017**, *121* (16), 4023–4039.

(30) Yin, J.; Fenley, A. T.; Henriksen, N. M.; Gilson, M. K. Toward Improved Force-Field Accuracy Through Sensitivity Analysis of Host-Guest Binding Thermodynamics. *J. Phys. Chem. B* **2015**, *119* (32), 10145–10155.

(31) Yin, J.; Henriksen, N. M.; Muddana, H. S.; Gilson, M. K. Bind3P: Optimization of a Water Model Based on Host–Guest Binding Data. *J. Chem. Theory Comput.* **2018**, *14* (7), 3621–3632.