

## Membranes, motors, and molecular modeling

My research interests center on uncovering the physical principles that underlie biological and chemical phenomena. I try to create more accurate molecular simulations by improving physics-based force fields and combining new simulation procedures with nonequilibrium analytical tools. I will extend my previous work on biological membranes to elucidate driving forces behind the formation of membrane domains featuring phosphoinositides. I will also run simulations of synthetic molecular motors, such as those designed by Ben Feringa, to characterize their mechanical properties. 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. These investigations will provide a detailed accounting of the physical principles that are at work in biological systems. My undergraduate degree in physics combined with my graduate work in biochemistry and biophysics, together with my postdoctoral training in computational chemistry, makes me well suited to this research plan. Each aim of my detailed research plan focuses on a physical or chemical question at the heart of molecular motion.

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.](#)

### Summary of previous research

#### 1 Improving free energy calculations

Accurate binding thermodynamics calculations remain one of the central goals of biomolecular simulations. While in the Gilson group, I contributed to an overall evaluation of more than ten different methods, from research groups around the world, used to compute host-guest binding free energies and enthalpies<sup>1</sup>. As part of that work, I extended our own method (attach-pull-release<sup>2</sup>) by completely rewriting the code in Python and adding several advanced features enabling larger and more challenging systems to be used (this code is available on [GitHub](#)). More recently, I have also worked on assessing the efficiency and reproducibility of how different algorithms converge toward free energy estimates (forthcoming publication).

#### 2 The evolution of molecular motors

Biological motors are enzymes that use chemical energy to drive directional motion. Although motor proteins must have arisen through random mutation and natural selection, it is not clear how the evolutionary leap from non-motor enzymes to molecular motors could have occurred. I showed that any chiral molecule driven out of

equilibrium should undergo cycles of conformational change<sup>3</sup>. This work was highlighted by [UCSD](#) and “[New and Notable](#)” in Biophysical Journal.

### 3 The design of new force fields

As part of the open force field consortium—a collaboration with Dr. David Mobley, Dr. John Chodera, Dr. Michael Shirts, Dr. Lee-Ping Wang, and Dr. Chris Bayly— I participated in the effort to build new molecular mechanics force fields based on direct chemical perception instead of indirect, atom-type based approaches (forthcoming publication). I have also contributed to the design, coding, and implementation of a new force field format—SMIRKS Native Open Force Field (SMIRNOFF)—based on a hierarchical structure that supports customized combining rules, partial bond orders, polarizability, and other advanced features. The tools and data are open source and available on [GitHub](#).

### 4 Simulations of complex membranes

Continuing the work that I started during graduate school at the University of Pennsylvania<sup>4–8</sup>, I’m interested in simulating increasingly complex, and thus more physiologically relevant, models of biological membranes. We recently reported (forthcoming publication) how the interplay between lipid species, cholesterol, counterions, and water near the membrane interface can lead to the formation of nano-scale lipid clusters. A natural next step is to understand the role proteins play in stabilizing these clusters, as discussed below.

#### Detailed research plan

##### 1 How are highly enriched clusters of negatively charged phospholipids stabilized?

Phosphoinositides are a minor class of lipids located on the inner leaflet of the plasma membrane, found at about 0.5 mole percent of all phospholipids in mammalian cells (10–100  $\mu\text{M}$ ), that have an outsized biological role. This class of lipids participates in myriad cell processes including attachment of the cytoskeleton, membrane fusion, vesiculation, and both activation and inhibition of enzymes<sup>9</sup>. Disruption in signaling pathways associated with the phosphoinositides have been implicated in several cancers as well as Charcot-Marie-Tooth disease along with other neurodegenerative disorders<sup>10–15</sup>. Among all phospholipids,  $\text{PtdIns}(4,5)\text{P}_2$  may be the most common regulatory lipid: hundreds of cytosolic proteins have been found to bind  $\text{PtdIns}(4,5)\text{P}_2$  *in vitro*<sup>16</sup> including proteins that cause or sense membrane curvature<sup>17,18</sup>.

$\text{PtdIns}(4,5)\text{P}_2$  is one of the most highly charged molecules in the cell, carrying a net charge from -3 to -5 at physiological pH, owing to a phosphodinoester and two phosphomonoester groups<sup>19</sup>. Despite their high charge, through the use of super-resolution microscopy, there is mounting evidence for nano-scale clusters of  $\text{PtdIns}(4,5)\text{P}_2$  within cells. In neuronal cells, clusters of  $\text{PtdIns}(4,5)\text{P}_2$  have been visualized using fluorescent phosphoinositide-binding domains and phosphoinositide-specific antibodies<sup>20,21</sup>. These clusters are circular, roughly 70–90 nm in diameter, and composed of greater than 80%  $\text{PtdIns}(4,5)\text{P}_2$ . A separate study found clusters of around 60 nm for  $\text{PtdIns}(4,5)\text{P}_2$ , and larger clusters for  $\text{PtdIns}(3,4,5)\text{P}_3$ , in the plasma membrane<sup>22</sup>. Even in the absence of proteins, *in vitro* assays in the Janmey group have demonstrated  $\text{PtdIns}(4,5)\text{P}_2$  clusters of 40–50 nm, by AFM and TEM imaging, after the addition of as little as  $\sim 1 \mu\text{M}$   $\text{Ca}^{2+}$ <sup>23</sup>. Another study using fluorescence self-quenching and Förster resonance energy transfer assays noted  $\text{PtdIns}(4,5)\text{P}_2$  self association at concentrations as low as 0.05 mole percent, with divalent and trivalent metal ions, in model membranes and liposomes<sup>24</sup>.

- Our simulations to date have not only yielded an accurate characterization of  $\text{PtdIns}(4,5)\text{P}_2$  (at the native protonation state and bound to different ions) in a physiologically relevant bilayer environment, but also several important structural and energetic metrics – head group angle, molecular area, solvation in the

immediate vicinity of  $\text{PtdIns}(4,5)P_2$ , and spatial ion distribution functions for the entire bilayer.

- However, I have yet to study clusters of  $\text{PtdIns}(4,5)P_2$  and the interactions between  $\text{PtdIns}(4,5)P_2$  and proteins, two biologically-relevant components.

### 1.1 Understand how divalent cations grow and stabilize clusters

The finite and well-defined size of  $\text{PtdIns}(4,5)P_2$  clusters suggests there is a balance between the mutual electrostatic and steric repulsion of the negatively charged lipids and the attraction mediated by  $\text{Ca}^{2+}$ . A key piece of evidence is the observation from molecular dynamics simulations that a single  $\text{Ca}^{2+}$  ion is able to bring together groupings of up to three lipids and lead to changes in membrane packing (as detailed in a publication currently under review). In simulations, we find that only  $\text{Ca}^{2+}$  has the ability to nucleate clusters; in experiments, clusters are detected only at higher concentrations of  $\text{Mg}^{2+}$  and not at all with the monovalent ions  $\text{Na}^+$  or  $\text{K}^+$ <sup>25</sup>.

Building upon an automated toolchain for constructing heterogeneous membrane compositions for molecular simulations<sup>26,27</sup>, I will construct and run simulations of  $\text{PtdIns}(4,5)P_2$  clusters. As in my previous work, I will test several variables independently by varying the counterion species, counterion concentration, and presence or absence of proteins.

- What is the critical concentration for  $\text{PtdIns}(4,5)P_2$  to spontaneously form clusters?
- $\text{PtdIns}(4,5)P_2$  preferentially partitions into the liquid disordered phase.
- As the concentration of  $\text{Ca}^{2+}$  in the aqueous phase increases, the gel phase domains increase in size and irregularity; the gel phase domains have excess charge density and repel each other with such energy that they are capable of forming a Wigner lattice (the lowest energy spatial configuration of negatively charged particles in a positive electrostatic potential such that the Coulombic interactions dominate the kinetic energy) [47].
- Do  $\text{PtdIns}(4,5)P_2$  clusters grow in size over the course of the simulation?
- What happens on the boundary of the cluster? Are certain lipids more likely than others? Where is cholesterol?
- That the clusters so far visualized in experiments are round suggests the domains are fluid, rather than crystalline.
- I am going to test the hypothesis that  $\text{PtdIns}(4,5)P_2$  clusters increase the height of the membrane, are more laterally stiff, and display curvature.
- What is the diffusion coefficient inside the cluster?
- Quantify the Gaussian curvature.

### 1.2 Determine how the ability of $\text{PtdIns}(4,5)P_2$ to bind proteins changes inside a cluster

I am going to test the hypothesis that  $\text{PtdIns}(4,5)P_2$  clusters can form even in the absence of proteins, and that the clustered  $\text{PtdIns}(4,5)P_2$  has a greater affinity for many proteins.

- It is thought that the different physical properties inside clusters have significant biological effects.
- As many  $\text{PtdIns}(4,5)P_2$ -binding proteins interact through an unstructured polybasic stretch, their interactions with the membrane depend on its surface potential and are therefore sensitive to the local density of  $\text{PtdIns}(4,5)P_2$ , which is the most highly charged lipid in the bilayer.
- To stereospecifically bind  $\text{PtdIns}(4,5)P_2$ , PH domains require a nearly extended form of  $\text{PtdIns}(4,5)P_2$  (this

can be seen in some crystal structures).

- Actin regulatory proteins bind  $\text{PtdIns}(4,5)P_2$  and are likely to respond to changes in  $\text{PtdIns}(4,5)P_2$  packing.
- $\text{PtdIns}(4,5)P_2$  clustering is not reduced when actin nucleation is disrupted, suggesting the clusters precede the proteins.
- We will build smaller ( $40 \times 40 \text{ nm}^2$ ) bilayer patches with varying levels of  $\text{PtdIns}(4,5)P_2$  concentration (including 0, 4, 10, and 30%) in the “inner” leaflet to study the effects of  $\text{PtdIns}(4,5)P_2$  on the protein binding mechanism.
- By studying the binding of polybasic protein domains under varying concentrations, we will characterize the interactions governing the protein binding affinity, and report the molecular mechanism of the PIP<sub>2</sub>-density dependence of polybasic protein binding.
- Many of the results of this aim can be compared directly to *in vitro* experiments of  $\text{PtdIns}(4,5)P_2$  in model membranes and *in vivo* imaging studies along the lines of van den Bogaart.
- Diffusion is very slow.
- Induction of pores or local regions of high curvature.
- What is the Gaussian curvature?
- What is the modulus of elasticity?
- What is the equivalent radius of curvature?
- Match AFM or scattering experiments.
- Saffron-Delbruk model of diffusion.
- The results can be compared to AFM, IR, neutron scattering, or other biophysical experiments on membranes.
- As I reported previously, the angle between the head group and the surface of the membrane affects the lipid’s ability to bind proteins.

## 2 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>

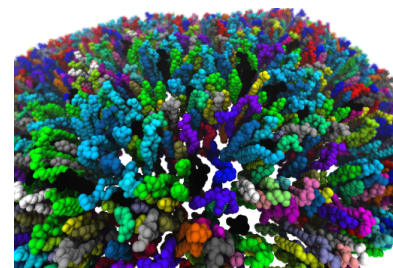
Despite advances in super-resolution microscopy, such as PALM and STORM, *in vivo* characterization of membrane heterogeneity remains challenging due to the high spatiotemporal resolution required to study fluctuating nanoscale assemblies of lipids and proteins in living cells. However, a variety of 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<sup>28–32</sup>. Even in the absence of proteins, *in vitro* assays in the Janmey group have demonstrated clusters of negatively charged phospholipids of on the order of 50 nm after the addition of as little as 1  $\mu\text{M}$   $\text{Ca}^{2+}$ <sup>23</sup>.

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. In those simulations, I found that  $\text{Ca}^{2+}$  can act as “molecular glue,” forming aggregates of three lipids that remain together for at least 100 ns, much longer than other ion-lipid bond lifetimes. Extending beyond the all-atom simulations, I will use the [methods](#) we created to construct and equilibrate multi-component membrane systems to build models containing rich mixtures of phospholipids (Figure ??), 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<sup>30</sup>. Simple, physical models based primarily on electrostatics<sup>33</sup> 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<sup>23</sup>.

Experiments pulling thin membrane tethers have shown that phase-separated phospholipid domains form precisely at the junction of highly curved regions<sup>34</sup>, implying a relationship between membrane morphology and lipid sorting. I will compare the bending modulus and mean curvature of clusters with data from pipette aspiration of membrane tubules<sup>30,32,34</sup>. 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.

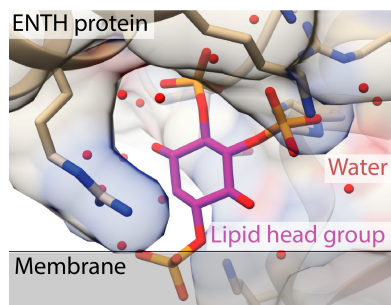
Finally, I will study whether the act of being in a cluster alters the ability of certain lipids to perform their biological roles. Many of the lipids implicated in cluster formation, such as  $\text{PtdIns}(4,5)\text{P}_2$ , have been shown to bind hundreds of proteins<sup>16</sup>, and several integral membrane proteins have been shown to partition into lipid raft-like domains<sup>35</sup>.  $\text{PtdIns}(4,5)\text{P}_2$  binds (some) proteins in a highly stereospecific manner (Figure ??) requiring a fully extended and accessible head group. In order to evaluate the biological consequences of clustering, I will run



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



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 PLC $\delta$ ). I will test whether these proteins bind “free” lipid molecules as well as they bind clustered ones by using steered molecular dynamics.



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

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.

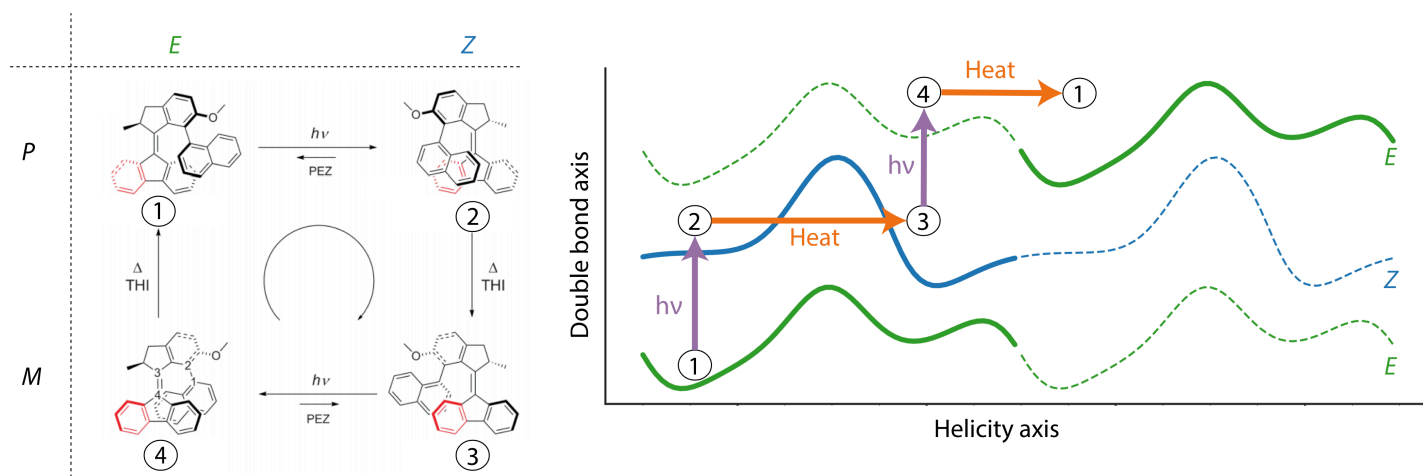
### 3 What are the mechanical properties of synthetic molecular motors?

The ability of light-driven molecular motors to switch between two or more states makes them suitable for new forms of optical data storage<sup>36</sup>, as “wheels” on nano-scale cars<sup>37</sup>, trains<sup>38</sup>, worms<sup>39</sup>, and walkers<sup>40</sup>, or in new forms of responsive materials<sup>41</sup>. Unlike switches, whose work is reversed after every full cycle, molecular motors can be used to progressively move systems away from thermodynamic equilibrium<sup>42</sup>. 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 ??).

#### 3.1 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<sup>43</sup>. 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 ?? . 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.

Previously we used microsecond-scale molecular dynamics simulations to determine the energy landscapes based on equilibrium population distributions. How-



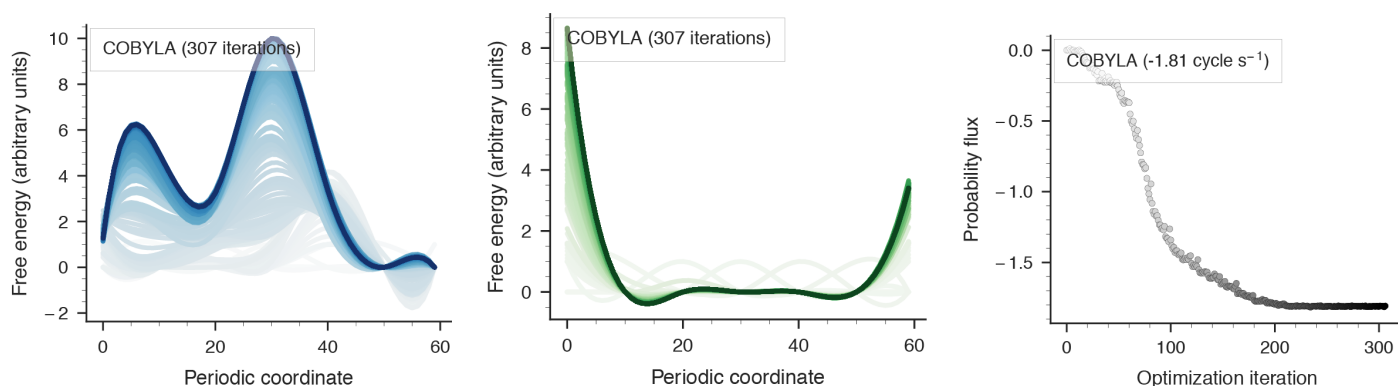
**Figure 4:** On the left, the four ground state conformations of a second generation motor, adapted from Štacko et al<sup>44</sup>. On the right, the same four states placed on free energy profiles. The energy profiles are periodic, with two cycles shown for both isomerization state. A clone of the lower *E* energy surface is shown above, for clarity, to demonstrate the progression from state 4 to state 1 requires energy.

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 ??, for a given input level of light irradiation.

### 3.2 Optimize the design of artificial molecular motors

As demonstrated by Richard Feynman in a lecture on Brownian ratchets, the challenge of designing molecular motors is not how to create movement, but how to control the directionality of the omnipresent – and random – motion on the molecular scale<sup>45</sup>. This has been referred to as the “gating” of stochastic motion<sup>46</sup>. In my previous work<sup>3</sup>, I showed that the gating can arise purely from the complementary shape of two potential energy surfaces, whereby an unpassable barrier on one surface can be circumnavigated by motion on the other surface (see Figure 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.

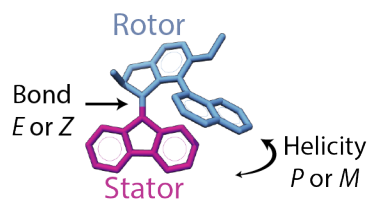
It has been empirically determined that adding electron donating and electron withdrawing moieties has the potential to affect the rotation rate of the motor<sup>47</sup>. Polarizing the molecule by placing electron withdrawing and electron donating groups on either end of the central double bond, has 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 ?? 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.

#### 4 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.



**Figure 6:** The two degrees of freedom in a synthetic molecular motor.

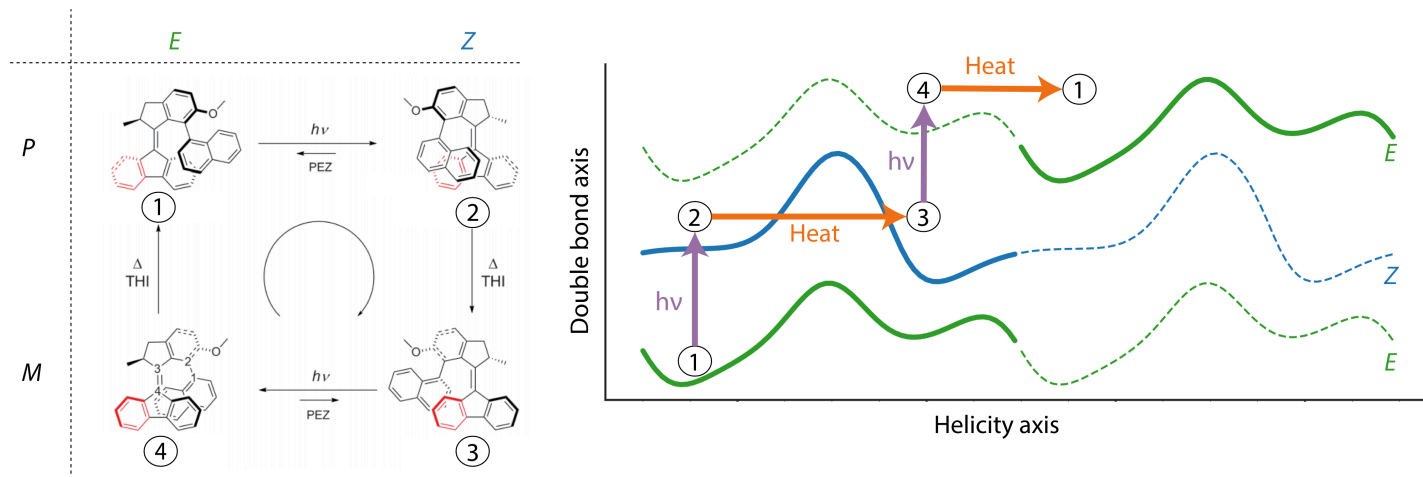
The ability of light-driven molecular motors to switch between two or more states makes them suitable for new forms of optical data storage<sup>36</sup>, as “wheels” on nano-scale cars<sup>37</sup>, trains<sup>38</sup>, worms<sup>39</sup>, and walkers<sup>40</sup>, or in new forms of responsive materials<sup>41</sup>. Unlike switches, whose work is reversed after every full cycle, molecular motors can be used to progressively move systems away from thermodynamic equilibrium<sup>42</sup>. 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 ??).



#### 4.1 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<sup>43</sup>. 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 ?? . 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 → 2, 3 → 4) and two thermal steps (2 → 3, 4 → 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.



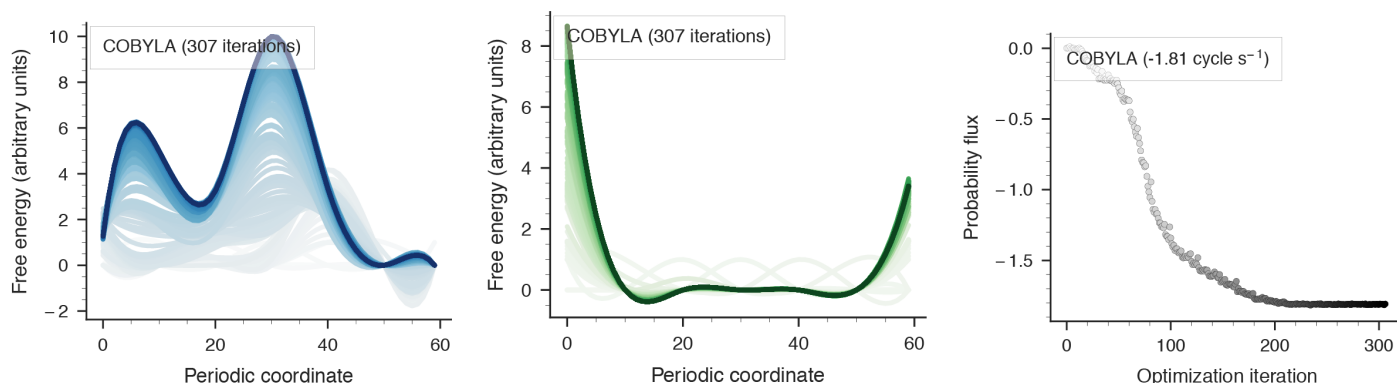
**Figure 7:** On the left, the four ground state conformations of a second generation motor, adapted from Štacko et al<sup>44</sup>. On the right, the same four states placed on free energy profiles. The energy profiles are periodic, with two cycles shown for both isomerization state. A clone of the 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 determine the energy landscapes based on equilibrium population distributions. However, 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 → 1), as illustrated by the series of purple and orange arrows in Figure ??, for a given input level of light irradiation.

## 4.2 Optimize the design of artificial molecular motors

As demonstrated by Richard Feynman in a lecture on Brownian ratchets, the challenge of designing molecular motors is not how to create movement, but how to control the directionality of the omnipresent – and random – motion on the molecular scale<sup>45</sup>. This has been referred to as the “gating” of stochastic motion<sup>46</sup>. In my previous work<sup>3</sup>, I showed that the gating can arise purely from the complementary shape of two potential energy surfaces, whereby an unpassable barrier on one surface can be circumnavigated by motion on the other surface (see Figure 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<sup>47</sup>. 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 ?? 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.

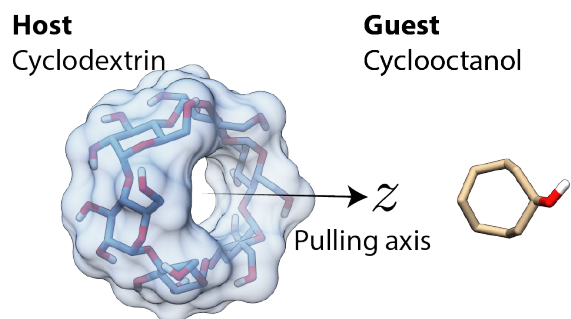


**Figure 8:** Example optimization (using the constrained optimization by linear approximation algorithm, COBYLA) of motor surfaces for maximum flux. Later iterations are darker colors. On the left, the  $Z$  surface; in the middle, the  $E$  surface; on the right, the net probability flux corresponding to each optimization iteration.

## 5 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<sup>48,49</sup>. 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<sup>50</sup>. The root mean squared error (RMSE) associated with such calculations is in the range of 1-4 kcal/mol<sup>51</sup>, yet even a modest improvement of the prediction accuracy of  $\sim 1$  kcal/mol would lead to a substantial decrease in the number of compounds that must be manually screened by nearly an order of magnitude<sup>52</sup>. 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<sup>53</sup>. 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.



**Figure 9:** An example host-guest system,  $\alpha$ -cyclodextrin with cyclooctanol (unbound) showing the pulling coordinate along the  $z$  axis.

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,  $\Delta G^\circ$ .

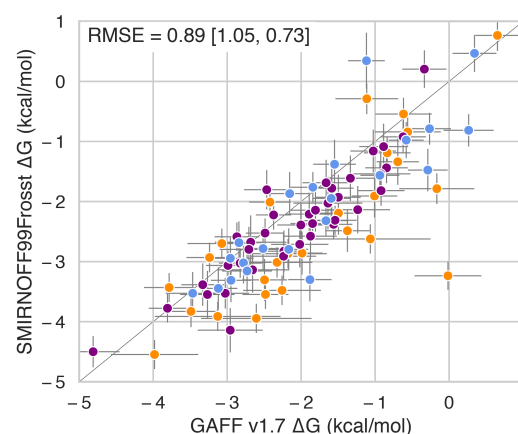
### 5.1 Incorporate host-guest binding data into force field development

APR has consistently been ranked among the most accurate methods for predicting binding thermodynamics in blind challenges<sup>1,54</sup>. 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<sup>55</sup>.

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<sup>56</sup>, 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<sup>57</sup>. 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<sup>58</sup>, and it appears that SMIRNOFF99Frosst reduces this tendency. To begin, I will use the “benchmark set” of Mobley et al.<sup>59</sup>. There, experimental data are available for rigid cucurbituril 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<sup>60</sup>. 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  $\tau$ , 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.



**Figure 10:** A comparison of binding free energies between SMIRNOFF99Frosst and GAFF v1.7 for a series of cyclodextrin hosts and guests (unpublished results). Points are colored according to guest functional group.

## 5.2 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<sup>61</sup> and protein force fields<sup>62</sup>. 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<sup>63</sup>, 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<sup>64</sup>.

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  $i$ ,  $\partial G_{\text{SEM}}^{\circ} / \partial N_{\text{frames}}^i$ . 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.

## 6 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<sup>48,49</sup>. 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<sup>50</sup>. The root mean squared error (RMSE) associated with such calculations is in the range of 1-4 kcal/mol<sup>51</sup>, 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<sup>52</sup>. 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<sup>53</sup>. 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,  $\Delta G^\circ$ .

### 6.1 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<sup>1,54</sup>. 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<sup>55</sup>.

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<sup>56</sup>, 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<sup>57</sup>. The first prototype force field from this effort was released 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<sup>58</sup>, and it appears that SMIRNOFF99Frosst reduces this tendency. To begin, I will use the “[benchmark set](#)” of Mobley et al.<sup>59</sup>. There, experimental data are available for rigid curcubituril 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<sup>60</sup>.

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  $\tau$ , 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.

## 6.2 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<sup>61</sup> and protein force fields<sup>62</sup>. 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<sup>63</sup>, 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<sup>64</sup>.

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  $i$ ,  $\partial G_{\text{SEM}}^{\circ} / \partial N_{\text{frames}}^i$ . 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**

## References

- (1) 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.
- (2) Yin, J.; Henriksen, N. M.; Slochower, D. R.; Gilson, M. K. The SAMPL5 Host–Guest Challenge: Computing Binding Free Energies and Enthalpies from Explicit Solvent Simulations by the Attach–Pull–Release (APR) Method. *J Comput Aided Mol Des* **2016**, *31* (1), 133–145.
- (3) Slochower, D. R.; Gilson, M. K. Motor-Like Properties of Non-motor Enzymes. *Biophysical Journal* **2018**, *114* (9), 2174–2179.
- (4) Slochower, D. R.; Huwe, P. J.; Radhakrishnan, R.; Janmey, P. A. Quantum and All-Atom Molecular Dynamics Simulations of Protonation and Divalent Ion Binding to Phosphatidylinositol 4,5-Bisphosphate (PIP<sub>2</sub>). *J. Phys. Chem. B* **2013**, *117* (28), 8322–8329.
- (5) Slochower, D. R.; Wang, Y.-H.; Tourdot, R. W.; Radhakrishnan, R.; Janmey, P. A. Counterion-Mediated Pattern Formation in Membranes Containing Anionic Lipids. *Advances in Colloid and Interface Science* **2014**, *208*, 177–188.
- (6) Wang, Y.-H.; Slochower, D. R.; Janmey, P. A. Counterion-Mediated Cluster Formation by Polyphosphoinositides. *Chemistry and Physics of Lipids* **2014**, *182*, 38–51.
- (7) Slochower, D. R.; Wang, Y.-H.; Radhakrishnan, R.; Janmey, P. A. Physical Chemistry and Membrane Properties of Two Phosphatidylinositol Bisphosphate Isomers. *Phys. Chem. Chem. Phys.* **2015**, *17* (19), 12608–12615.
- (8) Slochower, D. R.; Wang, Y.-H.; Radhakrishnan, R.; Janmey, P. A. Lipid Membrane Shape Evolution and the Actin Cytoskeleton. In *Handbook of Lipid Membranes, Molecular and Materials aspects*; Safinya, C., Rädler, J., Series Eds.; Taylor & Francis Publishers, 2018.
- (9) Di Paolo, G.; De Camilli, P. Phosphoinositides in Cell Regulation and Membrane Dynamics. *Nature* **2006**, *443* (7112), 651–657.
- (10) Faiderbe, S.; Chagnaud, J.-L.; Geffard, M. Anti-Phosphoinositide Auto-Antibodies in Sera of Cancer Patients: Isotypic and Immunochemical Characterization. *Cancer Letters* **1992**, *66* (1), 35–41.
- (11) Lin, W.; Leung, L. W.; Bae, Y. S.; Bittman, R.; Arthur, G. Effects of a Water-Soluble Antitumor Ether Phosphoinositide, d-Myo-Inositol 4-(Hexadecyloxy)-3(S)-Methoxybutanephosphonate (C4-Pi), on Inositol Lipid Metabolism in Breast Epithelial Cancer Cell Lines. *Biochemical Pharmacology* **1999**, *57* (10), 1153–1158.
- (12) Katso, R.; Okkenhaug, K.; Ahmadi, K.; White, S.; Timms, J.; Waterfield, M. D. Cellular Function of Phosphoinositide 3-Kinases: Implications for Development, Immunity, Homeostasis, and Cancer. *Annu. Rev. Cell Dev. Biol.* **2001**, *17* (1), 615–675.



- (13) Engelman, J. A. The Role of Phosphoinositide 3-Kinase Pathway Inhibitors in the Treatment of Lung Cancer. *Clinical Cancer Research* **2007**, *13* (15), 4637s–4640s.
- (14) Miled, N.; Yan, Y.; Hon, W.-C.; Perisic, O.; Zvelebil, M.; Inbar, Y.; Schneidman-Duhovny, D.; Wolfson, H. J.; Backer, J. M.; Williams, R. L. Mechanism of Two Classes of Cancer Mutations in the Phosphoinositide 3-Kinase Catalytic Subunit. *Science* **2007**, *317* (5835), 239–242.
- (15) Bunney, T. D.; Katan, M. Phosphoinositide Signalling in Cancer: Beyond PI3K and PTEN. *Nat Rev Cancer* **2010**, *10* (5), 342–352.
- (16) Lemmon, M. A. Membrane Recognition by Phospholipid-Binding Domains. *Nat Rev Mol Cell Biol* **2008**, *9* (2), 99–111.
- (17) McLaughlin, S.; Wang, J.; Gambhir, A.; Murray, D. PIP2 and Proteins: Interactions, Organization, and Information Flow. *Annu. Rev. Biophys. Biomol. Struct.* **2002**, *31* (1), 151–175.
- (18) Bradley, R.; Radhakrishnan, R. Curvature-Undulation Coupling as a Basis for Curvature Sensing and Generation in Bilayer Membranes. *Biophysical Journal* **2017**, *112* (3), 220a–221a.
- (19) Kooijman, E. E.; King, K. E.; Gangoda, M.; Gericke, A. Ionization Properties of Phosphatidylinositol Polyphosphates in Mixed Model Membranes. *Biochemistry* **2009**, *48* (40), 9360–9371.
- (20) van den Bogaart, G.; Meyenberg, K.; Risselada, H. J.; Amin, H.; Willig, K. I.; Hubrich, B. E.; Dier, M.; Hell, S. W.; Grubmüller, H.; Diederichsen, U.; et al. Membrane Protein Sequestering by Ionic Protein–Lipid Interactions. *Nature* **2011**, *479* (7374), 552–555.
- (21) Honigsmann, A.; van den Bogaart, G.; Iraheta, E.; Risselada, H. J.; Milovanovic, D.; Mueller, V.; Müller, S.; Diederichsen, U.; Fasshauer, D.; Grubmüller, H.; et al. Phosphatidylinositol 4,5-Bisphosphate Clusters Act as Molecular Beacons for Vesicle Recruitment. *Nat Struct Mol Biol* **2013**, *20* (6), 679–686.
- (22) Wang, J.; Richards, D. A. Segregation of PIP2 and PIP3 into Distinct Nanoscale Regions Within the Plasma Membrane. *Biology Open* **2012**, *1* (9), 857–862.
- (23) Wang, Y.-H.; Collins, A.; Guo, L.; Smith-Dupont, K. B.; Gai, F.; Svitkina, T.; Janmey, P. A. Divalent Cation-Induced Cluster Formation by Polyphosphoinositides in Model Membranes. *J. Am. Chem. Soc.* **2012**, *134* (7), 3387–3395.
- (24) Wen, Y.; Vogt, V. M.; Feigenson, G. W. Multivalent Cation-Bridged PI(4,5)P<sub>2</sub> Clusters Form at Very Low Concentrations. *Biophysical Journal* **2018**, *114* (11), 2630–2639.
- (25) Levental, I.; Christian, D. A.; Wang, Y.-H.; Madara, J. J.; Discher, D. E.; Janmey, P. A. Calcium-Dependent Lateral Organization in Phosphatidylinositol 4,5-Bisphosphate (PIP<sub>2</sub>)- and Cholesterol-Containing Monolayers. *Biochemistry* **2009**, *48* (34), 8241–8248.
- (26) BioPhysCode <https://github.com/biophyscode>.
- (27) BioPhysCode Portal <https://biophyscode.github.io/>.

- (28) Gidwani, A.; Holowka, D.; Baird, B. Fluorescence Anisotropy Measurements of Lipid Order in Plasma Membranes and Lipid Rafts from RBL-2H3 Mast Cells†. *Biochemistry* **2001**, *40* (41), 12422–12429.
- (29) Swamy, M. J.; Ciani, L.; Ge, M.; Smith, A. K.; Holowka, D.; Baird, B.; Freed, J. H. Coexisting Domains in the Plasma Membranes of Live Cells Characterized by Spin-Label ESR Spectroscopy. *Biophysical Journal* **2006**, *90* (12), 4452–4465.
- (30) Levental, I.; Byfield, F.; Chowdhury, P.; Gai, F.; Baumgart, T.; Janmey, P. Cholesterol-Dependent Phase Separation in Cell-Derived Giant Plasma-Membrane Vesicles. *Biochem. J.* **2009**, *424* (2), 163–167.
- (31) Sengupta, P.; Holowka, D.; Baird, B. Fluorescence Resonance Energy Transfer Between Lipid Probes Detects Nanoscopic Heterogeneity in the Plasma Membrane of Live Cells. *Biophysical Journal* **2007**, *92* (10), 3564–3574.
- (32) Baumgart, T.; Hammond, A. T.; Sengupta, P.; Hess, S. T.; Holowka, D. A.; Baird, B. A.; Webb, W. W. Large-Scale Fluid/Fluid Phase Separation of Proteins and Lipids in Giant Plasma Membrane Vesicles. *Proceedings of the National Academy of Sciences* **2007**, *104* (9), 3165–3170.
- (33) Ellenbroek, W.; Wang, Y.-H.; Christian, D.; Discher, D.; Janmey, P.; Liu, A. Divalent Cation-Dependent Formation of Electrostatic PIP2 Clusters in Lipid Monolayers. *Biophysical Journal* **2011**, *101* (9), 2178–2184.
- (34) Heinrich, M.; Tian, A.; Esposito, C.; Baumgart, T. Dynamic Sorting of Lipids and Proteins in Membrane Tubes with a Moving Phase Boundary. *Proceedings of the National Academy of Sciences* **2010**, *107* (16), 7208–7213.
- (35) Raghunathan, K.; Kenworthy, A. K. Dynamic Pattern Generation in Cell Membranes: Current Insights into Membrane Organization. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2018**, *1860* (10), 2018–2031.
- (36) Feringa, B.; Koumura, N.; van Delden, R.; ter Wiel, M. Light-Driven Molecular Switches and Motors. *Appl Phys A* **2002**, *75* (2), 301–308.
- (37) Kudernac, T.; Ruangsapichat, 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.
- (38) 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.
- (39) Sasaki, T.; Tour, J. M. Synthesis of a New Photoactive Nanovehicle: A Nanoworm. *Org. Lett.* **2008**, *10* (5), 897–900.
- (40) von Delius, M.; Leigh, D. A. Walking Molecules. *Chem. Soc. Rev.* **2011**, *40* (7), 3656.

- (41) 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.
- (42) 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.
- (43) 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.
- (44) 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.
- (45) Browne, W. R.; Feringa, B. L. Making Molecular Machines Work. *Nature Nanotech* **2006**, *1* (1), 25–35.
- (46) Astumian, R. D. Stochastic Pumping of Non-Equilibrium Steady-States: How Molecules Adapt to a Fluctuating Environment. *Chem. Commun.* **2018**, *54* (5), 427–444.
- (47) 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).
- (48) van Vlijmen, H.; Desjarlais, R. L.; Mirzadegan, T. Computational Chemistry at Janssen. *J Comput Aided Mol Des* **2016**, *31* (3), 267–273.
- (49) 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.
- (50) 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.
- (51) 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.
- (52) 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.
- (53) 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.
- (54) 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.

(55) 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.

(56) 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.

(57) 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.

(58) 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.

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

(60) 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.

(61) 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.

(62) 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.

(63) 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.

(64) 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.