UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Theory and Modeling of Molecular Motion out of Equilibrium

Permalink

https://escholarship.org/uc/item/12x7b7s9

Author

Feng, Mudong

Publication Date

2022

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA SAN DIEGO

Theory and Modeling of Molecular Motion out of Equilibrium

A Dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Chemistry

by

Mudong Feng

Committee in charge:

Professor Michael K. Gilson, Chair
Professor Rommie Amaro, Co-Chair
Professor J. Andrew McCammon
Professor Douglas Smith
Professor Wei Wang

2022

Copyright

Mudong Feng, 2022

All rights reserved.

The Dissertation of Mudong Feng is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ACKNOWLEDGEMENTS

I acknowledge Dr. Michael Gilson for his excellent support and mentorship during my time at UCSD. His passion and rigor in research impressed me during our fruitful hours of in-depth scientific discussions. I enjoyed a high degree of autonomy in exploring research directions, and great work-life balance. My enjoyable experience at Gilson lab motivates me to continue working as a researcher in the field of computational chemistry after graduation.

I acknowledge all the friends, colleagues and researchers I've met during my PhD years, such as Drs. David Slochower, Jamie Schiffer, Lee-Ping Wang, Niel Henriksen, Germano Heinzelmann, and so on. They greatly helped my research career in various ways.

I acknowledge my wife and my best friend, Jianghua Wang, whose company has been a blessing.

Chapter 1, in full, is a reprint of the material as it appears in "Feng M, Gilson MK. Enhanced Diffusion and Chemotaxis of Enzymes. Annual Review of Biophysics. 2020". The dissertation author was the primary author of this publication.

Chapter 2, in full, is a reprint of the material as it appears in "Feng M, Gilson MK. A Thermodynamic Limit on the Role of Self-Propulsion in Enhanced Enzyme Diffusion. Biophysical Journal. 2019". The dissertation author was the primary author of this publication.

Chapter 3, in full, is a reprint of the material as it appears in "Feng M, Gilson MK. Mechanistic analysis of light-driven overcrowded alkene-based molecular motors by multiscale molecular simulations. Phys Chem Chem Phys. 2021". The dissertation author was the primary author of this publication.

VITA

2016 Bachelor of Science, Peking University

2022 Doctor of Philosophy, University of California San Diego

ABSTRACT OF THE DISSERTATION

Theory and Modeling of Molecular Motion out of Equilibrium

by

Mudong Feng

Doctor of Philosophy in Chemistry

University of California San Diego, 2022

Professor Michael K. Gilson, Chair

Molecules at temperatures above 0K are always in motion, translating, rotating, and undergoing conformational changes. In systems that are out of equilibrium, these motions often become more intense and complex, leading to interesting phenomena, including the existence of life. This dissertation presents theoretical and computational modeling for some of these phenomena. First, many enzymes appear to diffuse faster in the presence of their substrates and to drift along concentration gradients of their substrate, phenomena known respectively as enhanced enzyme diffusion and enzyme chemotaxis. Here, experimental findings and proposed mechanisms for these observations are critically reviewed, then we

propose a kinematic and thermodynamic analysis to serve as a validity check for any mechanism that attributes enhanced enzyme diffusion to self-propulsion. Second, overcrowded alkene-based molecular motors, a class of synthetic small molecules designed for light-driven rotation of its rotor part relative to its stator part, exhibit fast rotation in the microsecond timescale. Here, the full rotation process is modeled by quantum surface-hopping molecular dynamics simulations coupled with classical molecular dynamics simulations. This study proposes a novel rotation pathway, as well as providing computational predictions for rotation rate and maximal power output. Encouraging agreement with experiments are found, after fitting critical forcefield parameters to reference quantum mechanical energy surfaces. In conclusion, these efforts contribute to better understanding of molecular motions out of equilibrium and how to conceptualize and model them.

CHAPTER 1: ENHANCED DIFFUSION AND CHEMOTAXIS OF ENZYMES

Annu. Rev. Biophys. 2020.49:87-105. Downloaded from www.annualreviews.org
Access provided by University of California - San Diego on 02/11/22. For personal use only.

ANNUAL REVIEWS

Annual Review of Biophysics

Enhanced Diffusion and Chemotaxis of Enzymes

Mudong Feng[1] and Michael K. Gilson[1,2]

[1]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, California 92093, USA; email: mgilson@ucsd.edu

[2]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, California 92093, USA

Annu. Rev. Biophys. 2020. 49:87–105

First published as a Review in Advance on January 27, 2020

The *Annual Review of Biophysics* is online at biophys.annualreviews.org

https://doi.org/10.1146/annurev-biophys-121219-081535

Copyright © 2020 by Annual Reviews. All rights reserved

ANNUAL REVIEWS CONNECT

www.annualreviews.org

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

Keywords

Fokker-Planck equation, microfluidics, fluorescence correlation spectroscopy, FCS, urease, catalase, aldolase

Abstract

Many enzymes appear to diffuse faster in the presence of substrate and to drift either up or down a concentration gradient of their substrate. Observations of these phenomena, termed enhanced enzyme diffusion (EED) and enzyme chemotaxis, respectively, lead to a novel view of enzymes as active matter. Enzyme chemotaxis and EED may be important in biology and could have practical applications in biotechnology and nanotechnology. They are also of considerable biophysical interest; indeed, their physical mechanisms are still quite uncertain. This review provides an analytic summary of experimental studies of these phenomena and of the mechanisms that have been proposed to explain them and offers a perspective on future directions for the field.

Annu. Rev. Biophys. 2020.49:87-105. Downloaded from www.annualreviews.org
Access provided by University of California - San Diego on 02/11/22. For personal use only.

Contents

1. INTRODUCTION

Studies over the past decade suggest that nonmotor enzymes engage in active, translational motion. Two phenomena have been noted. One, termed enhanced enzyme diffusion (EED), is an increase in the enzymes' diffusion constant induced, typically, by provision of substrate. The other, enzyme chemotaxis, is a tendency for enzymes to move up or down a substrate concentration gradient. These phenomena have been seen for both fast exothermic enzymes like urease and slow endothermic enzymes like aldolase. These results open a new perspective of enzymes as active matter (27) and have potential practical and biological implications. For example, enzyme chemotaxis can be used to separate catalytically active and inactive enzymes (12) and might contribute to the assembly of intracellular metabolons (69, 75) and to intracellular signaling (65). Several mechanisms have been proposed for these phenomena, but their physical basis is still a subject of active research.

In this review, we analyze relationships among the various experimental and theoretical studies and discuss general theoretical considerations and frameworks for this field. [Related reviews also summarize recent work on EED and chemotaxis (1, 4, 13, 73).] First for EED and then for chemotaxis, we review experimental findings and analyze molecular mechanisms that have been proposed to account for these findings. Central questions include how to explain observations of EED and chemotaxis in general, whether—or when—EED is an active process, and how to reconcile apparently inconsistent experimental observations. We conclude with a summary of key points and potential directions for future work to further elucidate these intriguing phenomena.

Enhanced enzyme diffusion (EED): an apparent speedup of translational enzyme diffusion, usually reported as an increase in the diffusion coefficient

Enzyme chemotaxis: apparent drift of an enzyme up (attractive) or down (repulsive) a chemical gradient

Active mechanism: proposed mechanism of EED or chemotaxis is active if it requires the release of chemical energy by enzyme catalysis

2. EXPERIMENTAL STUDIES OF ENHANCED ENZYME DIFFUSION

This section summarizes experimental reports of increases in enzyme diffusion coefficients, focusing on enzymes for which an active mechanism, such as execution of swimming motions by the

Table 1 Turnover rates, hydrodynamic radii, minimum thrust speeds, and required reaction free energies of enzymes reported to show enhanced enzyme diffusion

Enzyme	Turnover rate (s^{-1})	Radius (nm)	v_{min} ($m.s^{-1}$)	$-\Delta G^{\circ}_{req}$ ($kJ.mol^{-1}$)
T4 DNA polymerase	0.5 (7)	4.6 (9)	1×10^{-2}	1×10^{7}
Aldolase	5 (25)	4.9 (11)	9×10^{-3}	8×10^{5}
T7 RNA polymerase	4 (39)	6.0[a]	6×10^{-3}	6×10^{5}
Hexokinase	300 (75)	6.3 (38)	5×10^{-3}	6,000
ATP synthase	1,000 (23)	6.6 (8)	5×10^{-3}	2,000
Alkaline phosphatase	3,000 (46)	7.7 (15)	3×10^{-3}	400
Catalase	10,000 (46)	5.3 (11)	7×10^{-3}	300
Urease	10,000 (46)	7.0 (18)	4×10^{-3}	100
Acetylcholinesterase	20,000 (19)	8.8 (47)	3×10^{-3}	40

[a]Calculated with HYDROPRO (44) from Protein Data Bank structure 4RNP (58).
For details, see Reference 17. Citations for individual cells are in parentheses.

Annu. Rev. Biophys. 2020.49:87-105. Downloaded from www.annualreviews.org
Access provided by University of California - San Diego on 02/11/22. For personal use only.

enzyme, has been proposed. We also discuss cases in which EED was expected but not observed and consider potential experimental artifacts.

2.1. Positive Observations

Apparent EED has been reported for various enzymes, including ATPase (8), T7 RNA polymerase (71), T4 DNA polymerase (55), hexokinase (75), aldolase (75), alkaline phosphatase (46), acetylcholinesterase (28), jack bean urease (27, 28, 43, 46, 54, 70), and catalase (29, 46, 54). The relative diffusion enhancements, measured in homogeneous solution at the highest tested substrate concentrations, range from 15% to 80% (**Table 1**). Although these increases are similar, the turnover rates of these enzymes span orders of magnitude (**Table 1**). This discrepancy argues against an active mechanism as a general explanation of EED. Most of these studies used fluorescence correlation spectroscopy (FCS), which measures enzyme diffusion rates in homogeneous solution. However, one (71) used fluorescence recovery after photobleaching, and another used a relatively novel electrochemical method to support its FCS observation of EED (29).

Two FCS studies conducted by Jee and coworkers (27, 28) provide FCS data at increased spatial and temporal resolution. By combining superresolution microscopy and FCS, these authors were able to reduce the horizontal diameter of the observation region—the beam waist—to 50 nm, versus more typical values of approximately 800 nm (43). As the beam waist fell below approximately 100 nm, an initially unimodal distribution of transit times was resolved into two peaks. The peak corresponding to longer transit times was attributed to conventional diffusion through the waist, and the peak corresponding to shorter times was attributed to fast ballistic motions induced by catalysis.

Another study measured the diffusion of fluorescently labeled urease by single-molecule tracking with total internal reflectance fluorescence microscopy (70). This method detects molecules only while they are approximately 300 nm from the planar glass coverslip, so the enzyme molecules were confined in this layer by addition of methylcellulose to the solution. This study yielded a remarkable 300% increase in the diffusion coefficient upon addition of 1 mM urea. Interestingly, the enhanced diffusion coefficient observed here, $\sim 3 \times 10^{-13}$ m^2/s, is approximately 200-fold less than the nonenhanced value measured by FCS for urease in homogeneous solution (54). This drop in the diffusion coefficient might result from increased viscosity due to the methylcellulose. It is perhaps relevant that increasing the viscosity reduces the power required to achieve a given

Annu. Rev. Biophys. 2020.49:87-105. Downloaded from www.annualreviews.org
Access provided by University of California - San Diego on 02/11/22. For personal use only.

relative increment in the translational diffusion coefficient via self-propulsion (see Section 3.1, Equation 2) (17). However, such a large change in the baseline diffusion coefficient raises the possibility that the physics of diffusion in this confined setting is significantly different from that in bulk solution.

2.2. Evidence For and Against an Active Mechanism of Enhanced Diffusion

The concept that EED results from an active process has gained support from FCS and dynamic light scattering (DLS) studies indicating that catalytically active urease, catalase, and aldolase generate increased motion of passive tracer particles in solution (28, 74), with a magnitude and reaction-rate dependence similar to those of the enzymes' own enhanced diffusion. Similarly, immobilized DNA polymerase exhibits EED and generates fluid flow proportional to the catalytic rate (55). The observation of a 20% increase in the diffusion coefficient of passive tracers is particularly striking for aldolase given that its turnover rate is only approximately 5/s and that it was present at a concentration of 10 nM.

If EED results from catalysis, then enzyme inhibitors should prevent EED. Accordingly, EED has not been observed in catalase in the presence of both substrate and the inhibitor cyanide (54) or azide (46). However, at least three enzymes described as having catalysis-induced EED have also been reported to show EED in the absence of catalysis. First, Yu and coworkers (71) reported that RNA polymerase has 25% EED when catalytically active and 15% EED when substrate is provided but the required cofactor Mg^{2+} is withheld. Second, heat release had been proposed as a mechanistic requirement for EED (46); however, Illien and coworkers (25) later showed 30% EED for the enzyme aldolase, although this enzyme catalyzes an endothermic reaction (75). This led to the idea that catalysis itself might not be required for EED by aldolase, and the same authors used FCS to show that aldolase's competitive inhibitor pyrophosphate could generate nearly the same level of EED as its substrate, fructose-1,6-bisphosphate. Third, urease did not show EED in the presence of the inhibitor pyrocatechol alone at 1 mM concentration but did show attenuated EED in the presence of both pyrocatechol and substrate (43). Another study of urease found that a 1 mM concentration of the substrate urea sufficed to cause EED, while the urease inhibitor boric acid began to cause EED only at higher concentrations (approximately 100 mM) (27). It was therefore proposed that substrate and inhibitor cause EED by two different mechanisms, to account for inconsistencies among experimental results as to when or whether EED requires catalysis (27).

2.3. Negative Observations and Possible Experimental Artifacts

Significant inconsistencies have emerged in studies of aldolase across multiple techniques. Although an FCS study (25) indicated approximately 30% EED in the presence of either substrate or a competitive inhibitor, Zhang and coworkers (72) studying aldolase using DLS found no EED in the presence of either substrate or inhibitor. Günther and coworkers (22) also observed no EED when studying aldolase using a third technique, diffusion ordered spectroscopy (DOSY) nuclear magnetic resonance (NMR). It is not yet clear how to reconcile all of these results, but another study from Günther and coworkers (21) highlights potential artifacts and complexities of the widely used FCS technique.

One potential source of error in FCS is that there is always some free fluorophore, so if more protein binds to the glass over time, then the relative contribution of the fast-diffusing fluorophore to the measured diffusion coefficient increases, leading to an artifactual increase in the apparent diffusion coefficient (21). However, the results of Illien and coworkers (25) argue

against this artifact: They found that the elevated diffusion coefficient of aldolase in the presence of substrate returned to baseline once the substrate was consumed.

Quenching of the fluorophore by substrate or product could also lead to errors, as suggested by Bai & Wolynes (6). Indeed, Günther and coworkers (21) showed that this artifact can account for apparent EED of the enzyme alkaline phosphatase (46) when the quenching substrate nitrophenyl phosphate is used because the apparent EED disappears when a nonquenching substrate is used. Given that EED in catalase has been studied by FCS, it is worth noting that its substrate, hydrogen peroxide, can also act as a quencher (21). In contrast, an FCS study of EED in urease argues against the quenching artifact by confirming that urea does not reduce the fluorescence lifetime of the fluorescent label (27).

Finally, FCS measurements are typically carried out at enzyme concentrations roughly 1,000 times lower than those of DLS and DOSY measurements (21, 72). This makes it more probable that some of the multimers have dissociated in the FCS studies. Given that binding of substrate and/or inhibitor molecules sometimes promotes dissociation (8, 68), binding could increase the measured diffusion coefficient merely by causing enzymes to dissociate into faster-diffusing subunits. In any case, because most reports of EED, for all enzymes studied, rely exclusively on FCS, it would seem important to track down the cause of the discrepancy for aldolase and/or to apply alternative experimental methods to other enzymes.

Annu. Rev. Biophys. 2020.49:87-105. Downloaded from www.annualreviews.org
Access provided by University of California - San Diego on 02/11/22. For personal use only.

3. POTENTIAL MECHANISMS OF ENHANCED ENZYME DIFFUSION

In this section, we analyze proposed mechanistic explanations of EED. These are divided into two categories, propulsive and nonpropulsive. Before discussing specific mechanisms, we consider the thermodynamics of active self-propulsion and the possibility of a role for hydrodynamic interactions among enzymes in solution. It is worth emphasizing at the outset that, to be plausible, a mechanism must meet both qualitative and quantitative criteria. That is, the proposed mechanism must not only be physically workable but also capable of generating EED at the levels observed experimentally when realistic values of parameters, such as k_{cat} and hydrodynamic radius, are considered. For example, in principle, an enzyme's diffusion coefficient could rise due to heating of the solution by the enzyme-catalyzed reaction. However, under normal conditions, the temperature does not rise nearly enough to account for observed levels of EED (see Section 3.4.2).

3.1. A Thermodynamic Constraint on Enzyme Self-Propulsion

Varied physical mechanisms have been proposed by which chemical energy released via enzyme catalysis could lead to propulsion, thus increasing an enzyme's translational diffusion coefficient. Any propulsion mechanism necessarily leads to dissipation of chemical energy, such as by viscous drag opposing the propelled motion. Therefore, the entire class of self-propulsion mechanisms is plausible only when the catalyzed reaction provides enough power to match the unavoidable dissipation. We recently analyzed this generally applicable limit, quantitatively connecting theory with experimental data (17), as summarized in this section.

A fundamental aspect of self-propulsion mechanisms is that the enzyme is an asymmetric particle, which is considered to be propelled in a given direction within its own frame of reference. As a consequence, the enzyme's rotational diffusion causes its lab-frame propulsion direction to change stochastically, according to the rotational diffusion constant D_r. The overall translational diffusion of such a self-propelled particle results from normal Brownian motion combined with this randomly oriented self-propulsion. Using the Stokes-Einstein law, one can write an apparent diffusion coefficient as the sum of the nonenhanced translational diffusion coefficient D_t and a

contribution from propulsion with speed v:

$$D_{app} = D_t + \frac{v^2}{6D_r}\frac{t_p}{t_c}, \quad 1.$$

where $\frac{t_p}{t_c}$ is the mean fraction of the catalytic cycle during which the propulsive force is present. The minimal power required to achieve a certain ratio of diffusion enhancement, $R = \frac{D_{app}}{D_t} - 1$, is given by the Stokes drag dissipation rate (37, 48, 49, 59, 62):

$$P_{req} = 6\pi\eta a v^2 \frac{t_p}{t_c} = \frac{3}{4}\frac{R(kT)^2}{\pi\eta a^3}, \quad 2.$$

where a is the hydrodynamic radius, and η is viscosity. Using the hydrodynamic radii and turnover numbers of enzymes reported to undergo EED allows one to estimate the free energy per reaction required to generate experimental levels of diffusion enhancement (17). The required free energies (**Table 1**) are orders of magnitude larger than those available for the slower enzymes and considerably larger than those available even for the fastest enzymes, despite the use of conservative assumptions in the power analysis (17).

The strong inverse dependence of P_{req} on hydrodynamic radius a helps explain how self-propulsion can still lead to significant enhanced diffusion for larger particles with asymmetrically disposed catalytic sites, such as so-called Janus nanoparticles (36). Intuitively, the orders-of-magnitude slower rotational diffusion of these larger particles makes the self-propulsion trajectory much less tortuous and more effective at generating net displacement from an initial location.

As part of this analysis, we considered the chemical free energy available to power propulsion. In principle, the chemical power available to drive propulsion is the product of the reaction rate and the free energy change of the reaction under the experimental conditions studied. This free energy change depends on the concentrations of reactants and products. However, global concentrations do not couple to the local processes of substrate binding, chemical reaction, and product release, so the global free energy of reaction may not be what is available to drive propulsion; it may instead be a local free energy (3) that is relevant. We used a thermodynamic cycle to estimate local free energy changes of enzyme-catalyzed reactions and found that, fortuitously, these are generally close to the standard reaction free energies (17). Experimental measurements of enzyme diffusion when the reaction is at equilibrium or running in reverse might test these ideas and shed light more generally on mechanisms of EED.

3.2. The Insignificance of Hydrodynamic Interactions

The thermodynamic analysis in Section 3.1 assumes that the enzyme molecules in solution move independently. However, EED might be amplified if the motions of each enzyme molecule could be further driven by the motions of others. Because the enzyme solutions used in FCS measurements are dilute, typically 10 nM, the average distance between two enzyme molecules is approximately 550 nm. At this range, any intermolecular forces are extremely weak. For example, the Coulombic interaction between two enzymes of charge +10, assuming the dielectric constant of water, is less than 0.1 kcal·mol^{-1}, even neglecting ionic screening. However, hydrodynamic interactions among enzyme molecules might provide a mechanism for longer-ranged enzyme–enzyme correlations. Indeed, several studies have indicated that enzyme activity, including that of aldolase, can generate fluid flows (55) or enhanced diffusion of passive tracer molecules (74), possibly via hydrodynamic interactions. In addition, Sengupta and coworkers observed no

Annu. Rev. Biophys. 2020.49:87-105. Downloaded from www.annualreviews.org
Access provided by University of California - San Diego on 02/11/22. For personal use only.

increase in the diffusion coefficient of fluorescently labelled, inactivated catalysis in the presence of unlabeled, active catalase (54).

A theoretical study has shown that hydrodynamic interactions are proportional to the concentration of active enzyme, and it was argued that hydrodynamic interactions among enzymes could lead to significant increments in diffusion coefficients (40). However, the numerical calculation in this study assumed an enzyme concentration of 1,000 nM, far higher than those used in FCS measurements. If the typical concentration of 10 nM is used, then the predicted enhancement in diffusion coefficient by hydrodynamic interactions comes to only 6×10^{-13} m^2/s, much less than reported values of EED. From another perspective, the flow velocity field generated by an enzyme decays no slower than r^{-2} (14, 40). Thus, in a 10 nM solution, the flow velocity generated near the surface of one enzyme (radius ~10 nm) will have decayed at least approximately 3,000-fold at the position of another enzyme (distance ~550 nm). Thus, it seems unlikely that hydrodynamic interactions contribute significantly to EED. Future measurements examining the magnitude of EED as a function of enzyme concentration might offer further insight.

3.3. Self-Propulsion Mechanisms

In this section, we consider potential mechanisms of EED that involve self-propulsion powered by the chemical energy of the enzyme-catalyzed reaction. Note that the theoretical considerations discussed above apply to all of these possibilities.

3.3.1. Mechanical swimming. By mechanical swimming, we mean the generation of propulsion by a repeated cycle of conformational changes. Microorganisms engage in mechanical swimming (14, 35), but swimming by nonmotor enzymes is not well established. The scallop theorem implies that a simple cycle of forward and reverse conformational changes cannot generate net propulsion, as any motion induced by the forward step will be undone by the reverse step (34). However, the chiral character of enzymes means that motions driven by an out-of-equilibrium chemical reaction will be directional and hysteretic (56) and thus capable of generating net propulsion (56). Furthermore, even if the conformational changes were perfectly time reversable, they could generate an increase in the translational diffusion coefficient because rotational diffusion of the enzyme during the enzymatic cycle allows the motion generated by the forward step to be along a different lab-frame axis than the motion generated by the reverse step (33). Nonetheless, mechanical swimming is an unlikely explanation for EED, based on the thermodynamic argument provided in Section 3.1 and on another study indicating that any plausible enzyme motions are too weak to generate observed levels of EED (6).

3.3.2. Pressure waves. Riedel and coworkers (46) suggested that the rapid release of heat at the catalytic step of an exothermic, enzyme-catalyzed reaction could generate a pressure wave that produces an asymmetric force on the enzyme, leading to self-propulsion. Bai & Wolynes (6) argued against this mechanism by showing that an extremely large conformational motion, along the lines of complete unfolding and refolding, would be needed. Our interpretation of the original suggestion is that the pressure wave comes not from a fast conformational change but from sudden heating at the catalytic site due to the chemical reaction. This view might avoid the concern raised by Bai & Wolynes; however, it would still be unclear how passage of a single pressure pulse through the enzyme at the speed of sound could generate a large net displacement of the enzyme. (Note that the passage of sound waves through water or air does not lead to net displacement of the molecules forming the medium.) Golestanian (20) has also provided a theoretical argument against this proposed mechanism.

Annu. Rev. Biophys. 2020.49:87-105. Downloaded from www.annualreviews.org
Access provided by University of California - San Diego on 02/11/22. For personal use only.

Annu. Rev. Biophys. 2020.49:87-105. Downloaded from www.annualreviews.org
Access provided by University of California - San Diego on 02/11/22. For personal use only.

3.3.3. Bubble propulsion. The enzyme catalase has molecular oxygen as a product, and a sufficiently high density of catalase molecules on a surface can generate oxygen bubbles, leading to a propulsive force (13, 45, 50). However, generation of bubbles was, arguably, ruled out as a mechanism for self-propulsion of catalase by direct observation and by demonstration that active catalase does not increase the diffusion coefficient of nearby passive tracer molecules (54). In addition, most of the enzymes for which EED has been reported do not create a potentially gaseous product.

3.3.4. Phoretic self-propulsion. Phoretic mechanisms play an important role in self-propelled synthetic particles, such as Janus nanomotors (51), and the theory underlying these mechanisms is well-developed (42). Self-phoretic mechanisms include self-diffusiophoresis, self-electrophoresis, and self-thermophoresis, which result, respectively, from interactions of a particle with self-induced gradients of concentration, electrical potential, or temperature. These mechanisms are unlikely to explain EED because the thermodynamic limit on self-propulsion discussed above applies to phoretic self-propulsion. In addition to frictional dissipation, phoretic self-propulsion would require extra power to maintain the self-induced gradient (49), further increasing the gap between the required power and the power available from the chemical reaction. Nonetheless, it is informative to consider specific phoretic mechanisms that have been put forward.

Self-electrophoresis has been suggested as a mechanism of EED (43), but it seems unlikely. First, EED has been reported for enzymes whose substrates and products are electrically neutral, so they cannot set up an electric field. Second, self-electrophoresis would be influenced by ionic strength, but ionic strength was reported to have no influence on the apparent EED of RNA polymerase (71). Third, although self-electrophoresis has been reported for Janus nanomotors, these have spatially separated ionic flows at their cathodes and anodes and are thus well-suited to create ionic gradients and resulting electrical fields. In contrast, enzymes usually bind substrate and expel product at the same site and therefore are not as good at generating gradients.

Arguing in favor of a self-diffusiophoresis mechanism, Colberg & Kapral (10) presented simulations of enzyme-sized particles undergoing diffusiophoresis at high propulsion speeds of roughly 4 m/s. However, their study assumes a diffusion-controlled reaction with a rate constant of approximately 4×10^{10}/(M·s), which is much larger than the enzyme turnover rates listed in **Table 1**. It also assumes a concentration of substrate much higher than that in enzyme systems—the substrate was essentially a solvent—so the diffusiophoretic forces could be unrealistically large. Finally, the magnitudes of the enzyme–substrate and enzyme–product forces are arbitrary, rather than being chosen to reflect typical enzyme–substrate interactions. Thus, their model does not closely resemble an enzyme–substrate system and is of limited applicability.

Finally, Golestanian estimated the magnitude of the potential self-thermophoretic self-propulsion in catalase and found it to be negligible (20); self-thermophoresis should be even weaker for other enzymes that release less heat.

3.4. Nonpropulsive Mechanisms

In this section, we consider mechanisms of EED that do not involve self-propulsion and therefore avoid the concerns raised above regarding the inadequacy of the chemical power available.

3.4.1. Local pH change. Muddana and coworkers (43) considered whether EED could, for some enzymes at least, result from local changes in pH due to enzymatic activity. For example, ammonia, one of the products of the urease reaction, is basic and therefore can change the pH. However, measurement of the pH in the immediate vicinity of urease with a pH-sensitive fluorophore covalently bound to the enzyme revealed pH increases of up to only ~0.8, which were

judged insufficient to explain EED. Note, too, that many enzymes for which EED has been reported cannot change the pH.

3.4.2. Temperature increase. Another way for catalysis to increase the diffusion constant of an enzyme would be for the released heat (if any) to increase the temperature of the solution. However, even for an exothermic enzyme-catalyzed reaction, the increase in the bulk temperature of the solution is far too small to account for observed EED (20, 46, 60). Golestanian (20) has proposed a more refined mechanism, collective heating, which accounts for the nonequilibrium heat flow in the measuring container and for the increase in enzyme turnover with increasing temperature, and has argued that these factors can lead to a large enough temperature increase and viscosity decrease to account for EED. However, Golestanian's study used the thermal conductivity of air in its numerical analysis, and when the 30-fold higher value of water is used instead, as would seem appropriate, the predicted temperature increase no longer appears to be sufficient.

3.4.3. Changes in conformation, conformational fluctuations, and quaternary structure. If binding of substrate generates a new conformational distribution of the enzyme with a smaller mean hydrodynamic radius, then this would lead to EED. In addition, the degree of EED would correlate with the enzyme's catalytic rate, as is often observed, because increasing substrate concentration will increase both the catalytic rate and the fraction of bound enzyme. This explanation is appealing because it avoids the power requirements of propulsive mechanisms. It could also account for observations of EED induced by binding of an inhibitor or by addition of substrate to an enzyme that lacks an essential cofactor. However, experimental data compiled from the literature (73) and Brownian dynamics simulations (31) suggest that binding of a substrate or an inhibitor does not cause a large enough reduction in an enzyme's mean radius of gyration to account for observed levels of EED.

Recently, Illien and coworkers showed theoretically that the diffusion coefficient can be increased by a decrease in the thermal fluctuations of the particle radius (24, 25). They note that an enzyme may be stiffened by binding of another molecule, so addition of substrate or inhibitor could lead to concentration-dependent EED, as observed experimentally. However, the Brownian dynamics study mentioned above (31) suggests that this mechanism would not lead to observed levels of EED.

Finally, Günther and coworkers (21) have pointed out that many of the enzymes for which EED has been reported are multimeric, and that binding of substrate can lead to dissociation of multimeric enzymes. Because dissociation into smaller components would lead to an increase in the diffusion coefficient, averaged over the various multimers in solution, binding-induced dissociation could provide another nonpropulsive explanation for observations of EED. This mechanism may be particularly relevant for FCS measurements, which are typically run at low enzyme concentrations that shift the equilibrium toward dissociated states (34). Dimeric yeast hexokinase, for which EED has been observed via FCS (75), has a dissociation constant of 0.1–1.0 M, which is well above the 10-nM concentrations typically used in FCS experiments. Furthermore, three enzymes for which EED has been observed, hexokinase, urease, and acetylcholinesterase, were reported to dissociate as substrate concentrations rose above their respective k_M values (26). It is thus worth noting that, if different-sized multimers interconvert on an appropriate timescale (tens of microseconds), then this would cause an additional decay mode in the FCS autocorrelation function and thus could offer an alternative explanation for the appearance of a bimodal distribution of transit times in high-resolution FCS studies (27, 28). However, dissociation does not seem like a viable explanation for at least some observations of EED because the tetramer-to-dimer dissociation constant of aldolase is approximately 1 pM (72), so this enzyme should be quite stable as a multimer in most or all of the relevant experiments;

Annu. Rev. Biophys. 2020.49:87-105. Downloaded from www.annualreviews.org
Access provided by University of California - San Diego on 02/11/22. For personal use only.

Annu. Rev. Biophys. 2020.49:87-105. Downloaded from www.annualreviews.org
Access provided by University of California - San Diego on 02/11/22. For personal use only.

the single-molecule tracking study of urease by Xu and coworkers (70) reported no influence of the concentrations of urea or enzyme on dissociation or EED. In addition, there is evidence of EED by urease at substrate concentrations where little dissociation is evident (26).

4. EXPERIMENTAL OBSERVATIONS OF ENZYME CHEMOTAXIS

Motile bacteria and eukaryotic cells have evolved mechanisms for swimming up or down gradients of dissolved compounds. This directed movement of cells is called chemotaxis, and movement up or down a gradient is termed attractive or repulsive chemotaxis, respectively. Bacteria use a molecular memory system to determine whether their recent motions have taken them up or down the gradient and adapt accordingly, while eukaryotic cells use their size to sense the direction of a gradient across the cell in real time (52). In recent years, it has been reported that enzymes can also move preferentially up or down a substrate gradient (28, 54), with apparent drift speeds up to approximately 1 μm/s for attractive enzyme chemotaxis (54) and 10 μm/s for repulsive chemotaxis (28).

Enzymes cannot meet the strictest definition of chemotaxis (66, 67) because they lack the memory and/or size required to mimic either the bacterial or eukaryotic mechanisms. Instead, the apparent directional migration of enzymes may arise from factors such as space-dependent enzyme diffusivity or diffusiophoresis, as discussed in Section 5. In this review, we follow the literature by applying the term chemotaxis to all observations of enzymes moving preferentially up or down a concentration gradient.

4.1. Attractive Chemotaxis

An early report of attractive enzyme chemotaxis used fluorescence to detect preferential displacement of RNA polymerase up a gradient of its nucleotide triphosphate substrate in a millimeter-scale device (71). Subsequent experiments have measured fluorescence intensity profiles of labeled enzymes across microfluidic flow channels fed by incoming channels containing solutions with differing compositions (**Figure 1**). For example, a left feed might contain either plain buffer or substrate in buffer, and a right feed might contain enzyme in buffer; in this case, attractive chemotaxis could manifest by a tendency of the enzyme to move to the left more in the presence of substrate than in its absence (**Figure 1*a***). Because flow in microfluidic devices is laminar, diffusion, rather than convection, dominates the relaxation of the initial nonequilibrium concentration. Such devices showed preferential diffusion of catalase and urease toward their respective substrates (12, 54) and of DNA polymerase toward either its substrate or its Mg^{2+} cofactor in the presence of substrate (55). A technically similar study of mitochondrial malate dehydrogenase and citrate synthase showed chemotaxis toward substrate in both the presence and absence of required enzyme cofactors (69). Thus, catalysis is not always necessary for enzyme chemotaxis to be observed.

A potential weakness of studies like those above is that enhanced movement of enzyme into the region of the microfluidic channel with substrate might result only from faster diffusion in the presence of substrate, rather than from any directional preference (27). Studies by Sen and coworkers address this (41, 75). Thus, when hexokinase and its substrates D-glucose and ATP were injected into the center of the channel, with plain buffer on both sides, the spread of the active enzyme across the channel was slower than when the same experiment was done with inactive enzyme. In effect, the enzyme was retained in the central channel. This focusing result implies a real tendency of the enzyme to remain close to the substrate. In addition, catalysis appears to be important in this effect because chemotaxis was attenuated when mannose, a substrate with slower turnover, was used instead of D-glucose (41, 75), and chemotaxis was absent in situations with D-glucose but no ATP, and in situations with both ATP and L-glucose present, although the latter is not a substrate (41).

Annu. Rev. Biophys. 2020.49:87-105. Downloaded from www.annualreviews.org
Access provided by University of California - San Diego on 02/11/22. For personal use only.

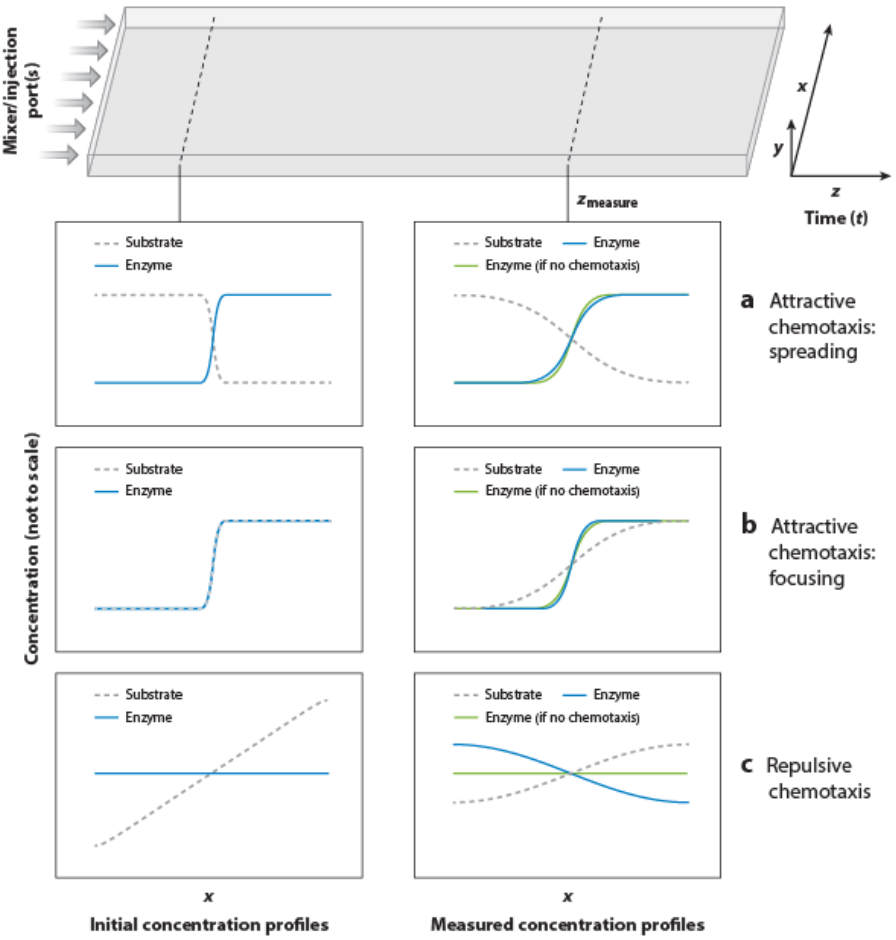

Figure 1

Generic diagram of a microfluidic chemotaxis experiment, with laminar flow from left to right, i.e., in the $+z$ direction. Concentration profiles of enzyme (*solid blue* or *green*) and substrate (*dashed gray*) across the channel (x axis) are initialized by injection ports and mixers (initial concentration profiles in the plot) and measured (measured concentration profiles in the plot) after the concentration profiles have relaxed for some time t, corresponding to distance $z_{measure}$ from the injection ports. (*a*) Concentration profiles corresponding to a transient lurch of enzyme toward substrate, suggestive of attractive chemotaxis (54). (*b*) Retention, or focusing, of enzyme in the high-substrate part of the channel, suggestive of attractive chemotaxis (75). (*c*) Repulsive chemotaxis in the context of an initially linear substrate profile (27).

4.2. Repulsive Chemotaxis

Paradoxically, Jee and coworkers (28), also using a microfluidic device, reported repulsive chemotaxis for urease and acetylcholinesterase, as an initially uniform enzyme concentration increased on the side of the channel with lower substrate concentration. Agudo-Canalejo et al. (2) have suggested that this difference, relative to prior experiments, might result from the use of a different range of substrate concentrations, but the observation was later confirmed over a

wider concentration range (27, 41). It may also be relevant that Jee and coworkers used a different experimental design, in which the enzyme was initially at uniform concentration across the channel (**Figure 1**). In addition, it appears that Jee and coworkers' measurements (28) allowed more interaction time (distance from start of channel × channel area/fluid flow rate) than did the studies that observed attractive chemotaxis (41, 54). Thus, the experiments of Jee and coworkers might reveal something closer to the shape of the ultimate steady-state distribution of the enzyme.

4.3. The Microfluidic Method

Two important features of the microfluidic experiments are worth highlighting. First, they do not give the steady-state distribution of an enzyme in the context of a time-invariant substrate gradient, because the concentration profiles of not only the enzyme but also the substrate change as the fluid progresses along the channel. Instead, they report a transient response to a time-varying substrate gradient. Moreover, the diffusion coefficient of the substrate is usually much higher than that of the enzyme, so the concentration profile of the substrate normally relaxes more quickly than that of the enzyme. This is important because some physical mechanisms can explain transient movement of the enzyme along a substrate gradient but lead to a uniform distribution at steady state under a time-invariant substrate gradient. The microfluidic experiments may reveal the transient response but not the steady-state response. It is perhaps worth mentioning that a microfluidic setup can, in another sense, be said to have reached its steady state after enough wall-clock time has passed for its flows and concentration profiles to have stabilized. This is different from the question of whether one has taken a measurement far enough down the channel—i.e., long enough after mixing has begun—for the concentration profiles to have stabilized as a function of the mixing time. Second, when interpreting the concentration distribution of enzyme across the channel, it should be borne in mind that laminar flow has a parabolic profile, so detailed interpretation requires accounting for the details of fluid flow by simulations, as has previously been done (41).

5. POTENTIAL MECHANISMS OF ENZYME CHEMOTAXIS

Potential mechanisms of enzyme chemotaxis may be grouped into two categories. On one hand, it is posited that a force directed parallel to the substrate (or inhibitor) concentration gradient drives either attractive or repulsive chemotaxis. The other potential mechanism is based on the idea that the enzyme's diffusion coefficient depends on the substrate concentration, leading to a position-dependent diffusion coefficient when there is a substrate concentration gradient. Importantly, as emphasized by Agudo-Canalejo and coworkers (2), both categories of mechanisms can be at work in the same system. Before considering specific mechanisms, however, we present a theoretical framework for defining and analyzing proposed mechanisms of enzyme chemotaxis.

5.1. A Theoretical Framework for Mechanisms of Enzyme Chemotaxis

Enzymes translate through solvent in a Brownian manner and can thus be modeled by the Fokker-Planck (FP) equation (76). This equation describes how an initial probability distribution (i.e., concentration) in space, $p(x, t = 0)$, evolves over time, and we simplify it by considering a one-dimensional system. For example, if x represents the distance of a point from one edge of a microfluidic channel (**Figure 1**), an initial step-function concentration profile across the inlet end of the channel decays over time to a sigmoid-like distribution and approaches uniformity as the flow progresses further down the channel. The FP equation obeys conservation of probability, which requires that $\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}J$, where $J = J(x, t)$ is the probability flux. If the diffusion coefficient,

Annu. Rev. Biophys. 2020.49:87-105. Downloaded from www.annualreviews.org
Access provided by University of California - San Diego on 02/11/22. For personal use only.

D, is constant, then the flux can be written as $J = \mu F p - D\frac{\partial p}{\partial x}$, where the first term represents directional drift induced by a force F, and μ is the enzyme's mobility (the reciprocal of the friction coefficient).

If an enzyme diffuses at a different rate in the presence of substrate, such as by EED, then the enzyme's diffusion coefficient will depend on position, so $D = D(x)$. [The diffusion coefficient could also be a function of time, $D = D(x, t)$, because the substrate gradient in a microfluidic device decays with time (Section 4.3). However, the mathematical consequences of this complication are not considered in current literature on enzyme chemotaxis and thus are not considered in this review.] Perhaps surprisingly, merely specifying $D(x)$ does not fully determine the correct form of FP equation. This is because the flux expression depends on the character of the microscopic process that causes $D(x)$ to vary with x—an issue known as the Ito-Stratonovich dilemma, which arises in systems with multiplicative noise (16, 32, 57, 61). A range of scenarios is captured by the following expression: $J = \mu F p - \alpha\frac{\partial D(x)}{\partial x} p - D(x)\frac{\partial p}{\partial x}$, where $0 < \alpha < 1$. This leads to the following relatively general form of the FP equation:

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}(\mu F p) + \alpha\frac{\partial}{\partial x}\left[p\frac{\partial D(x)}{\partial x}\right] + \frac{\partial}{\partial x}\left[D(x)\frac{\partial p}{\partial x}\right]. \quad 3.$$

In this case, $\alpha\frac{\partial D(x)}{\partial x} p$ is an additional term that could contribute to chemotaxis. In particular, if $\alpha = 1$, then the steady-state distribution of enzyme will be greater where the diffusion coefficient is lower (2, 27). In contrast, if $\alpha = 0$, then the steady-state distribution will be uniform in the absence of a force F (27, 61). The FP equation with $\alpha = 1$ is often termed the Ito form, while the FP equation with $\alpha = 0$ is termed the isothermal form (16, 57), and we use these names below.

A fundamental conclusion of this analysis is that enzyme chemotaxis cannot be mechanistically explained by merely positing a position-dependent diffusion coefficient induced by a substrate gradient. This is because the drift of the enzyme depends on α, which depends on the microscopic origin of the position dependence of $D(x)$. Nonetheless, the FP equation is a valuable framework for understanding diffusive motion, and the following section uses it to consider various possible mechanisms of enzyme chemotaxis.

5.2. Mechanisms Based on Force-Induced Drift

If a substrate gradient leads to a net force, F, on enzyme molecules in solution, then this will induce drift either up (attractive) or down (repulsive) the gradient, which could account for experimental observations of enzyme chemotaxis. This class of mechanism could explain the enzyme focusing result of Sen and coworkers (41, 75) and the evolution of an initially uniform enzyme profile into a nonuniform one in the presence of a substrate gradient (27). Force-induced drift mechanisms do not require a space-dependent diffusion coefficient, so multiplicative noise is not an issue. In this section, we consider two specific proposals for mechanisms in this class.

5.2.1. Thermodynamic force. One mechanism derives an expression for a time-averaged force on macromolecules arising from the thermodynamics of macromolecule–cosolute binding in the presence of a concentration gradient of the cosolute. Schurr and coworkers (53) included this concept in their analysis of chemotaxis of nonenzyme molecules, and good correlation with experiment was obtained for a case of enzyme chemotaxis (75). Mohajerani and coworkers (41) modified this theory for enzymes, arguing that catalysis-associated EED could magnify the effect by increasing the baseline diffusion coefficient of the enzyme and thus the speed of chemotaxis, and they reported agreement with experimental data. The fundamental picture of this model is that

Annu. Rev. Biophys. 2020.49:87-105. Downloaded from www.annualreviews.org
Access provided by University of California - San Diego on 02/11/22. For personal use only.

Annu. Rev. Biophys. 2020.49:87-105. Downloaded from www.annualreviews.org
Access provided by University of California - San Diego on 02/11/22. For personal use only.

a free enzyme molecule tends to move in the direction of higher substrate concentration when it binds, whereas an enzyme–substrate complex does not have any directional preference.

A concern with this proposed mechanism is that it equates the position-averaged force on the enzyme, which is computed from the thermodynamic gradient, with the time-averaged force on the enzyme, which is the quantity relevant for chemotaxis. It is the time-averaged force that is relevant because the instantaneous velocity of an overdamped enzyme molecule is proportional to the instantaneous force, and we are interested in the mean of this velocity over time. If the free energy fell linearly with position, then the time-averaged and space-averaged forces would be equal. However, the free energy falls only at moments when the enzyme is transitioning from the unbound state to the substrate-bound state. These transitions are short lived because the binding forces are short ranged, and these brief transitions are separated by long time intervals during which the enzyme feels no directing force. The lengths of these intervals are governed by the association and dissociation rate constants. Thus, an enzyme diffusing in solution feels the binding force for only a small fraction of the time, and the time-averaged force on the enzyme is expected to be far smaller than the position-averaged force, so it is not clear how well this proposed mechanism can account for enzyme chemotaxis. This reasoning appears analogous to that in a prior study showing that the Stokes efficiency of a molecular motor is less than one and may be very low indeed when the driving potential is not linear in the spatial variable along which the motor moves (64).

Agudo-Canalejo and coworkers (1) have offered additional points of concern regarding the thermodynamic model, such as the fact that it cannot account for observations of repulsive chemotaxis (Section 4.2). The possibility of further complexities with a mechanism based on the thermodynamics of enzyme–substrate binding is suggested by a prior observation that thermodynamics alone does not determine the phoretic speeds, or even the directions, of colloid particles (3, p. 94). More broadly, determining molecular motions requires knowing more than a thermodynamic tendency, i.e., an energy gradient; one must also know how chemical coordinates and mechanical coordinates are coupled to transduce this energy (5, 63).

5.2.2. Diffusiophoresis. Diffusiophoresis causes directional drift of colloid particles up or down the concentration gradient of a cosolute (3, 30). It results from net attractive or repulsive forces between the particles and the cosolutes, leading to attractive or repulsive chemotaxis. Diffusiophoresis is similar to self-diffusiophoresis (Section 3.3.4) except that the gradient is externally imposed, instead of being self-generated by catalysis.

The particle–cosolute interactions that cause diffusiophoresis are typically nonspecific, often involve long-ranged electrostatics, and may be averaged across the surface of a relatively large colloid particle (42, 53). We are not aware of studies deriving diffusiophoretic velocities for binding of an enzyme to its substrate, but Agudo-Canalejo and coworkers (2) have argued that diffusiophoresis driven by nonspecific enzyme–substrate interactions can play a key mechanistic role in enzyme chemotaxis. However, further work is needed to assess whether the nonspecific interactions between an enzyme and its substrate are in fact capable of driving enzyme chemotaxis at observed rates via a diffusiophoretic mechanism. It is also worth noting that the magnitude and direction of nonspecific interactions may depend significantly on whether the enzyme has a bound substrate, particularly if the substrate has a nonzero net electrical charge.

5.3. Mechanisms Based on a Position-Dependent Diffusion Coefficient

If an enzyme has a higher diffusion coefficient in the presence of its substrate, then the diffusion coefficient of the enzyme will be position dependent in the presence of a substrate concentration gradient. The consequences of a position-dependent diffusion coefficient for experimental

observations of chemotaxis are complex and case dependent (Section 5.1). For one thing, the transient effect of a position-dependent diffusion coefficient may be different from its steady-state effect. Transiently, enhanced diffusion in the presence of substrate may cause enzyme molecules introduced into the center of a microfluidic channel to diffuse preferentially to a side of the channel that has substrate, relative to a side that contains only buffer (**Figure 1**), as has previously been noted (27). This initial lurch in the direction of increasing diffusion coefficient has been termed pseudochemotaxis because it does not result from a directional preference of the diffusing particle (52), only from the fact that a nonuniform concentration distribution will relax faster where the diffusion coefficient is larger. By contrast, at steady state in the presence of a stable substrate gradient, the distribution of enzyme will depend on the value of α. If $\alpha = 0$ (isothermal form), then the steady-state concentration profile of the enzyme will be uniform, suggesting no chemotaxis. However, if $\alpha = 1$ (Ito form), then the steady-state concentration of the enzyme will be lower where the substrate is at a higher concentration (Section 5.1), matching experimental observations of repulsive chemotaxis (27, 28). In principle, the Ito form would also yield attractive chemotaxis at steady state, if the enzyme diffused more slowly in the presence of substrate, but this scenario has not been reported. Two microscopic mechanisms for a position-dependent diffusion coefficient and an Ito-form FP equation have been put forward.

One involves catalysis-driven self-propulsion of enzymes (28, 65); in effect, an enzyme gets extra propulsive kicks in the presence of substrate, leading to an increasing diffusion coefficient along a substrate gradient. This mechanism resembles that of a temperature gradient, where a particle gets more kinetic kicks in regions of higher temperature. Given that diffusion in a temperature gradient can lead to an Ito-form FP equation (32), it is likely that the self-propulsion mechanism does as well. However, the power available from enzyme-catalyzed reactions does not appear to be enough to account for the self-propulsion required by this proposed chemotaxis mechanism (Section 3.1).

The other mechanism does not involve self-propulsion, but instead involves a local equilibrium between substrate-bound and free forms of the enzyme having different diffusion coefficients. A novel derivation from Agudo-Canalejo and coworkers (2) shows that this scenario leads to the Ito-form FP equation with a position-dependent mean diffusion coefficient for the enzyme. Because this mechanism does not rely on chemical energy to power diffusive motion, it is not subject to the thermodynamic restriction of the self-propulsion mechanism.

It would be valuable to carry out more detailed and quantitative calculations that would test the ability of these mechanisms to account for the magnitudes of the effects seen experimentally. Specific questions include whether binding of substrate changes the hydrodynamic radius of urease enough to fit experiment, and whether the timescales of experiments showing repulsive chemotaxis (27) are long enough for enzyme molecules to drastically redistribute across the microfluidic channel due merely to space-dependent diffusion.

6. CONCLUSIONS

Enhanced diffusion and chemotaxis of enzymes have emerged in recent years as novel phenomena with potential implications in biology and biotechnology. They also pose intriguing puzzles whose resolution could yield new insights into molecular processes and experimental methods. Elucidation of the underlying mechanisms will require analysis of potentially subtle linkages among nonequilibrium processes spanning a range of scales. Ultimately, understanding the mechanisms should make it possible to design enzymes or other molecules to maximize these effects and put them to use. Continued work in this field promises new insights into the intricacies of molecular motions in out-of-equilibrium systems.

Annu. Rev. Biophys. 2020.49:87-105. Downloaded from www.annualreviews.org
Access provided by University of California - San Diego on 02/11/22. For personal use only.

Annu. Rev. Biophys. 2020.49:87-105. Downloaded from www.annualreviews.org
Access provided by University of California - San Diego on 02/11/22. For personal use only.

FUTURE ISSUES

1. What characteristics of enzymes—e.g., structural or catalytic—correlate with EED and/or chemotaxis?
2. Are there design principles to discover and even utilize?
3. Are different mechanisms at work in different enzymes?
4. What role, if any, does the catalytic release of chemical energy play in EED and enzyme chemotaxis?
5. Why do DOSY and DLS yield diffusion results that are so different from those of FCS for aldolase?
6. Is enzyme chemotaxis purely a transient phenomenon, or can it also be observed when the enzyme concentration profile is at steady state in a stable substrate gradient?
7. What determines whether a chemotactic enzyme will undergo attractive versus repulsive enzyme chemotaxis?

DISCLOSURE STATEMENT

M.K.G. has an equity interest in, and is a cofounder and scientic advisor of, VeraChem LLC.

ACKNOWLEDGMENTS

We thank Dr. Tsvi Tlusty for drawing our attention to the relevance of the various forms of the Fokker-Planck equation.

LITERATURE CITED

1. Agudo-Canalejo J, Adeleke-Larodo T, Illien P, Golestanian R. 2018. Enhanced diffusion and chemotaxis at the nanoscale. *Acc. Chem. Res.* 51:2365–72
2. Agudo-Canalejo J, Illien P, Golestanian R. 2018. Phoresis and enhanced diffusion compete in enzyme chemotaxis. *Nano Lett.* 18:2711–17
3. Anderson JL. 1989. Colloid transport by interfacial forces. *Annu. Rev. Fluid Mech.* 21:61–99
4. Astumian RD. 2014. Enhanced diffusion, chemotaxis, and pumping by active enzymes: progress toward an organizing principle of molecular machines. *ACS Nano* 8:11917–24
5. Astumian RD, Mukherjee S, Warshel A. 2016. The physics and physical chemistry of molecular machines. *ChemPhysChem* 17:1719–41
6. Bai X, Wolynes PG. 2015. On the hydrodynamics of swimming enzymes. *J. Chem. Phys.* 143:165101
7. Burrows JA, Goward CR. 1992. Purification and properties of DNA polymerase from *Bacillus caldotenax*. *Biochem. J.* 287:971–77
8. Brsch M, Turina P, Eggeling C, Fries JR, Seidel CA, et al. 1998. Conformational changes of the H^+-ATPase from *Escherichia coli* upon nucleotide binding detected by single molecule fluorescence. *FEBS Lett.* 437:251–54
9. Challberg MD, Englund PT. 1979. Purification and properties of the deoxyribonucleic acid polymerase induced by vaccinia virus. *J. Biol. Chem.* 254:7812–19
10. Colberg PH, Kapral R. 2014. Ångström-scale chemically powered motors. *Europhys. Lett.* 106:30004
11. de la Torre JG, Huertas ML, Carrasco B. 2000. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* 78:719–30

Annu. Rev. Biophys. 2020.49:87-105. Downloaded from www.annualreviews.org
Access provided by University of California - San Diego on 02/11/22. For personal use only.

12. Dey KK, Das S, Poyton MF, Sengupta S, Butler PJ, et al. 2014. Chemotactic separation of enzymes. *ACS Nano* 8:11941–49
13. Dey KK, Sen A. 2017. Chemically propelled molecules and machines. *J. Am. Chem. Soc.* 139:7666–76
14. Elgeti J, Winkler RG, Gompper G. 2015. Physics of microswimmers—single particle motion and collective behavior: a review. *Rep. Prog. Phys.* 78:056601
15. Ey PL, Ferber E. 1977. Calf thymus alkaline phosphatase: I. Properties of the membrane-bound enzyme. *Biochim. Biophys. Acta Enzymol.* 480:403–16
16. Farago O, Grønbech-Jensen N. 2014. Fluctuation–dissipation relation for systems with spatially varying friction. *J. Stat. Phys.* 156:1093–110
17. Feng M, Gilson MK. 2019. A thermodynamic limit on the role of self-propulsion in enhanced enzyme diffusion. *Biophys. J.* 116:1898–906
18. Follmer C, Pereira FV, da Silveira NP, Carlini CR. 2004. Jack bean urease (EC 3.5.1.5) aggregation monitored by dynamic and static light scattering. *Biophys. Chem.* 111:79–87
19. Froede HC, Wilson IB. 1984. Direct determination of acetyl-enzyme intermediate in the acetylcholinesterase-catalyzed hydrolysis of acetylcholine and acetylthiocholine. *J. Biol. Chem.* 259:11010–13
20. Golestanian R. 2015. Enhanced diffusion of enzymes that catalyze exothermic reactions. *Phys. Rev. Lett.* 115:108102
21. Günther JP, Börsch M, Fischer P. 2018. Diffusion measurements of swimming enzymes with fluorescence correlation spectroscopy. *Acc. Chem. Res.* 51:1911–20
22. Günther JP, Majer G, Fischer P. 2019. Absolute diffusion measurements of active enzyme solutions by NMR. *J. Chem. Phys.* 150:124201
23. Iino R, Hasegawa R, Tabata KV, Noji H. 2009. Mechanism of inhibition by C-terminal-helices of the subunit of *Escherichia coli* FoF1-ATP synthase. *J. Biol. Chem.* 284:17457–64
24. Illien P, Adeleke-Larodo T, Golestanian R. 2017. Diffusion of an enzyme: the role of fluctuation-induced hydrodynamic coupling. *Europhys. Lett.* 119:40002
25. Illien P, Zhao X, Dey KK, Butler PJ, Sen A, Golestanian R. 2017. Exothermicity is not a necessary condition for enhanced diffusion of enzymes. *Nano Lett.* 17:4415–20
26. Jee AY, Chen K, Tlusty T, Zhao J, Granick S. 2019. Enhanced diffusion and oligomeric enzyme dissociation. *J. Am. Chem. Soc.* 141:20062–68
27. Jee AY, Cho YK, Granick S, Tlusty T. 2018. Catalytic enzymes are active matter. *PNAS* 115:E10812–21
28. Jee AY, Dutta S, Cho YK, Tlusty T, Granick S. 2018. Enzyme leaps fuel antichemotaxis. *PNAS* 115:14–18
29. Jiang L, Santiago I, Foord J. 2017. Observation of nanoimpact events of catalase on diamond ultramicroelectrodes by direct electron transfer. *Chem. Commun.* 53:8332–35
30. Khair AS. 2013. Diffusiophoresis of colloidal particles in neutral solute gradients at finite Péclet number. *J. Fluid Mech.* 731:64–94
31. Kondrat S, Popescu MN. 2019. Brownian dynamics assessment of enhanced diffusion exhibited by fluctuating-dumbbell enzymes. *Phys. Chem. Chem. Phys.* 21:18811–15
32. Landauer R. 1983. Stability and relative stability in nonlinear driven systems. *Helv. Phys. Acta* 56:847–61
33. Lauga E. 2011. Enhanced diffusion by reciprocal swimming. *Phys. Rev. Lett.* 106:178101
34. Lauga E. 2011. Life around the scallop theorem. *Soft Matter* 7:3060–65
35. Lauga E, Powers TR. 2009. The hydrodynamics of swimming microorganisms. *Rep. Prog. Phys.* 72:096601
36. Lee TC, Alarcón-Correa M, Miksch C, Hahn K, Gibbs JG, Fischer P. 2014. Self-propelling nanomotors in the presence of strong Brownian forces. *Nano Lett.* 14:2407–12
37. Lighthill MJ. 1952. On the squirming motion of nearly spherical deformable bodies through liquids at very small Reynolds numbers. *Commun. Pure Appl. Math.* 5:109–18
38. Lilie H, Bar D, Kettner K, Weininger U, Balbach J, et al. 2011. Yeast hexokinase isoenzyme ScHxk2: stability of a two-domain protein with discontinuous domains. *Protein Eng. Des. Select.* 24:79–87
39. Martin CT, Coleman JE. 1987. Kinetic analysis of T7 RNA polymerase-promoter interactions with small synthetic promoters. *Biochemistry* 26:2690–96
40. Mikhailov AS, Kapral R. 2015. Hydrodynamic collective effects of active protein machines in solution and lipid bilayers. *PNAS* 112:E3639–44

Annu. Rev. Biophys. 2020.49:87-105. Downloaded from www.annualreviews.org
Access provided by University of California - San Diego on 02/11/22. For personal use only.

41. Mohajerani F, Zhao X, Somasundar A, Velegol D, Sen A. 2018. A theory of enzyme chemotaxis: from experiments to modeling. *Biochemistry* 57:6256–63
42. Moran JL, Posner JD. 2017. Phoretic self-propulsion. *Annu. Rev. Fluid Mech.* 49:511–40
43. Muddana HS, Sengupta S, Mallouk TE, Sen A, Butler PJ. 2010. Substrate catalysis enhances single-enzyme diffusion. *J. Am. Chem. Soc.* 132:2110–11
44. Ortega A, Amorós D, García de la Torre J. 2011. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys. J.* 101:892–98
45. Pantarotto D, Browne WR, Feringa BL. 2008. Autonomous propulsion of carbon nanotubes powered by a multienzyme ensemble. *Chem. Commun.* 13:1533–35
46. Riedel C, Gabizon R, Wilson CAM, Hamadani K, Tsekouras K, et al. 2014. The heat released during catalytic turnover enhances the diffusion of an enzyme. *Nature* 517:227–30
47. Rieger F, Bon S, Massoulie J, Cartaud J, Benda P. 1976. *Torpedo marmorata* acetylcholinesterase; a comparison with the *Electrophorus electricus* enzyme. Molecular forms, subunits, electron microscopy, immunological relationship. *Eur. J. Biochem.* 68:513–21
48. Sabass B, Seifert U. 2010. Efficiency of surface-driven motion: Nanoswimmers beat microswimmers. *Phys. Rev. Lett.* 105:218103
49. Sabass B, Seifert U. 2012. Dynamics and efficiency of a self-propelled, diffusiophoretic swimmer. *J. Chem. Phys.* 136:064508
50. Sanchez S, Solovev AA, Mei Y, Schmidt OG. 2010. Dynamics of biocatalytic microengines mediated by variable friction control. *J. Am. Chem. Soc.* 132:13144–45
51. Santiago I. 2018. Nanoscale active matter matters: challenges and opportunities for self-propelled nanomotors. *Nano Today* 19:11–15
52. Schnitzer MJ, Block SM, Berg HC. 1990. Strategies for chemotaxis. *Biol. Chemotactic Response* 46:15–34
53. Schurr JM, Fujimoto BS, Huynh L, Chiu DT. 2013. A theory of macromolecular chemotaxis. *J. Phys. Chem. B* 117:7626–52
54. Sengupta S, Dey KK, Muddana HS, Tabouillot T, Ibele ME, et al. 2013. Enzyme molecules as nanomotors. *J. Am. Chem. Soc.* 135:1406–14
55. Sengupta S, Spiering MM, Dey KK, Duan W, Patra D, et al. 2014. DNA polymerase as a molecular motor and pump. *ACS Nano* 8:2410–18
56. Slochower DR, Gilson MK. 2018. Motor-like properties of nonmotor enzymes. *Biophys. J.* 114:2174–79
57. Sokolov I. 2010. Ito, Stratonovich, Hnggi and all the rest: the thermodynamics of interpretation. *Chem. Phys.* 375:359–63
58. Sousa R, Chung YJ, Rose JP, Wang BC. 1993. Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. *Nature* 364:593–99
59. Stone HA, Samuel AD. 1996. Propulsion of microorganisms by surface distortions. *Phys. Rev. Lett.* 77:4102
60. Tsekouras K, Riedel C, Gabizon R, Marqusee S, Pressé S, Bustamante C. 2016. Comment on "Enhanced diffusion of enzymes that catalyze exothermic reactions" by R. Golestanian. arXiv:1608.05433 [physics.bio-ph]
61. Tupper PF, Yang X. 2012. A paradox of state-dependent diffusion and how to resolve it. *Proc. R. Soc. A* 468:3864–81
62. Wang H. 2005. Chemical and mechanical efficiencies of molecular motors and implications for motor mechanisms. *J. Phys. Condens. Matt.* 17:S3997–4014
63. Wang H, Oster G. 2002. Ratchets, power strokes, and molecular motors. *Appl. Phys. A* 75:315–23
64. Wang H, Oster G. 2002. The Stokes efficiency for molecular motors and its applications. *Europhys. Lett.* 57:134–40
65. Weistuch C, Pressé S. 2018. Spatiotemporal organization of catalysts driven by enhanced diffusion. *J. Phys. Chem. B* 122:5286–90
66. Wheat PM. 2011. *Collective behavior of swimming bimetallic motors in chemical concentration gradients*. Ph.D. thesis, Arizona State Univ., Tempe
67. Wilkinson PC. 1998. Assays of leukocyte locomotion and chemotaxis. *J. Immunol. Methods* 216:139–53
68. Woodfin BM. 1967. Substrate-induced dissociation of rabbit muscle aldolase into active subunits. *Biochem. Biophys. Res. Commun.* 29:288–93

69. Wu F, Pelster LN, Minteer SD. 2015. Krebs cycle metabolon formation: Metabolite concentration gradient enhanced compartmentation of sequential enzymes. *Chem. Commun.* 51:1244–47
70. Xu M, Ross JL, Valdez L, Sen A. 2019. Direct single molecule imaging of enhanced enzyme diffusion. *Phys. Rev. Lett.* 123:128101
71. Yu H, Jo K, Kounovsky KL, de Pablo JJ, Schwartz DC. 2009. Molecular propulsion: chemical sensing and chemotaxis of DNA driven by RNA polymerase. *J. Am. Chem. Soc.* 131:5722–23
72. Zhang Y, Armstrong MJ, Bassir Kazeruni NM, Hess H. 2018. Aldolase does not show enhanced diffusion in dynamic light scattering experiments. *Nano Lett.* 18:8025–29
73. Zhang Y, Hess H. 2019. Enhanced diffusion of catalytically active enzymes. *ACS Cent. Sci.* 5:939–48
74. Zhao X, Dey KK, Jeganathan S, Butler PJ, Córdova-Figueroa UM, Sen A. 2017. Enhanced diffusion of passive tracers in active enzyme solutions. *Nano Lett.* 17:4807–12
75. Zhao X, Palacci H, Yadav V, Spiering MM, Gilson MK, et al. 2017. Substrate-driven chemotactic assembly in an enzyme cascade. *Nat. Chem.* 10:311–17
76. Zwanzig R. 2001. *Nonequilibrium Statistical Mechanics*. Oxford, UK: Oxford Univ. Press

Annu. Rev. Biophys. 2020.49:87-105. Downloaded from www.annualreviews.org
Access provided by University of California - San Diego on 02/11/22. For personal use only.

Acknowledgements

Chapter 1, in full, is a reprint of the material as it appears in “Feng M, Gilson MK. Enhanced Diffusion and Chemotaxis of Enzymes. Annual Review of Biophysics. 2020”. The dissertation author was the primary author of this publication.

CHAPTER 2: A THERMODYNAMIC LIMIT ON THE ROLE OF SELF-PROPULSION IN ENHANCED ENZYME DIFFUSION

Biophysical Journal
Article

Biophysical Society

A Thermodynamic Limit on the Role of Self-Propulsion in Enhanced Enzyme Diffusion

Mudong Feng[1] and Michael K. Gilson[1,2,*]
[1]Department of Chemistry and Biochemistry and [2]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, California

ABSTRACT A number of enzymes reportedly exhibit enhanced diffusion in the presence of their substrates, with a Michaelis-Menten-like concentration dependence. Although no definite explanation of this phenomenon has emerged, a physical picture of enzyme self-propulsion using energy from the catalyzed reaction has been widely considered. Here, we present a kinematic and thermodynamic analysis of enzyme self-propulsion that is independent of any specific propulsion mechanism. Using this theory, along with biophysical data compiled for all enzymes so far shown to undergo enhanced diffusion, we show that the propulsion speed required to generate experimental levels of enhanced diffusion exceeds the speeds of well-known active biomolecules, such as myosin, by several orders of magnitude. Furthermore, the minimal power dissipation required to account for enzyme enhanced diffusion by self-propulsion markedly exceeds the chemical power available from enzyme-catalyzed reactions. Alternative explanations for the observation of enhanced enzyme diffusion therefore merit stronger consideration.

INTRODUCTION

The apparent diffusion coefficients of various enzymes, as measured typically by fluorescence correlation spectroscopy, have been observed to increase in the presence of substrate by as much as 15–80%, depending on the enzyme, at maximal substrate concentration. Examples include F0F1-ATP synthase (1), T7 RNA polymerase (2), T4 DNA polymerase (3), bovine catalase (4,5), jack bean urease (4–6), hexokinase (7), fructose biophosphatase aldolase (7,8), alkaline phosphatase (5), and acetylcholinesterase (9). However, the mechanisms underlying these observations remain largely unexplained. For some enzymes, further experimentation has ruled out certain potential mechanisms for this phenomenon of enhanced enzyme diffusion (EED), including one mediated by local pH changes (6) and propulsion by bubble formation (4). In a number of cases, the increase in diffusion coefficient relative to baseline has been found to be approximately proportional to the catalytic rate of the enzyme, with a Michaelis-Menten relationship to substrate concentration (5). This proportionality has naturally led to the suggestion that the chemical reaction catalyzed by the enzyme is a driver of the diffusion enhancement. Indeed, larger, synthetic Janus particles are propelled by the catalysis of reactions at one face of the particle and not the other (10). Accordingly, a number of possible mechanisms for catalysis-driven self-propulsion of enzymes—i.e., for the transduction of the reaction free energy into mechanical propulsion—have been proposed. These include mechanical swimming (4,11), pressure waves generated by exothermic reactions (5), and self-diffusiophoresis (12). However, these specific mechanisms of EED have been debated (8,13), and none have been proven. Here, we step back from specific propulsion mechanisms and instead analyze the kinematics and thermodynamics of enzyme self-propulsion generically.

METHODS

The degree to which translational diffusion is enhanced may be expressed as

$$\begin{aligned} D_{app} &= D_t + \Delta D \\ &= D_t(1+R) \end{aligned}, \qquad (1)$$

where $\Delta D = D_{app} - D_t$ is the difference between the observed, or apparent, diffusion constant, D_{app}, and the baseline diffusion constant in the absence of enhancement, D_t. Thus, R is the relative diffusion enhancement. We consider an enzyme that, within each catalytic cycle, self-propels for a time $t_p \leq t_c$, where t_c is the enzymologic turnover time and reciprocal of turnover rate. The magnitude of the propulsive force, F, is considered to be constant during t_p. (The consequences of a more complex time dependence are considered in the Appendix.) For an enzyme in liquid water, the Reynold's number is very low. Therefore, the dynamics of the enzyme are overdamped, and the propulsion velocity has a constant

Submitted November 21, 2018, and accepted for publication April 2, 2019.
*Correspondence: mgilson@ucsd.edu
Editor: David Sept.
https://doi.org/10.1016/j.bpj.2019.04.005
© 2019 Biophysical Society.

magnitude $v \propto F$ while the propulsion is active. The vector of the propulsive force and velocity is considered fixed within the enzyme's internal frame of reference, but it reorients continuously in the lab frame because of the rotational Brownian motion of the enzyme. The enzyme is modeled as a hard sphere with radius a, moving in liquid water with viscosity η, so that the Stokes-Einstein equations may be used to estimate D_t and the rotational diffusion coefficient D_r:

$$D_t = \frac{kT}{6\pi\eta a}, \quad D_r = \frac{kT}{8\pi\eta a^3}. \tag{2}$$

Analytical solutions of the overdamped Langevin equation for self-propelled particles have been developed by ten Hagen et al. (14) under the assumption that the Stokes-Einstein equations hold and that rotational and diffusional translation are not coupled to each other. In EED experiments, the diffusion coefficient is measured over times much greater than the enzyme's turnover time, which is in turn usually much greater than the rotational relaxation time of the enzyme, $\tau = (2D_r)^{-1} \in [10^{-9}\ \text{s}, 10^{-6}\ \text{s}]$ (15). In this setting, ten Hagen et al.'s Eq. 34 applies and yields the mean-square displacement as a function of time:

$$\langle \Delta \mathbf{r}^2 \rangle = 6D_t t + \left(\frac{4}{3}\frac{a^2 F}{kT}\right)^2 \left(\frac{t_p}{t_c}\right) D_r t, \tag{3}$$

where the first term gives the mean-square displacement in the absence of propulsion and the second term captures the effect of propulsion. We have inserted the term t_p/t_c to account for the fact that self-propulsion acts to raise the diffusion constant only during this fraction of the time (see Appendix). Recognizing that $D_{app} = \langle \Delta \mathbf{r}^2 \rangle / (6t)$, using Eq. 2, and employing Stokes' law, $F = 6\pi\eta a v$, to replace force with velocity, one may rewrite Eq. 3 as

$$D_{app} = D_t + \frac{v^2}{6D_r}\frac{t_p}{t_c}. \tag{4}$$

The first term is the contribution of normal Brownian motion, and the second term is the contribution from self-propulsion. The enhancement ratio, R, then is

$$R = \frac{v^2}{6D_t D_r}\frac{t_p}{t_c}. \tag{5}$$

Thus, the propulsion speed required to achieve a given level of diffusion enhancement R is given by

$$v = \frac{kT}{\pi\eta a^2}\left(\frac{R}{8}\frac{t_c}{t_p}\right)^{\frac{1}{2}}. \tag{6}$$

To determine the power required for a self-propelled particle to achieve observed levels of enhanced diffusion, it is necessary to address the energetic efficiency of the self-propulsion mechanism. Rather than make any mechanistic assumptions here, we make the most conservative assumption—i.e., the one requiring least power—by using the minimal energy dissipation theorem. This says that, at low Reynolds number, no propulsion mechanism is more efficient than dragging the particle by external force in a Stokes flow (16–20). Accordingly, we consider the power to drag an enzyme molecule in a Stokes flow at the propulsion speed required to generate enhanced diffusion with a specific value of R. Inserting v from Eq. 6 into Stokes' law, $F = 6\pi\eta a v$, we obtain the required power averaged over the full catalytic cycle:

$$P_{req} = 6\pi\eta a v^2 \frac{t_p}{t_c} = \frac{3}{4}\frac{R(kT)^2}{\pi\eta a^3}. \tag{7}$$

RESULTS

We first apply Eq. 6 to estimate the propulsion speeds needed to account for experimentally observed diffusion enhancements. The minimal thrust speed, v_{min}, that would explain the diffusion enhancement is obtained by setting $t_p = t_c$ because larger speeds are required when $t_p < t_c$. Given T = 298 K, the viscosity of liquid water, and a typical enzyme diffusion enhancement of $R = 0.2$ (5), one obtains $v_{min}(\text{m/s}) = 0.21a^{-2}$ (a in nm). This quantity depends only on the radius of the enzyme. For catalase, $a = 5.3$ nm (21), so the minimal propulsion speed $v_{min} = 7 \times 10^{-3}$ m/s. Similar values of v_{min} are obtained for the other enzymes that showed EED in experiments because their radii are similar to that of catalase (Table 1). These speeds, which amount to $\sim 10^6$ enzyme radii per second, are strikingly high. Furthermore, we anticipate that any thrust generated by enzymatic catalysis will persist only for a small fraction of the enzymologic turnover time; i.e., in all likelihood, $t_p \ll t_c$. As a consequence, based on Eq. 6, even higher propulsion speeds would be needed during the short t_p intervals to explain observed values of R.

Although implausibly high propulsion speeds would be needed to account for EED by self-propulsion, this analysis remains consistent with the observation that larger particles, e.g., Janus particles, can achieve substantial enhancements of diffusion via self-propulsion (10). This is because, for larger particles, a given propulsion velocity leads to higher

TABLE 1 Turnover Rates, Experimental Hydrodynamic Radii, Minimal Thrust Speeds, and Required Reaction Free Energies of Enzymes Reported to Show EED

Enzyme	Turnover rate/s^{-1}	Radius/ nm	v_{min}/ $m \cdot s^{-1}$[a]	$-\Delta G^*_{req}$/ $kJ \cdot mol^{-1}$[b]
T4 DNA polymerase	0.5 (47)[c]	4.6 (48)	1×10^{-2}	1×10^7
Aldolase	5 (8)[d]	4.9 (21)	9×10^{-3}	8×10^5
T7 RNA polymerase	4 (49)[c]	8.4 (50)	3×10^{-3}	2×10^5
Hexokinase	300 (7)[d]	6.3 (51)	5×10^{-3}	6000
ATP synthase	1000 (52)[c]	6.6 (1)	5×10^{-3}	2000
Alkaline phosphatase	3000 (5)[d]	7.7 (53)	3×10^{-3}	400
Catalase	10,000 (5)[d]	5.3 (21)	7×10^{-3}	300
Urease	10,000 (5)[d]	7.0 (29)	4×10^{-3}	100
Acetylcholinesterase	20,000 (54)[c]	8.8 (32)	3×10^{-3}	40

For multimeric enzymes, turnover rate is of the whole multimer. Citations are parenthesized.
[a]For $R = 20\%$ diffusion enhancement, using Eq. 6 with $t_c = t_p$.
[b]For $R = 20\%$ diffusion enhancement, using Eq. 7 divided by turnover rate.
[c]Turnover rates of these enzymes were not reported in the publications of their EED measurements, so we instead use the turnover rate when substrate concentration equals K_m, i.e, monomer k_{cat} times number of catalytic sites times 0.5.
[d]Turnover rate at $R = 20\%$, read from the corresponding publication of EED measurements.

values of R, mainly through the dependence of D_r on size. Intuitively, the longer the rotational correlation time, the greater the effect of propulsion on the root mean-square displacement. Thus, self-propulsion is much more effective at enhancing the diffusion of large particles than that of small particles, such as enzymes.

We now turn to the power required to explain EED and its relation to the chemical energy available from catalysis, which is approximated by the standard free energy of reaction, as explained in the third part of the Appendix. For catalase, with $a = 5.3$ nm, the result is $P_{req} = 3 \times 10^6$ kJ $\cdot$ s^{-1} $\cdot$ mol^{-1}. The turnover rate of catalase is about 10^4 s^{-1} under conditions that yield an $R = 20\%$ (5), so this power requirement corresponds to a minimal required reaction free energy of $\Delta G^{\circ}_{req} = -300$ kJ$\cdot$mol^{-1}. This is well above the standard free energy of reaction, $\Delta G^{\circ} = -95$ kJ $\cdot$ mol^{-1}, computed from the standard free energies of formation of the reactant and products (22). For enzymes with lower turnover rates, the required reaction free energies range up to 5×10^6 kJ $\cdot$ mol^{-1} (Table 1). These required reaction free energies are far larger than what is available from the free energy of the chemical reactions catalyzed by the enzymes. For example, for alkaline phosphatase, $\Delta G^{\circ} = -8.5$ kJ $\cdot$ mol^{-1}; for urease, $\Delta G^{\circ} = -20$ kJ $\cdot$ mol^{-1}; and for acetylcholinesterase, $\Delta G^{\circ} = -17$ kJ $\cdot$ mol^{-1} (23). The magnitudes of the reaction free energies in Table 1 may be put into perspective by considering that the standard free energy of hydrolysis of ATP, the cell's energy currency, is only about -32 kJ $\cdot$ mol^{-1} (24). Furthermore, as detailed in the Discussion, the power requirements derived here are conservative, and the actual power requirements probably exceed what is available by an even larger margin. Thus, it is unlikely that experimental observations of EED can be accounted for by catalysis-driven self-propulsion.

DISCUSSION

We now critically examine the approximations and assumptions used in this theory and consider the results in light of recent relevant experimental studies.

Three key assumptions in this analysis are conservative, in the sense of lowering the estimate of the power required to achieve a certain level of enhanced diffusion. First, we used the minimal energy dissipation theorem, based on the assumption of a Stokes flow around the enzyme, to estimate the minimal power required for a given propulsion velocity. Any real propulsion likely generates a non-Stokes flow field around the enzyme, resulting in higher viscous dissipation integrated over whole space than in the ideal Stokes flow and hence lower efficiency than assumed here. (Intuitively, if one replaces the enzyme by a bacterium, we computed the dissipation associated with pulling it through the water with an optical trap, rather than the greater dissipation associated with its using flagellae to swim at the same speed.) Indeed, a bacterial propulsion mechanism, which, unlike a nonmotor enzyme, has been optimized during evolution, was found to have only ~1% of the maximal propulsion efficiency associated with pure Stokes drag (18). Additionally, a propulsion mechanism might rely on local chemical gradients, imposing an additional entropy production term as the chemical gradients spontaneously dissipate. Thus, although we have used the maximal efficiency assumption, the true efficiency of any enzyme propulsion mechanism is probably orders of magnitude lower. This makes it even less probable that the required power could be provided by the available chemical energy.

Second, we assumed that the propulsion mechanism increases the apparent translational diffusion coefficient without increasing the enzyme's rotational diffusion coefficient, D_r. We are not aware of any experiments that report on the rotational diffusion rates of enzymes undergoing translational EED, but any translational propulsion mechanism would probably also increase the rate of rotational diffusion. This is because there is no reason to expect that a propulsive force will not also exert a torque and thus drive rotation. In fact, the rotational diffusion coefficient of 30 nm Pt-Au Janus particles increases by up to 70% when they are catalytically active and undergoing enhanced translational diffusion (10). This is relevant here because, as is evident from Eq. 4, increasing D_r would further increase the velocity v needed to achieve a given level of D_{app}. Therefore, even more power would be required, again making EED harder to explain on the basis of catalysis-driven self-propulsion. In addition, any chemical energy expended in driving rotational motion would become unavailable to drive translational motion.

Third, we assumed that the thrust speed v is constant during the interval t_p. Using Fourier analysis, we show in the Appendix that allowing v to be time-varying during t_p could only increase the thrust speed and power required to achieve a given degree of diffusion enhancement. The Appendix furthermore proves a more implicit but intuitive assumption we have used, that diffusion enhancement and power requirements scale linearly with the duty ratio t_p/t_c. This is found to hold as long as any significant high-frequency components in the thrust velocity are slow relative to the rotational diffusion time of the enzyme. Intuitively, if a high-frequency component of the thrust speed reverses direction before the enzyme has had time to reorient, the motion due to this component can be canceled in the lab frame, leading to a minimal contribution to the net translational displacement. In contrast, if the enzyme has time to rotate before the thrust component reverses, the reversed component will act in a different direction in the lab frame, leading to less cancellation and more net displacement.

Several other assumptions also deserve comment. Our use of the Stokes-Einstein equations with stick boundary conditions is justified by several considerations. First, changing to slip boundary conditions would merely replace the factor of

1/6 in the Stokes-Einstein equation by a factor of 1/4, which would not change our conclusions. Additionally, simulations of spherical macromolecule-sized particles in solution yield translational diffusion coefficients that are bracketed by the results of the Stokes-Einstein equation computed with stick-and-slip boundary conditions, using the geometric radii of gyration of the solutes (25). And if one assumes stick boundary conditions in mapping from measured translational diffusion coefficients of proteins in water to effective radii and then from radii to the predicted rotational diffusion coefficient, the results agree with experiments to within ~50% (21). Interestingly, the actual rotational diffusion constants tend to be higher, rather than lower, than those predicted by Eq. 2 (26). Correcting in this direction would only strengthen our conclusions because increasing D_r means that even more power is required for a given value of R. Finally, treating the enzymes for which enhanced diffusion has been observed as spherical is reasonable for these globular proteins; highly nonspherical, (e.g., rod-like) proteins may deserve further analysis.

Additionally, we have treated each enzyme molecule's motion as independent of the motions of the other enzymes in solution. We tested this assumption by applying the hydrodynamic interaction model of Mikhailov and Kapral (27) to the case of enzymes at the very low concentrations, ~10 nM, used in typical EED measurements. The resulting hydrodynamic interactions are found to be negligibly small.

It is worth considering this analysis in the context of recent, high-resolution experimental studies of EED. In two elegant studies, Jee et al. combined stimulated emission depletion microscopy with fluorescence correlation spectroscopy to study enzyme diffusion at very high spatial resolution (9,28). Intriguingly, when urease in the presence of urea was studied with a small beam waist (50–250 nm), a fast component of translational motion was revealed. The authors interpreted the fast component as being the result of propulsive motion powered by the urease reaction and argued that this self-propulsion could explain enhanced diffusion of urease. Perhaps the chief reason for the difference in their conclusion relative to ours is that their rotational diffusion time of 2.9–5.6 μs corresponds to a hydrodynamic radius $a = 10$–12 nm, which is considerably larger than the value of 7.0 nm reported in a prior experimental study (29) and used here. The smaller hydrodynamic radius used here is further supported by our analysis of the hexameric biological unit of urease (30) with the program HYDROPRO (31), which yields translational and rotational diffusion coefficients corresponding to hydrodynamic radii of 6.6 and 6.7 nm, respectively. In addition, one may infer the hydrodynamic radius of urease from the baseline translational diffusion coefficient of 29 $\mu m^2/s$ reported by Jee et al; the result is 7.5 nm, which is close to the value we used. Based on Eq. 7, going from a radius of 10–12 nm to the more plausible value of 7 nm used here leads to a three- to fivefold increase in the power requirement for a given degree of diffusion enhancement R. Given that Jee and co-workers' estimated value of the free energy required for each catalytic cycle, 25 kJ $\cdot$ mol^{-1} (28), is already slightly higher than the reaction free energy of this enzyme, 20 kJ $\cdot$ mol^{-1} (23), an upward adjustment based on this consideration makes it difficult to support the hypothesis that catalytic self-propulsion explains EED in urease. Interestingly, their reported rotational diffusion time of 44–46 μs for the enzyme acetylcholinesterase corresponds to a hydrodynamic radius of ~25 nm, which is about three times the experimentally determined hydrodynamic radius of the largest globular form of this enzyme (32,33). One may speculate that the abnormally low rotational reorientation rates inferred by Jee et al. could reflect extrinsic perturbations of the enzymes, such as fluid flows, varying on a timescale of about 10 μs.

CONCLUSIONS

Our analysis shows that the propulsion speeds required to explain experimentally observed levels of EED by the mechanism of catalytic self-propulsion are implausibly large. More fundamentally, the power levels needed to account for observed levels of diffusion enhancement by catalytic self-propulsion are greater than those available from enzyme-catalyzed chemical reactions. For most enzymes, the power requirement is orders of magnitude too great, and even for the faster enzymes, the power required is still considerably larger than that afforded by the reaction. Moreover, the power actually required to generate observed levels of diffusion enhancement is probably greater than our estimates because we have used conservative approximations that lead to lower estimates of the required power. However, because the power required for a given level of diffusion enhancement decreases sharply with increasing particle size, our results remain consistent with experimental observations that self-propulsion of micron-scale particles with surfaces coated with a metallic catalyst (10) or with immobilized enzymes (34) leads to significantly enhanced translational diffusion. The propulsion direction of larger particles randomizes more slowly, so the contribution of propulsion to translational diffusion is increased. We conclude that enhanced diffusion of enzymes cannot easily be explained by self-propulsion powered by the chemical energy of the catalyzed reactions.

It is of interest to consider other explanations for EED. One possibility is an increase in normal, thermally driven translational diffusion. This could result from a decrease of the mean hydrodynamic radius of the enzyme in the course of the catalytic cycle, as recently noted (8,35). Alternatively, it has been proposed (36) that the catalytic cycle might raise the temperature of nearby solvent enough to increase the enzyme's diffusion constant, through η and T in Eq. 2. However, the viability of this explanation appears

to rely on use of the thermal conductivity of air rather than water (36) because the effect becomes negligible when the thermal conductivity of water is used. Global heating of the solution due to release of chemical energy is also insufficient to explain observed diffusion enhancement (5,37). It is worth noting, too, that exothermicity, and even chemical catalysis itself, is not required for at least some reported instances of EED (2,8).

Thus, the mechanisms of EED remain obscure. Further experimental studies may help solve this puzzle. It has been suggested (13,38) that fluorescence correlation spectroscopy measurements may be subject to experimental artifacts, such as subunit dissociation and fluorophore quenching, so that further controls, such as those employed by Jee and co-workers (9,28), are of high value. Because the turnover rate is needed to convert the power requirement (Eq. 7) to $\Delta G^{\ddagger}_{\text{req}}$ and compare with standard thermodynamic data, it would also be helpful if the turnover rate of the enzymes being studied could be measured under the precise conditions of each diffusion study to avoid uncertainties that may result from literature data measured under different conditions and from reliance on an assumption of Michaelis-Menten kinetics. Alternative technologies for measuring diffusion enhancement may also provide different perspectives. For example, although fluorescence correlation spectroscopy studies of aldolase demonstrated EED (7,8), aldolase did not show enhanced diffusion when studied by dynamic light scattering (39) or by NMR (40). On the other hand, an electrochemical experiment has provided supporting evidence of catalase EED (41). Intriguingly, a study in which enzyme molecules were confined to an ~2D region to enable single-molecule tracking showed strong enhanced diffusion, though we note that interpretation of these data is complicated by the fact that the baseline diffusion coefficients were markedly reduced relative to their 3D values (42). Further direct tracking studies (43) could be useful both to confirm the phenomenon of EED and to provide details that might bear on mechanism.

APPENDIX: FOURIER ANALYSIS OF TIME-VARYING THRUST—GENERAL ANALYSIS

The derivation in the main text treats the self-propulsion thrust as constant during an interval t_p within each catalytic cycle of duration $t_c \geq t_p$. Here, we examine the consequences of a more general time-varying thrust. We make the reasonable assumption that the time over which the translational diffusion constant is measured, t_m, is much larger than the duration of the catalytic cycle, t_c (milliseconds to seconds), which in turn is much larger than the rotational relaxation time $\tau \equiv (2D_r)^{-1}$ of the enzyme (nanoseconds to microseconds). We address the effect of time-varying propulsion on translational diffusion by expanding the propulsion speed in a Fourier series, as previously done by Lauga in the context of reciprocal swimming (44), and extend the analysis to determine how time-variation affects the efficiency with which propulsive power generates enhanced diffusion.

Consider an enzyme with a time-dependent, self-propulsion speed $v(t)$, whose translational diffusion is evaluated from time $t = 0$ until the end of some experimental time, t_m. As in the main text, the direction of the propulsion is fixed in the enzyme's frame of reference and therefore reorients in the lab frame of reference because of rotational diffusion of the enzyme. After periodic extension, $v(t)$ can be expanded into a Fourier series:

$$v(t) = v_0\left[\frac{c_0}{2} + \sum_{n=1}^{\infty} c_n \cos\frac{2n\pi}{t_m}t + \sum_{n=1}^{\infty} d_n \sin\frac{2n\pi}{t_m}t\right]. \quad (8)$$

This time-varying propulsion speed generates an increment in the translational diffusion coefficient given by Lauga's Eq. 7 (44),

$$\Delta D = \frac{1}{3}\lim_{t_m\to\infty}\left\langle \int_0^{t_m} v(t)v(t')e^{-\frac{t-t'}{\tau}}dt' \right\rangle_t, \quad (9)$$

where we have inserted missing angle brackets, indicating an ensemble average over reference time t in the integral. This expression yields a well-defined result because $t_m \gg \tau$. Substitution of the Fourier series into this expression yields

$$\Delta D = \frac{v_0^2\tau}{3}\left[\frac{c_0^2}{4} + \frac{1}{2}\sum_{n=1}^{\infty}\frac{c_n^2}{1+\left(\frac{2n\pi\tau}{t_m}\right)^2} + \frac{1}{2}\sum_{n=1}^{\infty}\frac{d_n^2}{1+\left(\frac{2n\pi\tau}{t_m}\right)^2}\right]. \quad (10)$$

This equation decomposes the diffusion enhancement into contributions from each Fourier component. The mean power consumption, $\langle P\rangle = 6\pi\eta a\langle v(t)^2\rangle$, may similarly be decomposed into contributions from each frequency component,

$$\begin{aligned}\langle P\rangle &= 6\pi\eta a v_0^2\left[\frac{c_0^2}{4} + \sum_{n=1}^{\infty} c_n^2\left\langle \cos^2\frac{2n\pi}{t_m}t\right\rangle + \sum_{n=1}^{\infty} d_n^2\left\langle \sin^2\frac{2n\pi}{t_m}t\right\rangle\right]. \quad (11)\\ &= 6\pi\eta a v_0^2\left[\frac{c_0^2}{4} + \frac{1}{2}\sum_{n=1}^{\infty} c_n^2 + \frac{1}{2}\sum_{n=1}^{\infty} d_n^2\right]\end{aligned}$$

Here, we have used the orthogonality of the Fourier components to eliminate cross terms and have made the substitutions $\langle \cos^2(2n\pi/T)t\rangle = \langle \sin^2(2n\pi/t_m)t\rangle = 1/2$.

Comparing Eqs. 10 and 11 reveals that, given a set of amplitudes c_0, c_1, ..., c_n, d_1, d_2, ..., d_n, higher-frequency components (i.e., ones with larger subscripts) generate smaller contributions to the diffusion coefficient but equal contributions to the power consumption. The efficiency of diffusion enhancement, normalized to that for constant propulsion, is given by Eqs. 10 and 11 as $\xi \equiv (18\pi\eta a/\tau)(\Delta D/\langle P\rangle)$. It is apparent from our analysis that the efficiency is greatest when only the constant thrust component, c_0, is nonzero; i.e., when the thrust speed is constant during the enzyme's catalytic cycle, as assumed when considering the minimal thrust speed in the main text. Any variation in thrust over time can only reduce ξ to below one. Thus, "scheduling" the thrust cannot decrease the power needed

for a given level of diffusion enhancement to below the power needed for constant thrust.

FOURIER ANALYSIS OF TIME-VARYING THRUST—SQUARE-WAVE CASE

In the main text, we assumed a square-wave thrust schedule, with constant nonzero thrust during $t_p < t_c$ and zero thrust during the rest of t_c. We argued that the diffusion enhancement and the minimal power dissipation both scale linearly with the duty ratio t_p/t_c. For diffusion enhancement, it should be apparent that this holds because the ensemble average in Eq. 9 is proportional to the portion of time when $v(t)$ is nonzero. Nonetheless, it is of interest to confirm these arguments numerically within the Fourier analysis. To do this, we consider the speed to be $v(t) = v_0$ when $t \in (-(t_p/2), (t_p/2))$, and $v(t) = 0$ elsewhere in $t \in (-(t_c/2), (t_c/2))$. The corresponding Fourier series is

$$v(t) = v_0\left[\frac{c_0}{2} + \sum_{n=1}^{\infty} c_n cos\frac{2n\pi}{t_c}t\right], \tag{12}$$

$$c_0 = 2\frac{t_p}{t_c}, \tag{13}$$

$$c_n = \frac{2}{t_c}\int_{-\frac{t_c}{2}}^{\frac{t_c}{2}} v(t) cos\frac{2n\pi}{t_c}tdt = \frac{2}{n\pi}sin\left(n\pi\frac{t_p}{t_c}\right). \tag{14}$$

Inserting these expressions into Eq. 10, with $\tau/t_c = 0.01$, which corresponds to the case of urease, yields the expected linear variation of ΔD with t_p, as shown in Fig. 1.

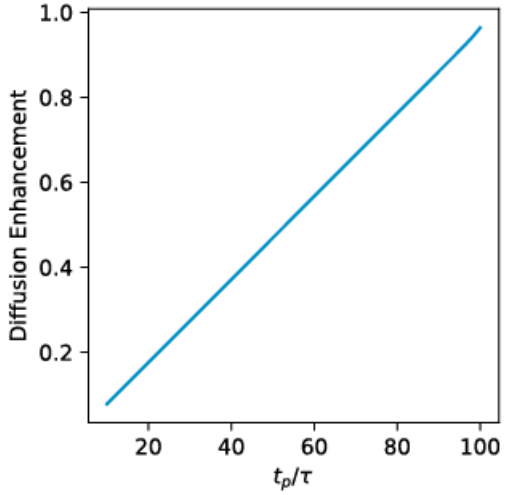

FIGURE 1 Numerical evaluation of the diffusion enhancement given by Eq. 10 confirms the linear relationship between the ratio t_p/τ and the diffusion enhancement for a square-wave thrust schedule. The enhancement is plotted relative to the case $t_p = t_c$. The value of t_c and τ correspond to urease from Table 1. To see this figure in color, go online.

We next examine the efficiency, ξ, for this square-wave thrust schedule:

$$\xi = \frac{\frac{c_0^2}{4} + \frac{1}{2}\sum_{n=1}^{\infty}\frac{c_n^2}{1+(2n\pi\tau/t_c)^2}}{\frac{c_0^2}{4} + \frac{1}{2}\sum_{n=1}^{\infty} c_n^2}. \tag{15}$$

For constant speed with $t_p = t_c$, this yields $\xi = 1$. The loss in efficiency when $t_p < t_c$ then is given by

$$1-\xi = \frac{\frac{1}{2}\sum_{n=1}^{\infty} c_n^2\frac{4n^2\pi^2\tau^2}{t_c^2+4n^2\pi^2\tau^2}}{\frac{c_0^2}{4} + \frac{1}{2}\sum_{n=1}^{\infty} c_n^2} = \frac{\frac{1}{2}\sum_{n=1}^{\infty} c_n^2\frac{4n^2\pi^2\tau^2}{t_c^2+4n^2\pi^2\tau^2}}{t_p/t_c}$$

$$= 6\frac{\tau}{t_c}\frac{\tau}{t_p}\sum_{n=1}^{\infty}\frac{\sin^2\left[n\pi\frac{\tau}{t_c}\left(\frac{\tau}{t_p}\right)^{-1}\right]}{1+4\pi^2\left(\frac{\tau}{t_c}\right)^2 n^2},$$

where we have used Parseval's theorem to evaluate the denominator and then inserted Eq. 14. For given values of τ and t_c, the maximal drop in efficiency is expected to happen when t_p is much smaller than t_c because this increases the weight of the high-frequency components of the thrust velocity. Focusing, then, on this low-efficiency limit, we can approximate the summation with an integral and then evaluate the integral using the residual theorem:

$$\lim_{\frac{t_p}{t_c}\to 0}(1-\xi) = \lim_{\frac{t_p}{t_c}\to 0} 16\frac{\tau}{t_c}\frac{\tau}{t_p}\left(-1 + \sum_{n=0}^{\infty}\frac{sin^2\left(\pi\frac{t_p}{t_c}n\right)}{1+4\pi^2\left(\frac{\tau}{t_c}\right)^2 n^2}\Delta n\right)$$

$$= 16\frac{\tau}{t_c}\frac{\tau}{t_p}\left(-1 + \frac{t_c}{t_p}\int_0^{\infty}\frac{sin^2(\pi x)}{1+4\pi^2\left(\frac{\tau}{t_p}\right)^2 x^2}dx\right)$$

$$= 16\frac{\tau}{t_c}\frac{\tau}{t_p}\left(-1 + \frac{t_c}{4t_p}\left(\int_{-\infty}^{\infty}\frac{dx}{1+4\pi^2\frac{\tau^2}{t_p^2}x^2} - \int_{-\infty}^{\infty}\frac{cos(2\pi x)}{1+4\pi^2\frac{\tau^2}{t_p^2}x^2}dx\right)\right)$$

$$= 16\frac{\tau}{t_c}\frac{\tau}{t_p}\left(-1 + \frac{t_c}{8\tau} - \frac{t_c}{8\tau}e^{-t_p/\tau}\right) \approx 2\frac{\tau}{t_p}.$$

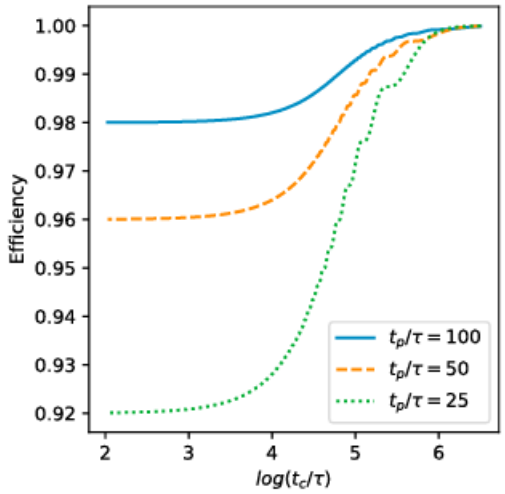

FIGURE 2 Relationship between efficiency ξ, and the ratio t_c/τ, plotted for three values of t_p/τ. To see this figure in color, go online.

Consequently, because $t_p \gg \tau$ is expected for most enzymes, the efficiency will remain near unity, even under the extreme assumption that $t_p \ll t_c$. This result supports our approximation in the main text that the diffusion enhancement caused by a square-wave thrust schedule is proportional to t_p. It also shows that our assumptions are conservative because not invoking this approximation would decrease the efficiency and further increase the power requirement. The analytical result in Eq. 16 analysis of the time-varying thrust square-wave case is elaborated by numerical calculations of the efficiency ξ, as drawn in Fig. 2. Here, t_c/τ spans the range of this ratio found for the enzymes in Table 1, from 100 for urease to 3×10^6 for DNA polymerase. Three values for t_p/τ are used, subject to the requirement that $t_p < t_c$.

The near proportionality of both ΔD and $\langle P \rangle$ to t_p may be understood more intuitively by reference to Eqs. 10 and 11. Because the denominator in Eq. 10, $1 + (n\tau\pi/2t_c)^2$, is near unity except for very large n, low-frequency components deviate only very slightly from the zeroth component in efficiency. On the other hand, high-frequency components with large n have negligible amplitudes because $c_n \leq (2/n\pi)$, so they do not alter efficiency either. Therefore, the diffusion enhancement and the power requirement both scale near-linearly with duty ratio t_p/t_c, leading to near-uniform efficiency. It is of interest to note, however, that efficiency would fall if there were significant oscillations in $v(t)$ on the timescale of τ or smaller. In this regime, the nonzero velocity components reverse direction before the enzyme has had time to rotate, so there is little net displacement due to the thrust. In contrast, when the nonzero velocity components do not reverse until the enzyme has had time to rotate, the net effect of the time-varying thrust is to generate randomly directed displacements, which contribute to the apparent diffusion constant.

FREE ENERGY FOR SELF-PROPULSION AVAILABLE FROM AN ENZYME-CATALYZED CHEMICAL REACTION

In the main text, we took the standard free energy of the reaction, ΔG°, to be the free energy from an enzyme-catalyzed chemical reaction that is available to power the enzyme's self-propulsion. A concern with this approach may be that, when the two sides of the chemical reaction have different numbers of solute molecules, ΔG° depends on the arbitrary standard concentration, C°, and the available free energy ought not depend on an arbitrary quantity. Here, we show that the standard reaction free energy is, in fact, a good approximation to the free energy available from the combined processes of substrate-enzyme binding, chemical reaction, and product release, so long as the standard concentration is set to its customary value of 1 mol/L. This section thus justifies the use of the standard concentration in the main text while also offering insight into how more refined estimates of the available free energy might be made.

First, it is instructive to consider whether it would be appropriate to take the free energy available for propulsion to be the free energy of reaction under the experimental conditions at which enzyme diffusion was studied; i.e., $\Delta G = \Delta G^\circ + RT\ln Q$, where Q is the experimental concentration quotient, assuming activity coefficients near unity. This approach is problematic because it would require a physical mechanism that could couple the macroscopic concentrations of substrate and product to the local events at a single enzyme molecule. Instead, if one considers the entire catalytic process, from enzyme-substrate encounter through release of product to the bulk, the only steps that could contribute free energy to enzyme propulsion

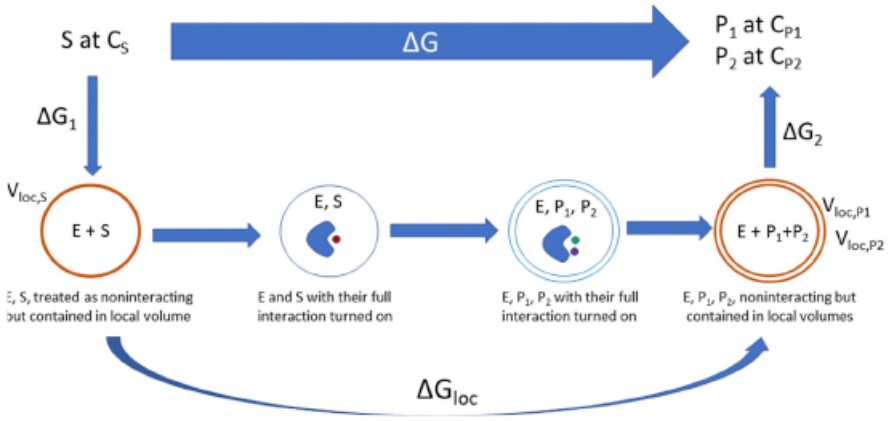

FIGURE 3 Definition of ΔG_{loc} via a schematized thermodynamic cycle. See text for details.

are those in which the enzyme interacts significantly with the substrate or product. Such interactions occur only when the substrate or product molecules are near the enzyme, so the free energy available for propulsion may be termed the local free energy, ΔG_{loc}.

The local free energy may be estimated with the thermodynamic cycle shown in Fig. 3, which illustrates a case in which one substrate molecule, S, present at concentration C_S, is converted to two product molecules, P1 and P2, present at concentrations C_{P1} and C_{P2}, respectively, with a free energy of reaction under experimental conditions of $\Delta G = \Delta G^{\circ} + RT\ln(C_{P1}C_{P2}/C_S C^{\circ})$, assuming ideal solutions. The lower route of the cycle breaks the process into five steps. In the first, step, the substrate molecule is, in effect, compressed into the region near the enzyme where enzyme-substrate interactions are non-negligible, under the artificial assumption that only steric interactions exist between the two molecules. The volume of this region is termed $V_{\mathrm{loc,S}}$, and the free-energy change associated with this step is $\Delta G_1 = -RT\ln(V_{\mathrm{loc,S}}C_S)$. The subsequent three steps are those for which the free-energy change, ΔG_{loc}, could contribute free energy to propulsion. Here, the nonsteric enzyme-substrate interactions are turned on, the substrate is converted to product, and then all nonsteric interactions between the enzyme and products are artificially turned off while the products are constrained to remain in the region where these interactions were non-negligible. For products P1 and P2, the volumes of these local regions are, respectively, $V_{\mathrm{loc,P1}}$ and $V_{\mathrm{loc,P2}}$. Finally, the constrained products are released to their solute concentrations with free-energy change $\Delta G_2 = RT\ln(V_{\mathrm{loc,P1}}V_{\mathrm{loc,P2}}C_{P1}C_{P2})$. Closing the thermodynamic cycle now allows one to show that $\Delta G_{\mathrm{loc}} = \Delta G^{*} - RT\ln((V_{\mathrm{loc,P1}}V_{\mathrm{loc,P2}}C^{*})/V_{\mathrm{loc,S}})$. Note that if V_{loc} is given in units of nm^3, then the 1 mol/L standard concentration should be written as 0.6 molecules/nm^3. The steps corresponding to the local free energy have the character of a unimolecular process, and this quantity is, accordingly, independent of the standard concentration, C°, because any change in C° causes equal and opposite changes in ΔG° and the second term of ΔG_{loc}. If the enzyme interacts with substrate and product molecules over similar ranges, we may write all three local volumes as the same quantity V_{loc}, and the local free energy takes the simpler form $\Delta G_{\mathrm{loc}} = \Delta G^{\circ} - RT\ln(V_{\mathrm{loc}}C^{\circ})$. A straightforward generalization to other stoichiometries yields $\Delta G_{\mathrm{loc}} = \Delta G^{\circ} - (N_P - N_S)RT\ln(V_{\mathrm{loc}}C^{\circ})$, where N_P and N_S are the numbers of product and substrate solutes, respectively. Again, although ΔG° depends on the standard concentration, this dependency is canceled by the factors of C° in the added term.

The quantity V_{loc} is the volume covered by the interaction range of substrate and product molecules with the enzyme. We estimate this quantity by considering the interaction region to be a hemisphere around the enzyme active site with a 1 nm radius typical of protein-ligand interaction ranges, as determined from molecular dynamics simulations (45,46). With these assumptions, $RT\ln(V_{\mathrm{loc}}C^{\circ}) = 0.6\ \mathrm{kJ \cdot mol^{-1}}$. Note that this quantity is rather insensitive to the precise choice of V_{loc} because of the logarithm. For urease, where one molecule of urea is decomposed into one carbon dioxide and two ammonia molecules, $N_P - N_S = 2$, so $\Delta G_{\mathrm{loc}} = \Delta G^{\circ} - 2RT\ln(V_{\mathrm{loc}}C^{\circ}) = -20\ \mathrm{kJ \cdot mol^{-1}} - 1.2\ \mathrm{kJ \cdot mol^{-1}} = -21.2\ \mathrm{kJ \cdot mol^{-1}}$. For acetylcholinesterase, where one molecule of acetylcholine is decomposed into one acetic acid and one choline, $N_P - N_S = 1$, so $\Delta G_{\mathrm{loc}} = \Delta G^{\circ} - RT\ln(V_{\mathrm{loc}}C^{\circ}) = -17\ \mathrm{kJ \cdot mol^{-1}} - 0.6\ \mathrm{kJ \cdot mol^{-1}} = -17.6\ \mathrm{kJ \cdot mol^{-1}}$. Thus, the local free energies available to drive propulsion remain close to the standard binding free energies appropriate to $C^{\circ} = 1$ mol/L, as was to be demonstrated. We note that this result is serendipitous because changing to a different standard concentration would not change ΔG_{loc} but would change ΔG°.

AUTHOR CONTRIBUTIONS

M.F. contributed to conceptualization, investigation, methodology, and review and editing and led in formal analysis, writing of the original draft, and software development. M.K.G. contributed to conceptualization, investigation, and review and editing; supported the methodology and formal analysis; and led in funding acquisition and supervision.

ACKNOWLEDGMENTS

We thank Dr. Hari Muddana for helpful discussions.

M.K.G. has an equity interest in, and is a cofounder and scientific advisor of, VeraChem LLC.

REFERENCES

1. Börsch, M., P. Turina, ..., P. Gräber. 1998. Conformational changes of the H+-ATPase from Escherichia coli upon nucleotide binding detected by single molecule fluorescence. *FEBS Lett.* 437:251–254.
2. Yu, H., K. Jo, ..., D. C. Schwartz. 2009. Molecular propulsion: chemical sensing and chemotaxis of DNA driven by RNA polymerase. *J. Am. Chem. Soc.* 131:5722–5723.
3. Sengupta, S., M. M. Spiering, ..., A. Sen. 2014. DNA polymerase as a molecular motor and pump. *ACS Nano.* 8:2410–2418.
4. Sengupta, S., K. K. Dey, ..., A. Sen. 2013. Enzyme molecules as nanomotors. *J. Am. Chem. Soc.* 135:1406–1414.
5. Riedel, C., R. Gabizon, ..., C. Bustamante. 2015. The heat released during catalytic turnover enhances the diffusion of an enzyme. *Nature.* 517:227–230.
6. Muddana, H. S., S. Sengupta, ..., P. J. Butler. 2010. Substrate catalysis enhances single-enzyme diffusion. *J. Am. Chem. Soc.* 132:2110–2111.
7. Zhao, X., H. Palacci, ..., A. Sen. 2018. Substrate-driven chemotactic assembly in an enzyme cascade. *Nat. Chem.* 10:311–317, Published online December 18, 2017.
8. Illien, P., X. Zhao, ..., R. Golestanian. 2017. Exothermicity is not a necessary condition for enhanced diffusion of enzymes. *Nano Lett.* 17:4415–4420.
9. Jee, A. Y., S. Dutta, ..., S. Granick. 2018. Enzyme leaps fuel antichemotaxis. *Proc. Natl. Acad. Sci. USA.* 115:14–18.
10. Lee, T. C., M. Alarcón-Correa, ..., P. Fischer. 2014. Self-propelling nanomotors in the presence of strong Brownian forces. *Nano Lett.* 14:2407–2412.
11. Golestanian, R. 2010. Synthetic mechanochemical molecular swimmer. *Phys. Rev. Lett.* 105:018103.
12. Moran, J. L., and J. D. Posner. 2017. Phoretic self-propulsion. *Annu. Rev. Fluid Mech.* 49:511–540.
13. Bai, X., and P. G. Wolynes. 2015. On the hydrodynamics of swimming enzymes. *J. Chem. Phys.* 143:165101.
14. ten Hagen, B., S. van Teeffelen, and H. Löwen. 2011. Brownian motion of a self-propelled particle. *J. Phys. Condens. Matter.* 23:194119.
15. Loman, A., I. Gregor, ..., J. Enderlein. 2010. Measuring rotational diffusion of macromolecules by fluorescence correlation spectroscopy. *Photochem. Photobiol. Sci.* 9:627–636.
16. Kim, S., and S. J. Karrila. 2005. Microhydrodynamics: Principles and Selected Applications. Courier Corporation, Chelmsford, MA.
17. Lighthill, M. J. 1952. On the squirming motion of nearly spherical deformable bodies through liquids at very small reynolds numbers. *Commun. Pure Appl. Math.* 5:109–118.
18. Stone, H. A., and A. D. Samuel. 1996. Propulsion of microorganisms by surface distortions. *Phys. Rev. Lett.* 77:4102–4104.
19. Sabass, B., and U. Seifert. 2010. Efficiency of surface-driven motion: nanoswimmers beat microswimmers. *Phys. Rev. Lett.* 105:218103.
20. Sabass, B., and U. Seifert. 2012. Dynamics and efficiency of a self-propelled, diffusiophoretic swimmer. *J. Chem. Phys.* 136:064508.
21. García De La Torre, J., M. L. Huertas, and B. Carrasco. 2000. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* 78:719–730.
22. Wagman, D. D., W. H. Evans, ..., R. L. Nuttall. 1982. The NBS tables of chemical thermodynamic properties: selected values for inorganic and C1 and C2 organic substances in SI units. *J. Phys. Chem. Ref. Data.* 11:2-37–2-38.

23. Goldberg, R. N., and Y. B. Tewari. 1994. Thermodynamics of enzyme–catalyzed reactions. Part 3. Hydrolases. *J. Phys. Chem. Ref. Data.* 23:1035–1103.

24. Alberty, R. A., and R. N. Goldberg. 1992. Standard thermodynamic formation properties for the adenosine 5′-triphosphate series. *Biochemistry.* 31:10610–10615.

25. Schmidt, J. R., and J. L. Skinner. 2003. Hydrodynamic boundary conditions, the Stokes–Einstein law, and long-time tails in the Brownian limit. *J. Chem. Phys.* 119:8062–8068.

26. Koenderink, G. H., H. Zhang, …, G. Nägele. 2003. On the validity of Stokes-Einstein-Debye relations for rotational diffusion in colloidal suspensions. *Faraday Discuss.* 123:335–354, discussion 401–421.

27. Mikhailov, A. S., and R. Kapral. 2015. Hydrodynamic collective effects of active protein machines in solution and lipid bilayers. *Proc. Natl. Acad. Sci. USA.* 112:E3639–E3644.

28. Jee, A. Y., Y. K. Cho, …, T. Tlusty. 2018. Catalytic enzymes are active matter. *Proc. Natl. Acad. Sci. USA.* 115:E10812–E10821.

29. Follmer, C., F. V. Pereira, …, C. R. Carlini. 2004. Jack bean urease (EC 3.5.1.5) aggregation monitored by dynamic and static light scattering. *Biophys. Chem.* 111:79–87.

30. Balasubramanian, A., and K. Ponnuraj. 2010. Crystal structure of the first plant urease from jack bean: 83 years of journey from its first crystal to molecular structure. *J. Mol. Biol.* 400:274–283.

31. Ortega, A., D. Amorós, and J. García de la Torre. 2011. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys. J.* 101:892–898.

32. Rieger, F., S. Bon, …, P. Benda. 1976. Torpedo marmorata acetylcholinesterase; a comparison with the Electrophorus electricus enzyme. Molecular forms, subunits, electron microscopy, immunological relationship. *Eur J. Biochem.* 68:513–521.

33. Bon, S., M. Vigny, and J. Massoulié. 1979. Asymmetric and globular forms of acetylcholinesterase in mammals and birds. *Proc. Natl. Acad. Sci. USA.* 76:2546–2550.

34. Patiño, T., X. Arqué, …, S. Sánchez. 2018. Fundamental aspects of enzyme-powered micro- and nanoswimmers. *Acc. Chem. Res.* 51:2662–2671.

35. Illien, P., T. Adeleke-Larodo, and R. Golestanian. 2017. Diffusion of an enzyme: the role of fluctuation-induced hydrodynamic coupling. *EPL.* 119:40002.

36. Golestanian, R. 2015. Enhanced diffusion of enzymes that catalyze exothermic reactions. *Phys. Rev. Lett.* 115:108102.

37. Tsekouras, K., C. Riedel, …, C. Bustamante. 2016. Comment on "Enhanced diffusion of enzymes that catalyze exothermic reactions" by R. Golestanian. *arXiv*, arXiv:1608.05433 https://arxiv.org/abs/1608.05433.

38. Günther, J. P., M. Börsch, and P. Fischer. 2018. Diffusion measurements of swimming enzymes with fluorescence correlation spectroscopy. *Acc. Chem. Res.* 51:1911–1920.

39. Zhang, Y., M. J. Armstrong, …, H. Hess. 2018. Aldolase does not show enhanced diffusion in dynamic light scattering experiments. *Nano Lett* Published online November 28, 2018. https://doi.org/10.1021/acs.nanolett.8b04240.

40. Günther, J.-P., G. Majer, and P. Fischer. 2018. Absolute diffusion measurements of active enzyme solutions by NMR. *arXiv*, arXiv:1812.08748 http://arxiv.org/abs/1812.08748.

41. Jiang, L., I. Santiago, and J. Foord. 2017. Observation of nanoimpact events of catalase on diamond ultramicroelectrodes by direct electron transfer. *Chem. Commun. (Camb.).* 53:8332–8335.

42. Xu, M., L. Valdez, …, J. L. Ross. 2018. Direct single molecule imaging of enhanced enzyme diffusion. *arXiv*, arXiv:1811.08483 http://arxiv.org/abs/1811.08483.

43. Sun, L., Y. Gao, …, C. Fan. 2017. Real-time imaging of single-molecule enzyme cascade using a DNA origami raft. *J. Am. Chem. Soc.* 139:17525–17532.

44. Lauga, E. 2011. Enhanced diffusion by reciprocal swimming. *Phys. Rev. Lett.* 106:178101.

45. Woo, H. J., and B. Roux. 2005. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. USA.* 102:6825–6830.

46. Heinzelmann, G., N. M. Henriksen, and M. K. Gilson. 2017. Attach-pull-release calculations of ligand binding and conformational changes on the first BRD4 bromodomain. *J. Chem. Theory Comput.* 13:3260–3275.

47. Burrows, J. A., and C. R. Goward. 1992. Purification and properties of DNA polymerase from Bacillus caldotenax. *Biochem. J.* 287:971–977.

48. Challberg, M. D., and P. T. Englund. 1979. Purification and properties of the deoxyribonucleic acid polymerase induced by vaccinia virus. *J. Biol. Chem.* 254:7812–7819.

49. Martin, C. T., and J. E. Coleman. 1987. Kinetic analysis of T7 RNA polymerase-promoter interactions with small synthetic promoters. *Biochemistry.* 26:2690–2696.

50. Falco, S. C., W. Zehring, and L. B. Rothman-Denes. 1980. DNA-dependent RNA polymerase from bacteriophage N4 virions. Purification and characterization. *J. Biol. Chem.* 255:4339–4347.

51. Lilie, H., D. Bär, …, T. Kriegel. 2011. Yeast hexokinase isoenzyme ScHxk2: stability of a two-domain protein with discontinuous domains. *Protein Eng. Des. Sel.* 24:79–87.

52. Iino, R., R. Hasegawa, …, H. Noji. 2009. Mechanism of inhibition by C-terminal alpha-helices of the epsilon subunit of Escherichia coli FoF1-ATP synthase. *J. Biol. Chem.* 284:17457–17464.

53. Ey, P. L., and E. Ferber. 1977. Calf thymus alkaline phosphatase. I. Properties of the membrane-bound enzyme. *Biochim. Biophys. Acta.* 480:403–416.

54. Froede, H. C., and I. B. Wilson. 1984. Direct determination of acetylenzyme intermediate in the acetylcholinesterase-catalyzed hydrolysis of acetylcholine and acetylthiocholine. *J. Biol. Chem.* 259:11010–11013.

Acknowledgements

Chapter 2, in full, is a reprint of the material as it appears in “Feng M, Gilson MK. A Thermodynamic Limit on the Role of Self-Propulsion in Enhanced Enzyme Diffusion. Biophysical Journal. 2019”. The dissertation author was the primary author of this publication.

CHAPTER 3: MECHANISTIC ANALYSIS OF LIGHT-DRIVEN OVERCROWDED ALKENE-BASED MOLECULAR MOTORS BY MULTISCALE MOLECULAR SIMULATIONS

Published on 19 March 2021. Downloaded by University of California - San Diego on 2/12/2022 6:45:52 AM.

PCCP

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Phys. Chem. Chem. Phys.*, 2021, **23**, 8525

Mechanistic analysis of light-driven overcrowded alkene-based molecular motors by multiscale molecular simulations†

Mudong Feng *[a] and Michael K. Gilson [b]

Received 29th December 2020,
Accepted 16th March 2021

DOI: 10.1039/d0cp06685k

rsc.li/pccp

We analyze light-driven overcrowded alkene-based molecular motors, an intriguing class of small molecules that have the potential to generate MHz-scale rotation rates. The full rotation process is simulated at multiple scales by combining quantum surface-hopping molecular dynamics (MD) simulations for the photoisomerization step with classical MD simulations for the thermal helix inversion step. A Markov state analysis resolves conformational substates, their interconversion kinetics, and their roles in the motor's rotation process. Furthermore, motor performance metrics, including rotation rate and maximal power output, are computed to validate computations against experimental measurements and to inform future designs. Lastly, we find that to correctly model these motors, the force field must be optimized by fitting selected parameters to reference quantum mechanical energy surfaces. Overall, our simulations yield encouraging agreement with experimental observables such as rotation rates, and provide mechanistic insights that may help future designs.

1 Introduction

Molecular motors are molecules that transduce optical or chemical energy into mechanical motions, such as translation and rotation, through cyclic conformational changes.[1] Naturally occurring molecular motors, such as ATP synthase and myosin, have evolved to support many essential biological functions.[2] It has also been argued that even enzymes not usually thought of as molecular motors, such as adenosine kinase, must exhibit at least weak motor-like properties, as a direct consequence of their being thrown back and forth between multiple chiral free energy surfaces corresponding to their apo and substrate-bound states.[3] The same fundamental principles also apply to non-biological molecules, and indeed, exciting progress has also been made in the design and synthesis of artificial molecular motors that may mimic or even surpass their biological counterparts.[1,4,5]

In particular, the overcrowded alkene-based motors[6,7] (abbreviated here as alkene motors) generate robust unidirectional rotation, which has been harnessed in several demonstration applications.[8–11] The alkene motors have a central double bond between two carbons (Fig. 1), which divides the molecule into a lower part, referred to as the stator, and an upper part, referred to as the rotor. If one views the motor, in the stereochemistry shown in Fig. 1, along the central bond, from the rotor to the stator, the rotor turns clockwise relative to the stator. In the first step, the central double bond undergoes *Z*–*E* photoisomerization, converting the stable form (A) to the metastable form (B). In the second step, thermal helix inversion (THI) converts B to another stable form (C). Here, the change in stator pucker from B to C resembles a flap of butterfly wings, putting the stator benzene rings from the back to the front. Another, similar, 180° rotation then converts C to A *via* D. Because the motors shown here have symmetric stators, A and C have identical energy, as do B and D. In contrast, A and D always have different stabilities, as do B and C, due to difference in ring pucker. Thus, the pucker of the rotor 5-member ring puts the methyl group in either axial (A, C) or equatorial (B, D) position, the latter giving larger repulsion with the stator, and hence higher energy.[12] As previously reported,[13–15] the metastable forms, B and D, are actually mixtures of metastable substates with distinct conformations, and hence are more complex than shown in Fig. 1.

An intuitive explanation of this process in terms of motions on the ground and excited state potential energy surfaces (PES) is provided in Fig. 2.[16] The motor starts in stable form (A) and is excited by light, undergoing a vertical transition (*i.e.*, one that does not involve a change in nuclear coordinates) to a Franck–Condon point on the excited state energy surface. It then rotates down the gradient of the excited state PES. During this

[a] *Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, 92093, USA. E-mail: mgilson@health.ucsd.edu*
[b] *Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, 9500 Gilman Drive, La Jolla, 92093-0751, USA*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0cp06685k

Published on 19 March 2021. Downloaded by University of California - San Diego on 2/12/2022 6:45:52 AM.

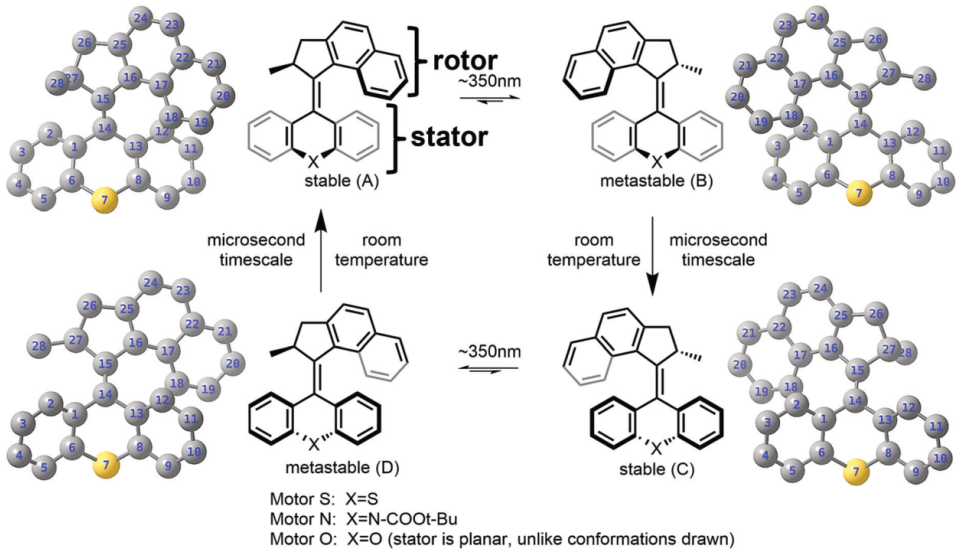

Fig. 1 Chemical structures and rotation processes of the molecular motors studied in this paper. The rate constants for their THI steps (vertical arrows) all are on the microsecond timescale (8×10^6 s^{-1}, 2×10^6 s^{-1}, 3×10^5 s^{-1} for Motors S, N, O respectively[12,16]). Although the metastable forms have multiple substates, only the lowest energy metastable substate is drawn here. The corresponding 3D structures provide atom numberings used throughout this study.

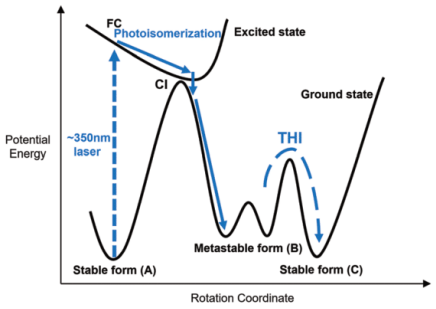

Fig. 2 Motor rotation dynamics on the ground state (lower) and excited state (upper) potential energy surfaces. FC: Frank–Condon point. CI: conical intersection. THI: thermal helix inversion. Stable and metastable forms are described in Fig. 1).

critical step, it "flies over" a large ground state energy barrier that would be very difficult to pass on the ground state PES. When it is near the conical intersection between the excited state and ground state surfaces, the molecule hops back to its ground state and continues down the energy gradient to the metastable form (B). Because there is now a high ground state energy barrier blocking a reverse motion back to A, the motor remains in B until thermal fluctuations kick it over the lower THI barrier to stable form (C). From C, the same process repeats, moving the motor through metastable form (D) and back to stable form (A).

These motors can be viewed as examples of flashing ratchets,[5,17] where switching between two different energy surfaces – here the ground state and the excited state – drives directional motion. In the absence of light, the motor would go to equilibrium on the ground state PES, and its conformational probabilities would follow the Boltzmann distribution. Consequently, the stable form would be at much higher concentration than the metastable form, and the principle of detailed balance[17] would mean that no net rotation occurs. In the presence of a constant light source, photo-excitations break detailed balance and create a non-Boltzmann steady-state distribution, the so-called photostationary state. Unlike in equilibrium, the steady state concentration of the metastable form can match or surpass the stable form.[16,18] Thus, the motor molecules can transduce light energy by letting photoisomerizations pump them to the metastable form, and releasing their stored energies during THI.

Given the ultimate goal of using molecular motors for practical applications, it is of interest to consider performance metrics to be optimized when designing them. One obvious metric is the rotation speed. There have been considerable efforts to make faster alkene motors, as reviewed from both experimental[6,7] and theoretical perspectives.[19] These have focused on accelerating THI, which is the rate-determining step of the rotational cycle, without sacrificing unidirectionality. The THI rate constant can be measured experimentally by monitoring the relaxation of the absorption spectrum after pulse excitation.[16] It can also be obtained from molecular simulations by determining the mean first passage time for the transition

Published on 19 March 2021. Downloaded by University of California - San Diego on 2/12/2022 6:45:52 AM.

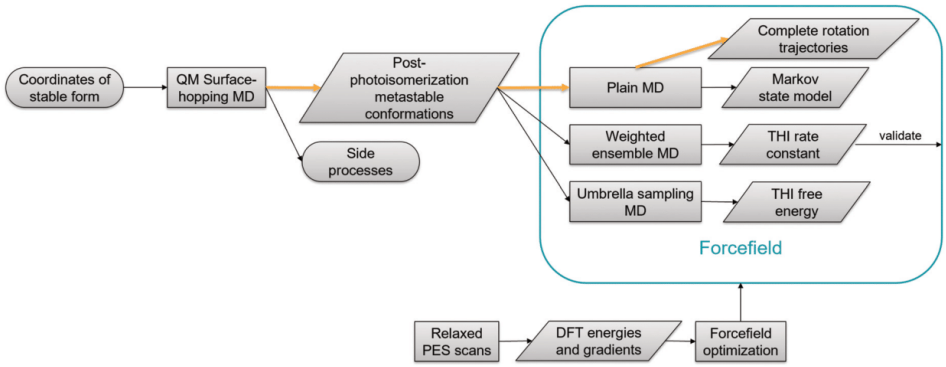

Fig. 3 Workflow describing the proposed computational framework. Rectangles represent processes, whereas non-rectangular parallelograms stand for results/data. All force field-based methods and their results are encompassed by a rounded rectangle standing for the force field used. Combining the MD results along the orange arrows give complete 180° rotation trajectories.

from the metastable form to the stable form. Thus, comparing to experiments provide a way to gauge simulation accuracy. Another performance metric, the power output, equals the rotation speed times the average work the motor does in each rotation cycle. However, the power output depends on factors extrinsic to the motor, notably the magnitude of load and how the load is coupled to the motor. A more general and intrinsic property is the thermodynamic maximum of the work per cycle, which is related to the free energy released when the motor undergoes THI.

Important prior studies have developed quantitative models of these molecular motors at varying levels of molecular detail. Geertsema *et al.* constructed a Markov state model (MSM) with rate constants fitted to NMR measurements,[18] and evaluated serveral performance metrics such as the average rotation speed and the degree of rotation unidirectionality. Their model provides a valuable perspective of concentration fluxes in the system at a macroscopic level, but does not offer microscopic details beyond the simple dichotomy of stable *vs.* metastable form. Other studies have considered the microscopic level, as follows. Computational studies have used static quantum mechanical (QM) methods, such as DFT geometric optimizations of energy minima and transition states, that can provide accurate energies,[14,20] but these are not fully informative about the statistical mechanics of the rotational process. Short molecular dynamics (MD) simulations at the picosecond timescale have also been used to examine the photoisomerization step but not the slower THI step.[21–24] However, classical MD can now routinely reach the microsecond timescale for such small systems, and thus is directly applicable to the subset of motors with THI rate constants on the microsecond time scale.[12,16,25] These motors are more challenging to characterize experimentally, because of their high rates, but they are easier to characterize with classical MD.

Here, we propose a multiscale modeling framework that couples QM surface-hopping MD, used to model the photoisomerization step, with classical MD, used to model the THI step (Fig. 3). This combination of methods allows modeling of the entire photoisomerization process, as the surface-hopping method includes a description of the electronic degrees of freedom in the ground and excited states and thus allows simulation of initial motions on the excited state energy surface and the stochastic hop back to the ground state, while the classical MD allows longer timescale simulations on the ground-state surface, which are required to model the slower THI process, and also allows use of a more detailed treatment of the solvent. The accuracy of our classical MD simulations is supported by optimization of selected force field parameters to fit QM energies and gradients for these specific compounds. The rotation mechanism is elucidated by full rotation trajectories (orange arrows in Fig. 3) and Markov state models that provide conformations and interconversion kinetics of the metastable substates. We also assess motor performance by using weighted ensemble simulations to compute the THI rate constant, and umbrella sampling to compute the free energy drop going from the metastable form to the stable form. Encouragingly, the THI rate constants computed agree well with experimental measurements. The results thus may benefit future design of similar molecular motors.

2 Methods

The computational framework in this study, illustrated in Fig. 3, has three main components. First, the photoisomerization process is modeled with QM surface-hopping MD; second, classical MD at the microsecond timescale is used to study the THI mechanism and obtain quantities related to performance, such as rate constants and free energies. Finally, to achieve realistic classical MD results, we adjust force field parameters against conformational energies from reference QM calculations. The three components are described in the following subsections. Additional details are provided in the ESI;† scripts

Published on 19 March 2021. Downloaded by University of California - San Diego on 2/12/2022 6:45:52 AM.

and input files to reproduce this study can be downloaded from a Github repository (github.com/fengmudong/motor-paper).

2.1 Surface-hopping MD simulation

We used QM surface-hopping MD, run in replicates with slightly different starting structures, as detailed below, to model the photoisomerization step of the motors in Fig. 1. These simulation replicates provide a diverse and physically relevant sample of post-photoisomerization structures that serve as initial structures of our classical MD simulations.

Surface-hopping simulations were carried out at the OM2/MRCI level, which is a semi-empirical QM method designed for fast excited state calculations[26] that gives competitive accuracy.[27] The software package MNDO[28] was used for these calculations. We first ran 10 ps of sampling in the ground state, starting from the optimized structure of the stable form (3D structure C in Fig. 1). Then, samples drawn at random from this initial ground state trajectory were set to the first excited state without any change in nuclear coordinates, if they met standard MNDO criteria such as successful mapping of the active space orbitals, and if they passed a stochastic selection based on the computed excitation probability.[28] About 70% of the initial samples passed and were excited. These excited samples represent an ensemble of Franck–Condon points, and from each we started a simulation replicate. Tully's fewest switches algorithm[29] was used to determine hopping to the ground state. Energies and gradients were evaluated on the fly at the OM2/MRCI level until the fixed simulation length of 4.5 ps was reached. A replicate was kept for further use if it had hopped to the ground state and ended the 4.5 ps run in the metastable form. The final coordinates of such replicates were used as initial coordinates of classical MD simulations. The other replicates, representing the side processes in Fig. 3, were not used to constitute the full rotation trajectory, but offer additional insights, as discussed in the Results and discussion section.

These surface-hopping simulations used a Langevin thermostat[30] targetting 90 K, the temperature used in the relevant photoisomerization experiments,[16] with a collision frequency of 0.2 ps^{-1}. The Langevin thermostat provides a random "kick" to each atom at every time-step, to model solvent impacts, and a drag on each atom, proportional to its velocity, to model the effect of solvent viscosity. The balance of these energy-adding and damping forces effectively sets the target simulation temperature. We also tried the Nose–Hoover thermostat,[30] an alternative to the Langevin thermostat, but this led to non-physical low-frequency kinetic energy oscillations not present in Langevin or constant energy simulations, even when a long Nose–Hoover chain was used, so we chose not to use this method.

2.2 Classical MD simulation

First, classical MD simulations were appended to the QM surface-hopping MD trajectories (Section 2.1) to construct full rotation trajectories of Motor S and Motor N. This was not done for Motor O, for reasons detailed in Section 2.3) The initial structure of each classical simulation was set as the final structure of one of the QM surface-hopping trajectories in the metastable form (Section 2.1). One motor molecule was solvated with 600 molecules of dichloromethane in a 40 Å cubic periodic box, using the software Packmol.[31] After energy minimization, the solute was restrained in position while the system at constant volume was heated to 300 K with the Berendsen thermostat. It was then equilibrated, with restraints still on, at 300 K and a pressure of 1 bar with the Langevin thermostat and the Monte Carlo barostat, mimicking the experimental conditions for microsecond THI.[16] Finally, restraints were removed for 1 μs production runs at the same temperature and pressure. In classical MD, the Langevin collision frequency we mainly use was 1 ps^{-1}, but different values were also used to evaluate the effect of this parameter on the kinetics (discussed later in Section 3.3.4). The trajectories that completed THI were concatenated with their corresponding surface-hopping MD trajectories to give continuous trajectories of the full 180° motor rotation.

Then, these production MD trajectories (whether they had completed THI or not) were used to construct a reversible Markov state model (MSM) of each motor. The trajectory segments in the stable form after THI were excluded when building the MSM, because the reverse THI process was not sampled. The high dimensional trajectories of atomic positions were first condensed to time series of all the dihedrals present in the topology file of the motor; each dihedral is associated with a force field term. After further dimension-reduction using time-lagged independent component analysis (TICA), the time series in continuous conformational space was discretized to conformational clusters using *k*-means clustering. Then the clusters were classified by Perron cluster analysis (PCCA)[32] into 4 metastable substates. We chose 4 substates because partitioning the metastable form into a larger number of substates led to some substates being overly similar to each other. The time series of substate classification were then analyzed to produce the MSM, as a 4×4 transition matrix.

Although plain MD yielded a few trajectories that completed THI, many more such runs would be needed to obtain better precision for THI rate constants.[33] Therefore, we instead used an enhanced sampling technique, the weighted ensemble method implemented in WESTPA,[34] to compute the THI rate constant. In WESTPA, a collection of MD replicates initiated from metastable form is managed by cloning new ones in regions of the conformational space with fewer replicates, and merging existing ones in regions with more replicates, to generate a more even distribution of replicates along desired progress coordinates. This allows an artificial increase in the number of replicates that sample higher energy transitional regions. Correct statistics is maintained by adjusting replicate weights when replicates are cloned or merged. Note that this method does not involve modifying the Hamiltonian. As detailed in Results and discussion, THI progress coordinates were defined for each motor in terms of key structural features, such as dihedrals describing rotation around the central double bond and rotor/stator puckering coordinates. Then the conformational space was divided into bins of different progress coordinate values, and WESTPA was used to monitor the number of MD replicates in each bin to decide when to clone or merge

Published on 19 March 2021. Downloaded by University of California - San Diego on 2/12/2022 6:45:52 AM.

replicates. When a replicate reached the bin defining the stable form of the motor, the replicate was restarted at one of the initial metastable conformations. As the probability distribution across bins converge to steady state distribution, the THI rate constant is calculated using the steady state flux from the metastable form to the stable form.

We calculated the free energy difference between the stable and metastable forms from potentials of mean force (PMF) obtained with umbrella sampling.[35] To do this, we used harmonic potentials of spring constant spring constant 200 kJ mol^{-1} to restrain the central double bond torsion angle in a series of 1 μs MD simulation windows. The torsion value at each window center ranges from −1.9 to 2.1 radian, with 0.1 radian spacing. For reasons given in Appendix B, the central torsion angle is defined here as 43-15-14-29, where 43 and 29 are hydrogens attached to C18 and C2 in Fig. 1, respectively. This range spans the metastable form, the stable form, and the barrier between them. Umbrella sampling simulations were set up using the open-source library PLUMED,[36,37] and the PMF along the torsion was obtained by Weighted Histogram Analysis Method (WHAM).[38,39] From the PMF $G(x)$, which is essentially the conditional probability distribution expressed as a free energy profile, the free energy difference between the metastable form and the stable form was calculated by integrating across the corresponding ranges of the torsion values: $G_m - G_s = RT \ln \int_s e^{-\beta G(x)} \mathrm{d}x - RT \ln \int_m e^{-\beta G(x)} \mathrm{d}x$. Here the s and m subscripts indicate the sampling windows that span the stable and metastable forms, respectively.

2.3 Force field optimization

As detailed in Results and discussion, we found that an initial "off the shelf" force field did not yield realistic results in classical MD simulations of these molecules. We therefore optimize selected force field parameters by fitting them to reference QM energy surfaces.

2.3.1 Generation of QM potential energy surfaces. Because the classical MD simulations address only ground state dynamics, ground state QM suffices to generate reliable fitting targets, including energies and energy gradients with respect to atomic positions. We chose DFTD3/DZVP level of theory, as it has been shown to strike a favorable balance between accuracy and computational cost in calculating conformational energies for molecules of this size.[40] The motor rotation process of interest involves conformations where the central carbon–carbon bond is twisted, so that atoms 16-15-14-13 are not coplanar. In such conformations, π orbital overlap is compromised, resulting in open shell character of the electronic structure; this motivated us to use unrestricted DFT orbitals.

We computed two dimensional PES by running constrained geometry optimizations at a grid of points along two predefined scan coordinates. Based on prior analyses of these motors[13,24] and our own trial results, we chose two combinations of scan coordinates: central torsion angle 18-15-14-10 combined with pyramidalization 14-1-13-15, and central torsion angle 18-15-14-10 combined with rotor pucker 27-15-16-25. In this pucker definition, positive/negative value of rotor pucker dihedral means equatorial/axial, respectively. The rationale for choosing these coordinates is provided in Appendix B.

We used the wavefront propagation algorithm implemented in TorsionDrive,[41] an open source package, to generated QM surfaces of motors S, N and O (Fig. 1). TorsionDrive runs geometry optimizations constrained at each grid point multiple times, from different propagating directions, to find the lowest energy structure at that point. TorsionDrive also benefits from using multiple starting structures to seed the propagation; the seed structures we used were the final metastable form structures from QM surface-hopping MD, plus the structures shown in Fig. 1. Relative to less exhaustive scanning methods in the literature, such strategy is less sensitive to convergence failures at some grid points, to hysteresis (*i.e.* different results when the same point is scanned from different direction), and to optimizations being trapped in local minima rather than the lowest energy structure at the grid point. As shown previously,[41] these problems can be particularly severe for alkene motors, likely due to nontrivial size (48 atoms for Motor S) and high rotation barriers, but problems are largely alleviated by the TorsionDrive method. Energies and gradients were computed by the Psi4 package[42] for use in TorsionDrive.

2.3.2 Parameter optimization. We used a Newton–Raphson algorithm implemented in the open-source package ForceBalance[43] to systematically optimize force field parameters, so that the force field replicates as much as possible the QM energy surfaces (Section 2.3.1). During optimization, force field energies and gradients of the unsolvated motor molecule were compared with the corresponding QM energies and gradients. The objective function is the sum of squared energy and gradient differences between the force field and QM across the grid of conformations. For Motor O, it proved necessary to employ additional regularization terms, *i.e.* penalty terms to discourage unphysically large changes of certain parameters, as described in Results and discussion.

As detailed in Results and discussion, we tested various choices of what parameters to optimize, and we found that optimizing dihedral (proper and improper) amplitudes alone sufficed for Motor S and N, whereas Motor O needed more changes. Our procedures ensure that these parameters remain non-negative during optimization. No changes were considered to the functional form of the force field, the multiplicities and phases of torsional terms, and the partial atomic charges. Initial force field parameters were taken from general Amber force field (GAFF) version 2.1[44,45] implemented in AmberTools18,[46] with AM1/BCC partial charges.[47] A subtlety of atom type selection for the motors is discussed in the Appendix C. For the dichloromethane solvent at room temperature and pressure, the GAFF2 parameters gave a density of 1.25 g cm^{-3}, in good agreement with the experimental value of 1.33 g cm^{-3},[48] so the GAFF2 parameters were used as is for the solvent.

2.4 Selection of reaction coordinates

It is often necessary to map chemically meaningful conformational differences to a few coordinates, which may be followed over time and through transitions. Here, we call these reaction

Published on 19 March 2021. Downloaded by University of California - San Diego on 2/12/2022 6:45:52 AM.

coordinates, but other synonyms are often used; *e.g.*, progress coordinates or collective variables. In general, intuition based on one coordinate alone, such as the central torsion angle, can be unreliable; it is preferable to account for all degrees of freedom in the system using methods like MD simulations, and then verify hypotheses about key coordinates through analysis of the resulting trajectories. For the motors in this study, the key conformational features are the rotation of the rotor relative to the stator, the pucker of the stator, the pucker of the rotor, and the pyramidalization of the two central carbons. We found that one well-chosen reaction coordinate suffices to differentiate between the metastable form and the stable form, as in the umbrella sampling calculations; however, at least two coordinates are needed to differentiate all the conformationally distinct and kinetically separated metastable substates. The approach used to define reaction coordinates in this study is detailed in the Appendix B.

3 Results and discussion

3.1 Modeling photoisomerization with QM surface-hopping MD

3.1.1 Analysis and fate of the simulation trajectories. For each of the three motors in Fig. 1, the photoisomerization step was simulated in multiple replicates that represent an ensemble of Franck–Condon points (Section 2.1). As summarized in Table 1, about 10% of the simulation replicates of Motor S and Motor O progressed to the ground state metastable form. Each of these simulations offers a chemically reasonable realization of photoisomerization. The trajectories are similar to those of other alkene motors in prior works,[22,24] in the sense that the conformations at which the excited to ground state hops occur are similar, as are the rotational processes. A video of a representative trajectory of Motor S is provided in the ESI† (mndo-amber.mp4). Final ground state metastable form conformations of Motor S and Motor O were used as starting conformations for the classical MD simulations detailed in Section 3.3. The starting conformations excluded a few that had not yet relaxed to a sufficiently low-energy conformation and thus were not suitable starting points for the classical simulations. This left 13 and 4 initial conformations for Motors S and O, respectively. For Motor N, the QM simulations did not yield any hops from the excited state to the ground state, despite running for similar durations and numbers of replicates.

In an interesting side-process, a few replicates of Motors S and O actually completed THI (Table 1), progressing through the metastable form to the stable form. In all three of these simulations for Motor S, the transition to the stable form is immediately preceded by flipping of the stator pucker to that of the target stable form, as seen in the representative trajectory (video mndo-whole.mp4 in the ESI†). The role of the stator pucker in THI is discussed further in Section 3.3. These rapid THI events appear to occur too quickly relative to experimental observations that THI occurs on the microsecond timescale.[12,16] On one hand, these rapid THI events may be artifacts of the simulation methodology. On the other hand, the computational result could be reconciled with experiment if the kinetics of THI is bimodal, with a fast mode that happens on the picosecond timescale and a slow mode that happens on the microsecond or millisecond timescale. In the fast mode, photoisomerization and THI would happen in quick succession, and the motor would not have time to dissipate much of the kinetic energy from the excited to ground state hop before crossing the THI energy barrier. In the slow mode, however, the motor has time to fully equilibrate, with the kinetic energy from photoisomerization dissipated into many solute and solvent degrees of freedom before THI. If most trajectories dissipate their energy before completing THI, thus following the slow mode, then the fast mode would be a side process which does not contribute much to the overall kinetics observed in experiment. The experimental data available for these motors[12,16] does not seem to allow one to determine whether THI in fact follows such bimodal kinetics.

As listed in Table 1, most replicates of all motors studied had not progressed to successful photoisomerization by the end of the simulations. Many of these replicates were still in the excited state, but there were also many that hopped to the ground state and fell back to the initial stable form, rather than progressing to the metastable form. These are unproductive replicates, in the sense that they would not lead to rotation of the motor. Quantum yields estimated from the fraction of productive replicates are on the same order of magnitude as reported experimentally for non-MHz alkene motors.[13]

3.1.2 Effects of thermostat and solvent properties. The present simulations use the Langevin thermostat to regulate temperature (Section 2.1). With this method, a given temperature can be modeled by larger random forces (‘‘kicks’’) and a larger drag coefficient, or by smaller kicks and a smaller drag coefficient. This scaling is controlled by the collision frequency parameter, which is connected with the viscosity. In the simulations described above, we used a collision frequency of 0.2 ps^{-1}, so that the kinetic energy released during photoisomerization dissipates into the bulk solvent on a several picosecond timescale, in agreement with experimental observations for a related

Table 1 Outcomes of surface-hopping molecular dynamics simulation replicates. Columns, from left to right, provide the motor name, the duration of each individual replicate simulation for this motor, the number of simulation replicates, the total number of replicates that failed to isomerize away from the initial stable form, the number of replicates remaining in the excited state, the number that completed photoisomerization and terminated in the metastable form, and the number that completed both photoisomerization and THI and thus terminated in the stable form (all replicates that completed photoisomerization also hopped from the excited state to the ground state)

			Not isomerized		Isomerized	
Motor	Time per replicate (ps)	Replicates	Total	Still in excited state	Metastable Form	THI Completed
S	4.5	194	170	120	21	3
O	4.5	71	64	55	6	1
N	5.0	71	71	71	0	0

Published on 19 March 2021. Downloaded by University of California - San Diego on 2/12/2022 6:45:52 AM.

photoexcited system.[49] For completeness, however, we also explored larger and smaller values of this parameter.

With a lower collision frequency of 0.02 ps^{-1}, corresponding to weaker kicks and lower viscous drag, the thermostat has less influence on the dynamics, so the kinetic energy added to the system by downhill motion on the excited state PES and by the subsequent hop to the ground state led to larger and longer-lasting rises in the effective temperature (kinetic energy) of the motors. Presumably because of the more extreme temperature rises, we observed more replicates that completed both photoisomerization and THI during the QM simulations. Interestingly, setting the collision frequency to zero, which decouples the motor from any heat bath and leads to a constant energy simulation, caused the photoisomerization success rate to drop, and a number of replicates ended in numerical failure, presumably due to motions too fast for the selected time-step. With a higher collision frequency of 2 ps^{-1}, the fraction of replicates that successfully hopped from the excited to the ground state within the simulation time was markedly reduced. Presumably, the increased viscosity slows rotation of the rotor toward the conical intersection, delaying return to the ground state.

Finally, the present simulations have neglected the dielectric properties of the solvent. The potential consequences of this are worth considering, as the commonly used solvent dichloromethane has a dielectric constant of about 9.[48] A prior QM study[50] reported that the excited state of a motor in this class has a dipole moment of about 1.1 Debye in the ground state and 6.8 Debye in the excited state. Therefore, the reaction field of a polar solvent should tend to stabilize the excited state preferentially and thus, presumably, reduce the amount of potential energy released as kinetic energy during photoisomerization. This might reduce the tendency of the motor to rapidly go through THI. However, the magnitude of any such effect is uncertain, and might be elucidated by more detailed calculations.

3.2 Optimization of force field parameters for overcrowded alkene motors

The accuracy of classical MD depends on the accuracy of the force field used. To our knowledge, no specialized force field exists for alkene motors, except one reparametrized from OPLS for classical MD simulation of photoisomerization, rather than THI.[21,22] In addition, our initial tests of GAFF2 force field yielded problematic results for these molecules, as detailed below. We therefore parameterized a specialized force field for each motor molecule by fitting to its unique QM potential energy surfaces and gradients, as detailed below.

3.2.1 Motor S and Motor N. The QM energy surfaces of Motor S provided by TorsionDrive are smooth and well-converged, as evident in Fig. 4A and B. The two deep minima in each energy surface correspond to the stable forms in Fig. 1, and the shallower minima correspond to the metastable form. This clean separation of stable form and metastable form was not achieved by other plausible choices of reaction coordinate, such as when the central torsion angle was defined by atoms 16-15-14-13 or when using the stator pucker as a scan coordinate. The lack of perfect 180° periodicity along the central torsion results from the geometric

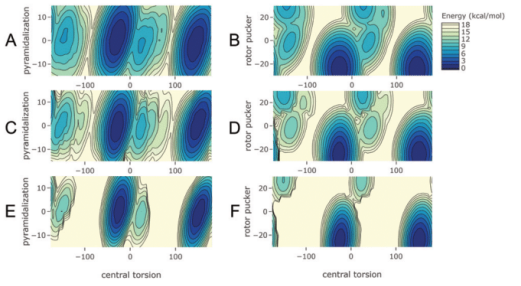

Fig. 4 Potential energy surfaces of Motor S. (A and B) Quantum mechanical (QM) energy. (C and D) Energy from force field after fitting to QM. (E and F) Energy from GAFF2 force field before fitting to QM. The scanned coordinates are the central torsion angle, 18-15-14-10; the pyramidalization improper dihedral, 14-1-13-15; and the rotor pucker dihedral, 27-15-16-25, with atom numberings from Fig. 1. The force field energy for each grid point was evaluated at the corresponding QM conformation without further relaxation.

properties of the torsion and the nonplanarity of the rotor and stator. The motors rotate from right to left in these graphs (negative direction of the central torsion) because the metastable to stable transition along negative x direction has a moderate barrier height, whereas the metastable to stable transition along the positive x direction has an extremely large barrier.

The GAFF2 force field yields energy surfaces (Fig. 4E and F) that deviate significantly from QM (Fig. 4A and B). The energy barriers between minima are higher and wider than QM, and the metastable local minimum at around (40, 0) in QM (Fig. 4B) is missed by GAFF2. Furthermore, we found that the unoptimized force field gave fundamentally incorrect conformations in MD simulations. For example, the stator heterocycle adopted a chair conformation, whereas experiments[13,16] and QM suggest a boat conformation. GAFF1 yielded similar problems, and we expect other general force fields, if not tuned for alkene motors, to be inadequate in quantitative studies, if not qualitatively wrong.

Following optimization, the force field energy surfaces (Fig. 4C and D) correctly represent all local minima and yield energy barriers similar to the QM results. It is notable that the optimization reduces the amplitudes of torsional energy terms associated with the central double bond and thus addresses the issue that the initial force field overestimated the barrier heights. In addition, optimization causes the various improper dihedral terms to have more varied amplitudes, ranging from essentially zero to 10.0 kcal mol^{-1}, whereas all the GAFF2 amplitudes are 1.1 kcal mol^{-1}. These changes presumably enable the optimized parameters to provide correct ring conformations. Note that only proper and improper dihedral terms were adjusted; other force field terms, such as bond-stretches and bond-angles, were left unchanged. For Motor N, the initial GAFF2 parameters again yield potential energy surfaces significantly different from the QM surfaces, and optimization of dihedral terms again yielded potential energy surfaces much closer to QM (Fig. S1 in the ESI†).

3.2.2 Motor O. Optimization of force field parameters for Motor O posed additional challenges. Although optimization of

Published on 19 March 2021. Downloaded by University of California - San Diego on 2/12/2022 6:45:52 AM.

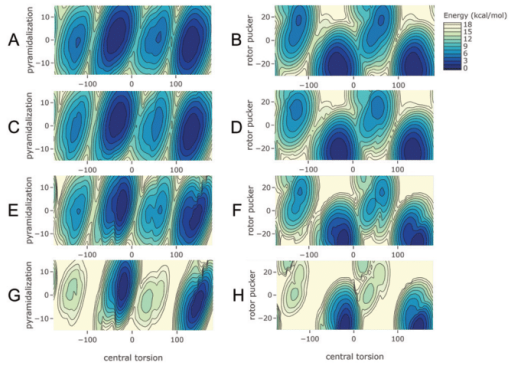

Fig. 5 Potential energy surfaces of Motor O. (A and B) Quantum mechanical (QM) energy. (E and F) Energy from force field after fitting only dihedral terms to QM. (C and D) Energy from force field after fitting dihedral terms, bond and angle terms, and terms of Buckingham potentials. (G and H) Energy from GAFF2 force field before fitting to QM. See Fig. 4 for definitions of the scan coordinates.

force field dihedral parameters improves agreement with QM energy surfaces (Fig. 5E and F), classical MD simulations with the resulting fitted parameters give no THI transitions at all for Motor O following optimization, implying an underestimation of the rate constant.

We therefore tried optimizing not only dihedral terms for Motor O, but also the σ and ε values of each Lennard-Jones (LJ) term, as well as the force constants and equilibrium lengths of bond-stretch and angle-bend terms. These terms strongly affect the steric interactions between the stator and rotor during THI. To avoid nonphysically large changes in LJ terms, flat bottomed restraints were applied, to allowed maximal changes of 0.1 Å in σ and 0.015 kJ mol^{-1} in ε. The resulting force field is softened, with σ and ε smaller in most cases, and also reduced bond-stretch and angle-bend force constants. As expected, using the softened LJ terms in classical MD simulations allowed a greater THI rate that matches experiments. However, in this case the optimization algorithm made bond and angle force constants unphysically small, causing the molecule to be overly flexible during simulation. For example, the force constant of the central double bond, which is conjugated and thus should have a bond order between one and two, fell from the GAFF2 value of 855 to 312 kJ mol^{-1} Å^{-2}. Note that even the GAFF2 value is probably on the low side, because the force constant for a benzene C–C bond, which also is of order between one and two, has been reported as 1124 kJ mol^{-1} Å^{-2}.[51]

We propose that the reason Motor O simulations face the dilemma of either underestimating the THI rate constant or using unrealistically soft bonds and angles is that the 12th power repulsion of the LJ functional form is unrealistically steep[52] and therefore overestimates the van der Waals repulsion between the rotor and the stator during THI. For Motor O, among the scanned QM conformations, the closest distance between the rotor methyl H and the stator H on C12 is as close as 2.0 Å, well below the minimal-energy distance of 3.0 Å for this interaction in the force field, and well into the range of distances where the 12th power repulsion becomes problematic. Motors S and N, in contrast, have a puckered and more flexible stator, and thus can alleviate this repulsion during THI, making the LJ repulsion acceptable for simulation of these motors.

We tested this explanation by replacing the LJ potential with the Buckingham potential,[52] which has a less steep and hence more realistic exponential repulsion term. The initial Buckingham parameters were taken from Table 3 of Engler of coworkers[53] and the B cofficients were optimized without restraint. Bond-stretch and angle-bends were allowed to vary, subject to flat bottom restraints that prevented relative changes greater than 10%. Optimization with the Buckingham potential led to improved agreement with QM, as shown in Fig. 5C and D, without abnormally small force constants. Thus, the Buckingham potential may replace LJ potential for simulations of more congested molecules like Motor O. Unfortunately, support for the Buckingham potential in simulation software is currently limited. For example, it is not implemented in AMBER and, although it is implemented in GROMACS, it is not available in the GPU version. This made it impractical for use in simulations of Motor O at the timescale needed for the present study, so no further results are presented for this molecule. Future progress in software infrastructure is needed to enable Buckingham potential simulations at longer timescales.

3.3 Classical MD simulation

3.3.1 Characterization of plain MD trajectories. Thirteen classical MD simulations of Motor S were initiated from metastable form ground state conformations available at the end of the QM surface-hopping simulations (Section 3.1), and each was run for a total of 1 μs. Because the surface-hopping simulations of Motor N did not yield any metastable conformations we could use to initiate classical MD (Section 3.1), for this motor we instead initiated four 1 μs classical MD simulations, one from each of the four metastable substates identified by the MSM (Section 3.3.2). No obvious pathologies were observed in the course of these simulations for either motor, and several interesting trends were observed.

First, 10 of the 13 simulations of Motor S and one of the four simulations of Motor N completed THI during their 1 μs duration. The full rotation process, consisting of excited state dynamics, decay to the ground state, ground state dynamics within the metastable form, and THI, is captured by the concatenation of a QM surface-hopping trajectory and its subsequent classical MD trajectory for Motor S in video mndo-amber.mp4 in the ESI.†

Intriguingly, just as seen for THI transitions during the QM surface-hopping MD simulations (Section 3.1), every THI transition in the classical simulations was immediately preceded by flipping of the stator pucker to that of the target stable form. This is not visible in video mndo-amber.mp4 (ESI†) because of the long time interval between frames, which was chosen to allow viewing of the full rotation process. Therefore, in videos amber-SnearTHI.mp4 and amber-NnearTHI.mp4 of the ESI,† we zoom in to animate the details near the THI transition.

Published on 19 March 2021. Downloaded by University of California - San Diego on 2/12/2022 6:45:52 AM.

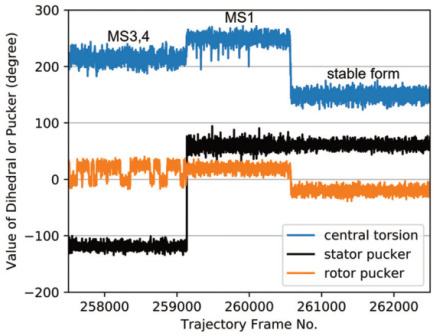

Fig. 6 Time series of the central torsion angle and puckers before, during, and after a THI transition simulated by classical MD. Forward rotation of the rotor relative to the stator corresponds to a decrease in the central torsion angle. The puckers are defined so that positive/negative rotor pucker mean equatorial/axial, respectively, and positive/negative stator pucker mean the same/opposite pucker as final stable form, respectively. Labels indicate which conformational state (Section 3.3.2) is occupied at each step.

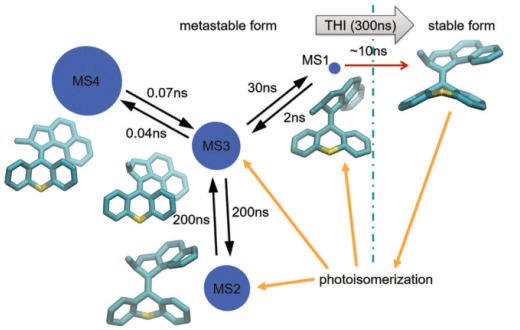

Fig. 7 Kinetic model of Motor S rotation at 300 K. Each metastable substate is represented as a circle of area proportional to steady state probability, along with a representative 3D structure. Black arrows are drawn for each nonzero transition matrix element of the MSM. Numbers on the arrows are corresponding transition time constants calculated with the MSM. A red arrow represents THI transitions that appear to always go through MS1 in classical MD; the associated time constant is calculated as the mean wait time in MS1 before transition to stable form. The gray arrow indicates the overall THI process and gives the overall THI rate constant estimated with WESTPA (Section 3.3.4). Additional orange arrows are drawn to represent photoisomerization to metastable substates seen at the end of the QM surface-hopping MD runs.

Fig. 6 shows the time series of the key coordinates, *i.e.* the central torsion angle and puckers, to illustrate the transitions in amber-SnearTHI.mp4 (ESI†). The first event is a sharp change in the pucker of the stator to that of the final stable form, along with a transient backward rotation of the rotor relative to the stator. This is followed by a sharp forward rotation of the rotor to reach the stable form. Meanwhile, the rotor pucker fluctuates rapidly between axial and equatorial until the stator pucker changes, at which point the rotor pucker becomes locked in the equatorial conformation until the rotor rotates to the stable form, at which point the rotor pucker becomes locked in the axial form. Sections 3.3.2 and 3.3.4 provide further details regarding the THI mechanism and kinetics.

3.3.2 Substate analysis of the metastable form *via* Markov state modeling. The conformational substates of the metastable form of Motor S and their transition network were obtained in the form of a Markov state model (MSM), through analysis of the thirteen classical MD trajectories. Representative conformations of the four resulting substates are depicted in Fig. 7, along with circles indicating their steady state probabilities estimated from the MSM transition matrix. The ESI† furthermore provides Cartesian coordinates of ten sample conformations for each Markov state. The transition time constants, *i.e.* inverse rate constants, are shown with black arrows representing the associated transitions between substates. Orange arrows represent photoisomerization. As noted in Sections 3.1 and 3.3, the transition to the stable form is always preceded by a flip of the stator pucker to that of the target stable form. The MSM groups the conformations with this pucker into a single substate, called metastable substate 1 (MS1), and an additional red arrow goes from MS1 to the stable form in Fig. 7. The 10ns time constant for this transition is estimated from the simulation time spent in MS1 for all 13 trajectories, divided by the number of THI transitions observed. The large arrow from the metastable form as a whole to the stable form is labeled with rate constant computed by WESTPA simulation (Section 3.3.4). The states and labelled arrows in Fig. 7 define a complete kinetic model of the whole motor rotation process, which is analyzed quantitatively in Section 3.3.3.

Our kinetic model in Fig. 7 indicates that MS1, despite its low population, is a necessary stop on the path from the metastable to the stable form. Thus, THI can be viewed as taking place in two consecutive steps: change of the stator pucker to that of the final stable form, followed by rotation of rotor, relative to the stator, to complete THI. Metastable substates 3 and 4 (MS3, MS4), which hold the majority of the population, are similar in conformation, as they only differ by their rotor pucker, and they interconvert quickly. Metastable substate 2 (MS2) is the closest to the stable form in terms of its central torsion angle (defined as *e.g.* 18-15-14-10 or 17-18-2-1), but the model implies that MS2 is actually the furthest kinetically from the stable form. This finding highlights the potential drawback of choosing a single geometric feature as the reaction coordinate: using only the central torsion angle as the reaction coordinate risks giving the incorrect impression that a motor in MS1, MS3, or MS4, needs to go to MS2 as the obligatory last step before reaching the stable form. Exploring mechanisms by unconstrained MD simulations, instead of exploring along predefined reaction coordinates, helps avoid such problems. These high dimensional MD trajectories are free of bias or information loss from inexact assumptions based on low dimensional intuitions, yet the trajectories can be projected to informative lower dimensional pictures based on analysis of structure and kinetics.

Published on 19 March 2021. Downloaded by University of California - San Diego on 2/12/2022 6:45:52 AM.

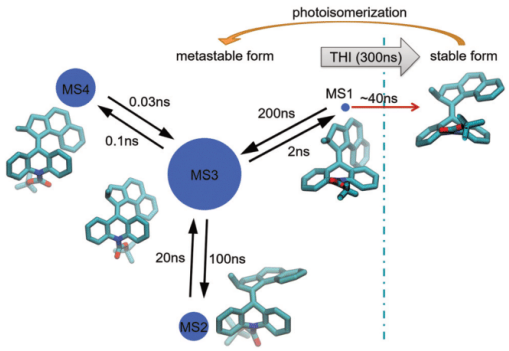

Fig. 8 Kinetic model of motor N rotation at 300K. Each metastable substate is represented as a circle of area proportional to steady state probability, along with a representative 3D structure. Black arrows are drawn for each nonzero transition matrix element of the MSM. Numbers on the arrows are corresponding transition time constants calculated with the MSM. A red arrow represents THI transitions that appear to always go through MS1 in classical MD; the associated time constant is calculated as the mean wait time in MS1 before transition to stable form. The gray arrow indicates the overall THI process and gives the overall THI rate constant estimated with WESTPA (Section 3.3.4), and the orange arrow indicates photoisomerization.

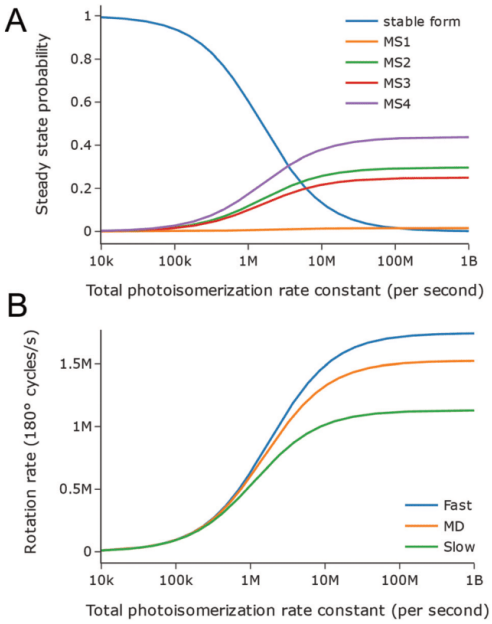

Fig. 9 Steady state kinetics of Motor S at 300 K. (A) Dependence of the steady state probabilities on the total photoisomerization rate. (B) Dependence of steady state rotation rate of the motor on the total photoisomerization rate, where the three curves correspond to three different distributions of metastable substates generated by photoisomerization: the distribution that leads to the fastest rotation (MS1 : MS2 : MS3 : MS4 = 1 : 0 : 0 : 0), the distribution that leads to the slowest rotation (MS1 : MS2 : MS3 : MS4 = 0 : 1 : 0 : 0), and the distribution inferred from our surface-hopping MD trajectories (MS1 : MS2 : MS3 : MS4 = 7 : 2 : 4 : 0). In A, the distribution from MD is used.

We also generated a MSM for Motor N from four 1 μs classical MD trajectories, each started from a conformation similar to one of the metastable substates of Motor S. The MSM was again constructed from the time series of all the dihedrals with associated force field terms, except those involving at least one atom of the –COO*t*Bu group. The resulting model (Fig. 8) is very similar to that of Motor S.

Prior studies using static QM methods also delineated metastable substates of a motor very similar to the present Motor S.[14,15] Although the conformations of those substates are similar to the ones reported here, the prior studies proposed a different pathway to the stable form. Rather than going through the relatively high energy MS1 as found here, their pathway goes to the stable form through the lower energy MS2. Potential explanations for the difference between the prior results and those presented here include the fact that the motor molecule studied previously had an extra methoxy group, relative to Motor S; the prior studies' use of high-level QM calculations throughout; the challenge of finding all relevant transition states by QM geometric optimization in the prior approach; and our use of a dynamical, rather than a static, method.

3.3.3 Steady state kinetics of Motor S. The diagram in Fig. 7 offers a kinetic model of Motor S and includes estimated rate constants for many of the interstate conversions. Although we do not have estimated rate constants for photoisomerization of the stable form into the various metastable substates (orange arrows), the properties of the motor may be explored by assuming a range of values for these missing quantities and solving for the steady state probabilities and rates, as detailed in the Appendix A. (The Appendix also considers the consequences of backward photoisomerization from the metastable to the stable form, which is omitted here.) The present results correspond to experiments where the motor is continuously illuminated while at a temperature that allows THI,[18,54] so that the motor may rotate continuously.

We first assume that the photoisomerization process deposits the system into each of the four metastable substates in the same proportions as observed in the QM surface-hopping MD calculations (Section 3.1), *i.e.*, MS1 : MS2 : MS3 : MS4 = 7 : 2 : 4 : 0. Thus, with an overall photoisomerization rate constant of k_{photo}, the rate constant for arriving in MS1 is $\frac{7}{13}k_{\text{photo}}$, and so on. The steady state probabilities of the system's states are plotted as a function of k_{photo} in Fig. 9A. At low k_{photo}, the stable form has a probability near unity and the metastable substates have probabilities near zero. As k_{photo} rises above $\sim 10^5\ \text{s}^{-1}$, the probability begins to shift from the stable form to the various metastable substates; and a plateau is approached at $k_{\text{photo}} \approx 10^8\ \text{s}^{-1}$, where the stable form probability nears zero.

The motor's rotation rate (Fig. 9B, orange curve) has a similarly sigmoid dependence on k_{photo}. The asymptotic maximal rotation rate, 1.5 MHz, is approached as the stable form spends a negligible amount of time waiting for photoisomerization, so that

Published on 19 March 2021. Downloaded by University of California - San Diego on 2/12/2022 6:45:52 AM.

THI is rate-limiting. The k_{photo} at which the rate is 50% of its asymptotic maximum, is $k_{photo,50} = 1 \times 10^6\ s^{-1}$.

Interestingly, the maximal rotation rate (Fig. 9B) does not depend dramatically on the ratio of metastable substates immediately following photoisomerization. Assuming that photoisomerization generates only MS1 (MS1 : MS2 : MS3 : MS4 = 1 : 0 : 0 : 0) increases maximal rotation rate by about 10% (Fig. 9B, blue), because MS1 is close to completing THI (Fig. 7). Conversely, assuming that photoisomerization generates only MS2 (MS1 : MS2 : MS3 : MS4 = 0 : 1 : 0 : 0) reduces maximal rotation rate by about 15%, because MS2 must go through a relatively slow conversion to MS3 before going to MS1 and then to stable form (Fig. 7). The insensitivity of the rotation rate to the details of photoisomerization supports the approach taken in this study, which focuses on THI process with less comprehensive modeling of photoisomerization. However, the rotation rates of other light-driven motors may not be so insensitive to the distribution of metastable substates generated by photoisomerization. For example, if a motor had the same transition network and time constants as in Fig. 7 except for a much longer time constant to go from MS1 to MS3, its rotation rate would be more sensitive to the distribution of metastable substates generated by photoisomerization, as shown in Fig. 16 of Appendix A.

It is worth clarifying that, although the photoisomerization process starting from the Franck–Condon point is very fast (on the picosecond timescale), the photoisomerization rate constant is determined by a longer timescale, *i.e.*, the wait time for photoexcitation. This wait time is determined not just by the intrinsic properties of the motor, but also by experimental settings, such as the intensity and spectrum of the light. Indeed, approaching the maximum rotation rate in this figure may require light intensities too high to be readily generated in experiments, as previously emphasized.[54] The need for such high intensities traces to the low molar absorptivities of these compounds, about 10 000 $molar^{-1}\ cm^{-1}$.[16]

3.3.4 THI rate constant from weighted ensemble MD simulations. THI is often the rate-limiting rotation step of the alkene motors, so its rate constant is a critical performance metric. Although some of the classical MD simulations discussed above led to THI transitions, more would be needed for a precise estimate of the THI rate constant. We therefore used weighted ensemble simulations in the software package WESTPA to sample THI transitions.

For WESTPA to facilitate the key conformational transitions, progress coordinates must be defined that clearly distinguish the slowly interconverting states. We chose two-dimensional progress coordinates – a rotor–stator rotation angle (17-18-2-1) and the stator pucker – that clearly distinguish among the slowly interconverting metastable substates MS1, MS2, and MS3/MS4, as well as the stable form, as shown in Fig. 10. WESTPA calculations were run for both Motor S and Motor N, using the same starting points as their respective plain classical MD simulations (Section 3.3.1). As shown in Fig. 11 and Fig. 12, the final computed THI rate constants are within an order of magnitude of the corresponding experimental results, as well as the modeling results in Section 3.3.3.

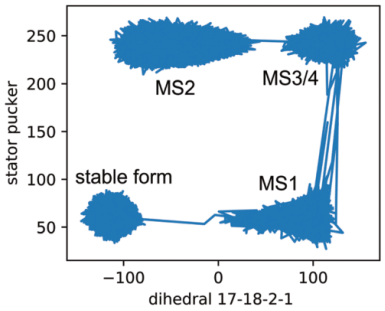

Fig. 10 Projection of one of the 13 plain MD simulations onto the two progress coordinates used in weighted ensemble simulation of Motor S. Clusters are labelled according to their conformational states (Fig. 7).

In experimental studies, increasing the solvent viscosity significantly decreased the THI rate constant.[55] Our simulations reproduced this trend (Fig. 13) by comparing THI rate constants from simulations using different Langevin collision frequencies, as the largest collision frequency, 10 ps^{-1}, which simulates the largest effective viscosity, resulted in a smaller THI rate constant than 1 ps^{-1} and 0.1 ps^{-1}. The smallest collision frequency, 0.01 ps^{-1}, also appears to decrease the THI rate, in agreement with a prior study about how collision frequency influences conformational sampling of a peptide.[56] Overall, however, the computed THI rate constant is not very sensitive to the somewhat arbitrary choice of collision frequency.

3.3.5 Free energy available from THI. The results presented above focus on the mechanisms and kinetics of the motor rotation. Here, we present results for the free energy difference between the metastable form and the stable form, and discuss the implications of this quantity as a performance metric. We

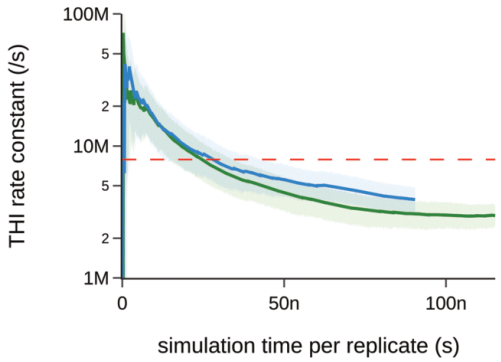

Fig. 11 THI rate constant of Motor S computed with WESTPA, as a function of simulation time per replicate, with shading to indicate 95% bootstrapping confidence interval. Results are shown for two runs with identical settings except different random number seeds. Dashed red line indicates experimental rate constant at 300 K, $7.9 \times 10^9\ s^{-1}$, read from Arrhenius plots of Klok *et al.*[16]

Published on 19 March 2021. Downloaded by University of California - San Diego on 2/12/2022 6:45:52 AM.

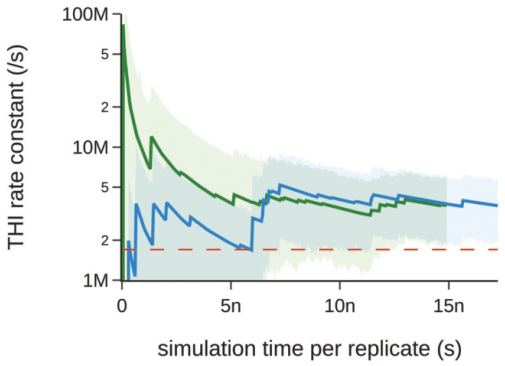

Fig. 12 THI rate constant of Motor N computed with WESTPA, as a function of simulation time per replicate, with shading to indicate 95% bootstrapping confidence intervals. Results are shown for two runs with identical settings except different random number seeds. Dashed red line indicates experimental rate constant at 300 K, 1.7×10^7 s^{-1}, from ref. 12.

used umbrella sampling to obtain the PMF for rotation around the central double bond, defined here by torsion 43-15-14-29. As shown in Fig. 14 for Motor S, the metastable form is separated from the stable form by a free energy barrier. Integration along the PMF (Section 2.2) gives a favorable free energy release of 7.8 kcal mol^{-1} on going from the metastable to the stable form (right to left in the figure). Note that the height and width of the barrier makes this free energy difference insensitive to the precise choice of boundary between the stable form's free energy well on the left and the metastable form's free energy well on the right. The corresponding PMF for Motor N is similar in shape (Fig. 15), and the free energy difference is slightly greater, at 8.8 kcal mol^{-1}. Thus, the available free energy from THI is similar to the standard free energy released by hydrolysis of ATP, the cell's energy currency, ~8 kcal mol^{-1}[57].

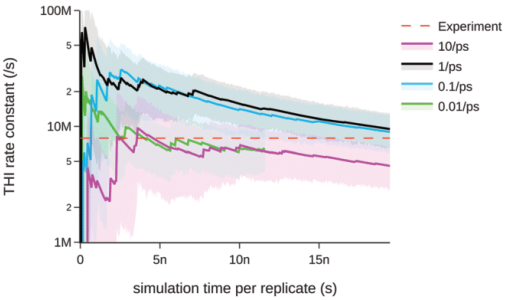

Fig. 13 THI rate constants of Motor S computed with WESTPA, as a function of simulation time per replicate, for four different collision frequences. Shaded areas indicate 95% bootstrapping confidence intervals. Dashed red line indicates experimental rate constant.[16]

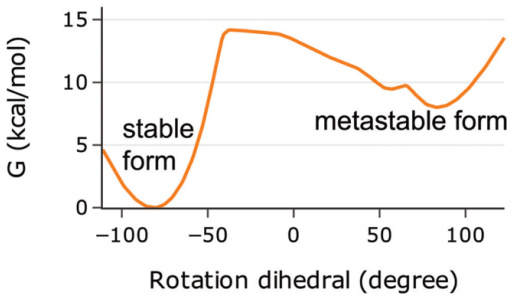

Fig. 14 Computed Gibbs free energy as a function of rotation dihedral (43-15-14-29) for Motor S.

The torque available from THI, averaged over the 180° rotation cycle of the motor, can be estimated as the free energy change, 7.8 kcal mol^{-1}, divided by π radians, which gives a torque of about 18 pN nm. This may be compared with 40 pN nm for the highly evolved F1-ATPase motor.[58] Note that this analysis omits any contribution from the photoisomerization step, which also runs in the forward direction and could in principle contribute more to the average torque.

We tested the numerical reliability of the free energy evaluations by rerunning all the simulations with different random number seeds, and obtained free energy differences within 0.3 kcal mol^{-1} of the original results. In addition, rerunning with umbrella sampling spring constant of 100 instead of 200 kj mol^{-1} affected the result by less than 0.3 kcal mol^{-1}, whereas varying WHAM parameters and the integral bounds within reason gave even smaller changes. Thus, the numerical uncertainties in these results appear to be well below 1 kcal mol^{-1}, and we expect that the force field is a larger source of error than the numerical uncertainty. Furthermore, these force field-based free energy differences are close to those previously obtained by purely QM methods for a similar motor.[15]

The maximal power available to drive an external load is less than the product of the maximal rotational rate and the free energy change, because loading the motor decreases its

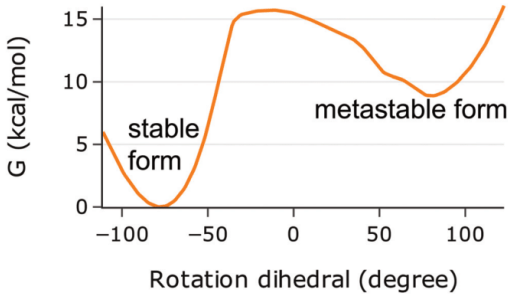

Fig. 15 Computed Gibbs free energy as a function of rotation dihedral (43-15-14-29) for motor N.

Published on 19 March 2021. Downloaded by University of California - San Diego on 2/12/2022 6:45:52 AM.

rotation rate. To get a sense for the power available from Motor S, we used WESTPA to recalculate the THI rate constant in the presence of a constant external torque along torsion angle 43-15-14-29. The external torque opposes forward rotation, with an energy difference (load) of 3.4 kcal mol^{-1} between the metastable form and the stable form. Adding this load dropped the THI rate constant to 1×10^5 s^{-1} (Fig. S2 in the ESI†), about an order of magnitude below the unloaded motor (Fig. 11). This loaded THI rate constant yields a power output of about 4×10^5 kcal mol^{-1} s^{-1}, when not bottlenecked by photoisomerization.[54] This potential power output is higher than typical biochemical reactions, because even fast enzymes, such as ATP synthase[59] (1×10^3 s^{-1}) and catalase[60] (1×10^4 s^{-1}), cycle much more slowly.

It is worth emphasizing that the $\sim$8 kcal mol^{-1} of free energy available from THI is only a fraction of the 80 kcal mol^{-1} energy provided by a photon from the 355 nm laser used in the experiments.[16] As a consequence, it is of interest to consider how to improve the energy efficiency. One possibility may be to modify the structure in a manner that makes the metastable form less stable, without degrading other key parameters. This would probably bring the additional benefit of lowering the THI barrier and thus increasing the rotation rate. It might also be possible to collect power not only from THI but also from the photoisomerization step, as noted above.

A previous study[55] obtained a far lower estimate of the work available from THI of Motor S. Using structures from QM calculations, the authors found that the stable form has a solvent-excluded volume 7 $Å^3$ larger than that of the metastable form, and they computed the pressure–volume (PV) expansion work associated with this expansion, which comes to only 0.006 kcal mol^{-1}. We obtained a similar volume change, 5 $Å^3$, for this step, by examining the time-series of system volumes in our explicit solvent simulations run at room pressure. However, it is the rotary work available from these motors, not the PV work that should be of interest. This may be understood by considering that capturing work done by the motor involves coupling an external load to the angle between the stator and rotor, not to a change in volume. More broadly, it is also worth noting that the Gibbs free energy change is the maximal non-PV work in a constant-temperature, constant-pressure, closed system.

4 Conclusions

We have studied a class of fast, light-driven overcrowded alkene-based molecular motors with a novel multiscale computational framework that combines quantum mechanical (QM) surface-hopping molecular dynamics (MD) simulations with classical MD simulations. The main conclusions are as follows.

(1) Trajectories of the rotation process (Videos in the ESI†) give a rich picture of the full rotation dynamics of these motors.

(2) Kinetic models derived from simulation trajectories provides a detailed rotation mechanism (Fig. 7); notably,

• A flip of the stator pucker, generating metastable substate 1 (MS1), immediately precedes and appears to enable THI transition to the stable form.

• The maximal rotation speed can be calculated from the model and depends on the substate interconversion rates, as well as the distribution of metastable substates generated by the photoisomerization process.

(3) Experimentally measured THI rate constants are replicated to within an order of magnitude using weighted ensemble MD simulations with optimized force field parameters. Thus, this technology may be useful in designing new molecular motors with desired rotation rates.

(4) The maximal available work from THI is estimated as $\sim$8 kcal mol^{-1}, which corresponds to an average torque of $\sim$18 pN nm. Simulated under half load, the motor still has a high THI rate constant 1×10^5 s^{-1} and the maximal power output is about 4×10^5 kcal mol^{-1} s^{-1}. These magnitudes are similar to and sometimes higher than biological motors, such as the F1-ATPase.

(5) Classical MD simulations with a widely used general force field gave incorrect conformations and dynamics. Far better results were obtained after fitting selected force field parameters to QM potential energy surfaces.

(6) For Motor O, with its planar stator, the Lennard-Jones term appeared inadequate to model van der Waals interactions during THI, probably due to its unrealistically steep repulsion potential. The less steep and hence more realistic Buckingham potential provided better results.

The computational and theoretical framework used here is modular, so future studies could use refined forms of each component method. For example, photoisomerization could be modeled with higher level QM techniques, such as *ab initio* Multiple Spawning methods,[61] and the classical simulations could use improved functional forms of the force field, such as a more realistic van der Waals potential or a more detailed description of torsional energy profiles. The present framework can also be extended to other classes of fast light-driven molecular motors, especially those that share the same overcrowded alkene core.[25,62]

Conflicts of interest

MKG has an equity interest in and is a cofounder and scientfic advisor of VeraChem LLC.

Appendix A: derivation of steady state rotation rate under continuous excitation

When the motor molecules are simultaneously excited by light and heated, they will rotate cycle after cycle. A steady state can be reached under constant conditions, and here we derive the steady state probabilities of all the conformational states drawn in Fig. 7, and the corresponding probability flux, *i.e.*, the steady state rotation rate. These results were used to produce Fig. 9.

Published on 19 March 2021. Downloaded by University of California - San Diego on 2/12/2022 6:45:52 AM.

The MSM transition matrix associated with Fig. 7 can be viewed as a discretization of continuous-time differential equations of chemical kinetics, written below.[63]

$$\frac{dp_s}{dt} = k_{1s}p_1 - (k_{s1} + k_{s2} + k_{s3})p_s \quad (1)$$

$$\frac{dp_1}{dt} = k_{s1}p_s + k_{31}p_3 - k_{1s}p_1 - k_{13}p_1 \quad (2)$$

$$\frac{dp_2}{dt} = k_{s2}p_s + k_{32}p_3 - k_{23}p_2 \quad (3)$$

$$\frac{dp_3}{dt} = k_{s3}p_s + k_{23}p_2 + k_{13}p_1 - k_{32}p_3 - k_{31}p_3 \quad (4)$$

$$p_4 = p_3\frac{k_{34}}{k_{43}} \quad (5)$$

The last equation posits rapid equilibrium between MS3 and MS4, based on the much faster interconversion time between them. Here, p indicates a probability, an s subscript indicates the stable form and an integer subscript indicates the corresponding metastable substate, and the rate constant from state a to b is denoted k_{ab}. Setting the derivatives to zero and solving these equations with the constraint $p_1 + p_2 + p_3 + p_4 + p_s = 1$ gives steady state probabilities π_1, π_2, π_3, π_4, π_s as functions of the rate constants. Then the steady state rotation rate (flux) equals $k_{1s}\pi_1$. With good MSM discretization, each rate constant in the MSM should match the corresponding rate constant in these differential equations, so we borrow the rate constants from Fig. 7 to use here. Scanning over a range of photoisomerization rate constants and considering different distributions of metastable substates generated by photoisomerization produce Fig. 9, as detailed in Section 3.3.3.

We anticipated that the motor rotation rate would become more sensitive to the distribution of metastable substates generated by photoisomerization if k_{13} was smaller. To test this, we reduced k_{13} by a factor of 10 000, while leaving other rate constants unchanged. As shown in Fig. 16, the rotation rate now is, in fact, highly sensitive to the distribution of metastable substates generated by photoisomerization. Intuitively, the drop in k_{13} means that virtually all trajectories initiated in MS1 will quickly undergo THI, rather than being delayed by going to MS3. A Jupyter notebook in our Github repository contains the source code for this calculation and plotting.

In the above analysis, we have omitted backward photoisomerization, *i.e.*, photoisomerization from B to A or from D to C (Fig. 1), because including this would have required making additional arbitrary assumptions about the rate constants for backward photoisomerization from each of the metastable substates. This omission should not be problematic when the illumination is weak enough and/or the temperature is high enough, so that the backward photoisomerization rate is small compared to the forward THI rate. However, in settings where backward photoisomerization is substantial relative to forward THI, we expect the model given by the above equations to underestimate the steady state probabilities of the stable form,

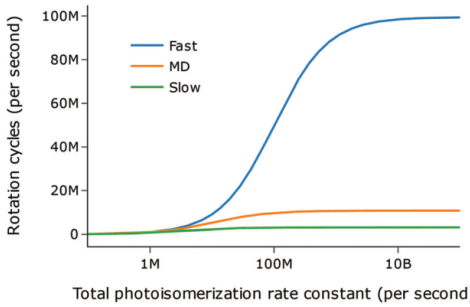

Fig. 16 Dependence of the motor rotation rate in steady state on the total photoisomerization rate, assuming 10 000 times smaller k_{13} than its value obtained from MSM and used in Fig. 9. The three curves correspond to three different distributions of metastable substates generated by photoisomerization: the distribution that leads to the fastest rotation (MS1 : MS2 : MS3 : MS4 = 1 : 0 : 0 : 0), the distribution that leads to the slowest rotation (MS1 : MS2 : MS3 : MS4 = 0 : 1 : 0 : 0), and the distribution inferred from our surface-hopping MD trajectories (MS1 : MS2 : MS3 : MS4 = 7 : 2 : 4 : 0).

and overestimate the rotation rate. This is because backward photoisomerization provides a nonproductive channel by which probability can drain away from the metastable form instead of going through forward THI. Although the quantitative details will depend on the backward photoisomerization rate constants from each of the metastable substates, the maximal rotation rate approached asymptotically with increasing levels of illumination should be approximately the maximal rotation rate from the present model multiplied by the probability of the metastable form in the photostationary state, which is usually on the order of 20% for the similar Motor O.[55] A prior study[54] provides valuable detail on how backward photoisomerization affects the steady state rotation rate, but without resolving the metastable form into substates.

Appendix B: selection of reaction coordinates

The rotation of the rotor relative to the stator is the defining feature of these motors. The most common way to describe this rotation is a torsion angle defined with 4 consecutive atoms around the central double bond, such as 16-15-14-13. However, we found that such dihedrals are insensitive to the ground state conformational changes, as they remain rather close to either 0° or 180°. Instead, non-sequential dihedrals involving two rotor and two stator atoms, such as 18-15-14-10 and especially 43-15-14-29 or 17-18-2-1, show larger and more definitive changes as the rotor naphthalene atoms snap past the stator benzene atoms to complete THI. The Cremer-Pople definition of the stator pucker[64] provides a valuable second reaction coordinate.

Upon excitation, the central double bond becomes more single bond-like, leading to carbon pyramidalization; in contrast, during ground state simulations we found the degree of pyramidalization to be small. Consequently, pyramidalization coordinate is only

Published on 19 March 2021. Downloaded by University of California - San Diego on 2/12/2022 6:45:52 AM.

useful for characterizing photoisomerization or the whole rotation cycle including photoisomerization. Although the rotor pucker was a useful reaction coordinate for the QM potential energy surfaces (Fig. 4), it was a less informative coordinate than the stator pucker when focusing on THI, except for Motor O, whose stator is planar. One reason is evident in Fig. 7, which shows that interconversion between different rotor puckers (MS3 and MS4) is much faster than interconversion between different stator puckers.

It is worth noting that a reaction coordinate based on an atom that is more rigidly linked to the rest of the molecule, such as an sp^2 carbon, is generally better than one based on a more mobile atom, such as an sp^3 carbon, at a similar location. This is because rigidly linked atoms tend to give smaller intra-state fluctuations of the reaction coordinate values, making the inter-state differences in reaction coordinate values more prominent. Similarly, uninteresting fluctuations in a reaction coordinate may be suppresed by defining it based on the center of mass of several atoms, instead of on one atom.

Appendix C: atom typing the central double bond

According to the atom typing rules of GAFF,[44] carbon 14 could logically be assigned atom type cc, an ''inner sp2 carbon in conjugated ring systems'', or atom type ce, an ''inner sp2 carbon in conjugated chain systems''. The parameter assignment program used in this study, Antechamber, would assign cc to all motors considered here. This may be appropriate for Motor O, with its planar stator, but the ring to which carbon 14 belongs is not conjugated for motors S and N, so type ce seems more appropriate for these two molecules. Indeed, for Motor S and Motor N, using type ce yielded a slightly better value of the objective function in the force field optimization procedures discussed above. In addition, the THI rate constants we computed using ce were generally better than those computed with type cc. Therefore we used ce for motors S and N, with their non-planar stators, and cc for Motor O, with its planar stator.

Acknowledgements

We thank Drs David Slochower, Jeffry Setiadi, Lee-Ping Wang, and Adrian Roitberg for helpful discussions, and Drs Jie Liu, Axel Koslowski and the late Walter Thiel for their help with MNDO. MKG acknowledges funding from National Institute of General Medical Sciences (GM061300). These findings are solely of the authors and do not necessarily represent the views of the NIH.

Notes and references

1 S. Kassem, T. van Leeuwen, A. S. Lubbe, M. R. Wilson, B. L. Feringa and D. A. Leigh, *Chem. Soc. Rev.*, 2017, **46**, 2592–2621.
2 A. B. Kolomeisky, *J. Phys.: Condens. Matter*, 2013, **25**, 463101.
3 D. R. Slochower and M. K. Gilson, *Biophys. J.*, 2018, **114**, 2174–2179.
4 M. Baroncini, S. Silvi and A. Credi, *Chem. Rev.*, 2019, **120**(1), 200–268.
5 S. Erbas-Cakmak, D. A. Leigh, C. T. McTernan and A. L. Nussbaumer, *Chem. Rev.*, 2015, **115**, 10081–10206.
6 B. L. Feringa, *J. Org. Chem.*, 2007, **72**, 6635–6652.
7 B. L. Feringa, *Angew. Chem., Int. Ed.*, 2017, **56**, 11060–11078.
8 V. García-López, D. Liu and J. M. Tour, *Chem. Rev.*, 2020, **120**, 79–124.
9 J. Chen, J. C. M. Kistemaker, J. Robertus and B. L. Feringa, *J. Am. Chem. Soc.*, 2014, **136**, 14924–14932.
10 R. Eelkema, M. M. Pollard, J. Vicario, N. Katsonis, B. S. Ramon, C. W. M. Bastiaansen, D. J. Broer and B. L. Feringa, *Nature*, 2006, **440**, 163.
11 T. Kudernac, N. Ruangsupapichat, M. Parschau, B. Maciá, N. Katsonis, S. R. Harutyunyan, K.-H. Ernst and B. L. Feringa, *Nature*, 2011, **479**, 208–211.
12 A. A. Kulago, E. M. Mes, M. Klok, A. Meetsma, A. M. Brouwer and B. L. Feringa, *J. Org. Chem.*, 2010, **75**, 666–679.
13 A. Cnossen, J. C. M. Kistemaker, T. Kojima and B. L. Feringa, *J. Org. Chem.*, 2014, **79**, 927–935.
14 B. Oruganti, C. Fang and B. Durbeej, *Phys. Chem. Chem. Phys.*, 2015, **17**, 21740–21751.
15 B. Oruganti and B. Durbeej, *J. Mol. Model.*, 2016, **22**, 219.
16 M. Klok, N. Boyle, M. T. Pryce, A. Meetsma, W. R. Browne and B. L. Feringa, *J. Am. Chem. Soc.*, 2008, **130**, 10484–10485.
17 R. D. Astumian, *Faraday Discuss.*, 2016, **195**, 583–597.
18 E. M. Geertsema, S. J. van der Molen, M. Martens and B. L. Feringa, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 16919–16924.
19 B. Oruganti, J. Wang and B. Durbeej, *Int. J. Quantum Chem.*, 2018, **118**(1), e25405.
20 F. Liu and K. Morokuma, *J. Am. Chem. Soc.*, 2012, **134**, 4864–4876.
21 A. Kazaryan, J. C. M. Kistemaker, L. V. Schäfer, W. R. Browne, B. L. Feringa and M. Filatov, *J. Phys. Chem. A*, 2010, **114**, 5058–5067.
22 A. Kazaryan, Z. Lan, L. V. Schäfer, W. Thiel and M. Filatov, *J. Chem. Theory Comput.*, 2011, **7**, 2189–2199.
23 A. Nikiforov, J. A. Gamez, W. Thiel and M. Filatov, *J. Phys. Chem. Lett.*, 2016, **7**, 105–110.
24 X. Pang, X. Cui, D. Hu, C. Jiang, D. Zhao, Z. Lan and F. Li, *J. Phys. Chem. A*, 2017, **121**, 1240–1249.
25 J. Vachon, G. T. Carroll, M. M. Pollard, E. M. Mes, A. M. Brouwer and B. L. Feringa, *Photochem. Photobiol. Sci.*, 2014, **13**, 241–246.
26 P. O. Dral, X. Wu, L. Spörkel, A. Koslowski, W. Weber, R. Steiger, M. Scholten and W. Thiel, *J. Chem. Theory Comput.*, 2016, **12**, 1082–1096.
27 D. Tuna, Y. Lu, A. Koslowski and W. Thiel, *J. Chem. Theory Comput.*, 2016, **12**, 4400–4422.
28 W. Thiel, *MNDO99 program*.
29 J. C. Tully, *J. Chem. Phys.*, 1990, **93**, 1061–1071.
30 E. Braun, J. Gilmer, H. Mayes, D. Mobley, S. Prasad, D. Zuckerman and J. Monroe, *J. Comput. Mol. Sci.*, 2018, **1**, 5957.

Published on 19 March 2021. Downloaded by University of California - San Diego on 2/12/2022 6:45:52 AM.

31 L. Martínez, R. Andrade, E. G. Birgin and J. M. Martínez, *J. Comput. Chem.*, 2009, **30**, 2157–2164.
32 C. Wehmeyer, M. K. Scherer, T. Hempel, B. E. Husic, S. Olsson and F. Noé, *J. Comput. Mol. Sci.*, 2019, **1**, 5965.
33 B. Mostofian and D. M. Zuckerman, *J. Chem. Theory Comput.*, 2019, **15**(6), 3499–3509.
34 M. C. Zwier, J. L. Adelman, J. W. Kaus, A. J. Pratt, K. F. Wong, N. B. Rego, E. Suárez, S. Lettieri, D. W. Wang, M. Grabe, D. M. Zuckerman and L. T. Chong, *J. Chem. Theory Comput.*, 2015, **11**, 800–809.
35 J. Kästner, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2011, **1**, 932–942.
36 M. Bonomi, G. Bussi, C. Camilloni, G. A. Tribello, P. Banáš, A. Barducci, M. Bernetti, P. G. Bolhuis, S. Bottaro and D. Branduardi, *Nat. Methods*, 2019, **16**, 670–673.
37 G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Comput. Phys. Commun.*, 2014, **185**, 604–613.
38 A. Grossfield, WHAM: the weighted histogram analysis method, version 2.0.10, http://membrane.urmc.rochester.edu/wordpress/?page_id=126.
39 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 1011–1021.
40 J. Řezáč, D. Bím, O. Gutten and L. Rulíšek, *J. Chem. Theory Comput.*, 2018, **14**, 1254–1266.
41 Y. Qiu, D. G. A. Smith, C. D. Stern, M. Feng, H. Jang and L.-P. Wang, *J. Chem. Phys.*, 2020, **152**, 244116.
42 R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonett, A. E. DePrince, E. G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. Di Remigio, R. M. Richard, J. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, M. Saitow, X. Wang, B. P. Pritchard, P. Verma, H. F. Schaefer, K. Patkowski, R. A. King, E. F. Valeev, F. A. Evangelista, J. M. Turney, T. D. Crawford and C. D. Sherrill, *J. Chem. Theory Comput.*, 2017, **13**, 3185–3197.
43 L.-P. Wang, T. J. Martinez and V. S. Pande, *J. Phys. Chem. Lett.*, 2014, **5**, 1885–1891.
44 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.
45 X. He, V. H. Man, W. Yang, T.-S. Lee and J. Wang, *The Journal of Chemical Physics*, 2020, **153**, 114502.
46 D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. Ghoreishi and M. K. Gilson, *AMBER 18*, University of California, San Francisco.
47 A. Jakalian, B. L. Bush, D. B. Jack and C. I. Bayly, *J. Comput. Chem.*, 2000, **21**, 132–146.
48 C. Product, *CRC Handbook of Chemistry and Physics. 92*, Taylor and Francis Group, 2012.
49 K. Iwata and H.-o. Hamaguchi, *J. Phys. Chem. A*, 1997, **101**, 632–637.
50 Y. Amatatsu, *J. Phys. Chem. A*, 2013, **117**, 3689–3696.
51 J. Medina, F. Avilés and A. Tapia, *Mol. Phys.*, 2015, **113**, 1297–1305.
52 F. Jensen, *Introduction to computational chemistry*, 3rd edn, John Wiley & Sons, Chichester, UK; Hoboken, NJ, 2017.
53 E. M. Engler, J. D. Andose and P. V. R. Schleyer, *J. Am. Chem. Soc.*, 1973, **95**, 8005–8025.
54 M. Klok, W. R. Browne and B. L. Feringa, *Phys. Chem. Chem. Phys.*, 2009, **11**, 9124.
55 Klok, B. L. Feringa and R. A. L. Jones, *Faraday Discuss.*, 2009, **143**(9), 319–334.
56 R. J. Loncharich, B. R. Brooks and R. W. Pastor, *Biopolymers*, 1992, **32**, 523–535.
57 R. A. Alberty and R. N. Goldberg, *Biochemistry*, 1992, **31**, 10610–10615.
58 K. Kinosita, R. Yasuda, H. Noji, S. Ishiwata and M. Yoshida, *Cell*, 1998, **93**, 21–24.
59 R. Iino, R. Hasegawa, K. V. Tabata and H. Noji, *J. Biol. Chem.*, 2009, **284**, 17457–17464.
60 C. Riedel, R. Gabizon, C. A. M. Wilson, K. Hamadani, K. Tsekouras, S. Marqusee, S. Pressé and C. Bustamante, *Nature*, 2014, **517**, 227–230.
61 B. F. E. Curchod and T. J. Martínez, *Chem. Rev.*, 2018, **118**, 3305–3336.
62 J. C. M. Kistemaker, P. Štacko, D. Roke, A. T. Wolters, G. H. Heideman, M.-C. Chang, P. van der Meulen, J. Visser, E. Otten and B. L. Feringa, *J. Am. Chem. Soc.*, 2017, **139**, 9650–9661.
63 D. M. Zuckerman, *Physical Lens on the Cell*, 2019.
64 D. Cremer and J. A. Pople, *J. Am. Chem. Soc.*, 1975, **97**, 1354–1358.

Acknowledgements

Chapter 3, in full, is a reprint of the material as it appears in “Feng M, Gilson MK. Mechanistic analysis of light-driven overcrowded alkene-based molecular motors by multiscale molecular simulations. Phys Chem Chem Phys. 2021”. The dissertation author was the primary author of this publication.