THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Perspective

The Next Frontier for Designing Switchable Proteins: Rational Enhancement of Kinetics

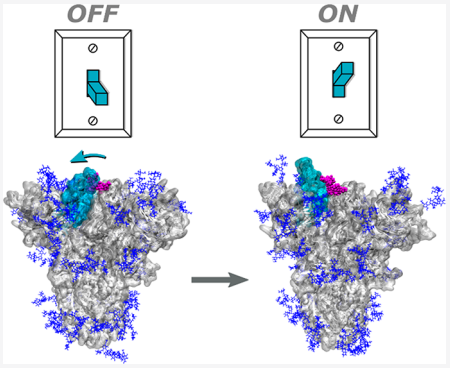
Anthony T. Bogetti, Maria F. Presti, Stewart N. Loh,* and Lillian T. Chong*

Cite This: *J. Phys. Chem. B* 2021, 125, 9069–9077

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Designing proteins that can switch between active (ON) and inactive (OFF) conformations in response to signals such as ligand binding and incident light has been a tantalizing endeavor in protein engineering for over a decade. While such designs have yielded novel biosensors, therapeutic agents, and smart biomaterials, the response times (times for switching ON and OFF) of many switches have been too slow to be of practical use. Among the defining properties of such switches, the kinetics of switching has been the most challenging to optimize. This is largely due to the difficulty of characterizing the structures of transient states, which are required for manipulating the height of the effective free energy barrier between the ON and OFF states. We share our perspective of the most promising new experimental and computational strategies over the past several years for tackling this next frontier for designing switchable proteins.



Downloaded via UNIV OF PITTSBURGH on September 22, 2021 at 21:17:27 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

INTRODUCTION

Protein conformational switches—proteins that adopt either active (ON) or inactive (OFF) conformations in response to signals such as ligand binding and incident light—have been exploited as the core machinery behind novel biosensors, therapeutic agents, and "smart" biomaterials.[1,2] The fundamental characteristics of a switch include its signal-to-noise ratio (the extent to which the switch converts between ON and OFF states), sensitivity (what levels of effector are required for activation), and response time (the time required for the switch to turn on and off). Signal-to-noise in biological switches can be a complex phenomenon that is sometimes modulated by agonists/antagonists that induce partial or alternate ON/OFF states, and improving signal-to-noise is an active subfield of its own in switch design. The response time has proven to be even more challenging to optimize. Most design strategies focus on stable states, specifically, the ON and OFF conformations. Switching mechanisms can consist of introducing a second stable state in a monomeric protein, creating new protein−protein or protein−ligand binding interfaces, and fusing protein domains such that they achieve input−output communication. These efforts are typically guided by structures of existing proteins or, more recently, by principles of de novo design.[3,4] Either way, they seek to define the structures and optimize the activities of the stable ON/OFF states of the protein. In general, it is left to chance that the stable states interconvert with reasonable rates.

The above scenario is illustrated by the free energy diagrams of Figure 1 by using the example of a protein biosensor into which two stable conformations have been engineered (represented by OFF and ON states with the latter binding the target ligand; Figure 1A). Some of the basic properties of the switch, for example, signal-to-noise (turn-on/turn-off ratio) and limit of detection, can be optimized by adjusting the relative thermodynamic stabilities of OFF and ON conformations and the ligand binding affinity (K_d) of the ON state, respectively (Figure 1A). Because the structures of OFF and ON states are typically known, these goals can be achieved by using well-established experimental and theoretical approaches. The response time (given by $[k_{ON} + k_{OFF}]^{-1}$) is proportional to the height of the transition state ensemble (TSE) between OFF and ON. Accelerating the response time can be accomplished by introducing interactions that stabilize the TSE but not the ground states (Figure 1B) or, more commonly, by deleting native interactions that are present in the ground states but absent in the TSE (Figure 1C). In either case the TSE must be characterized by experimental and/or computational means.

Here, we examine recent advances in tackling what we regard as the next frontier in the design of switchable proteins: the rational tuning of kinetics (i.e., turn-on and turn-off rates). Oftentimes this means making a switch cycle between ON and OFF sites more rapidly, so that it can react to conditions that

Received: May 7, 2021
Revised: July 5, 2021
Published: July 29, 2021

© 2021 American Chemical Society

https://doi.org/10.1021/acs.jpcb.1c04082
J. Phys. Chem. B 2021, 125, 9069–9077

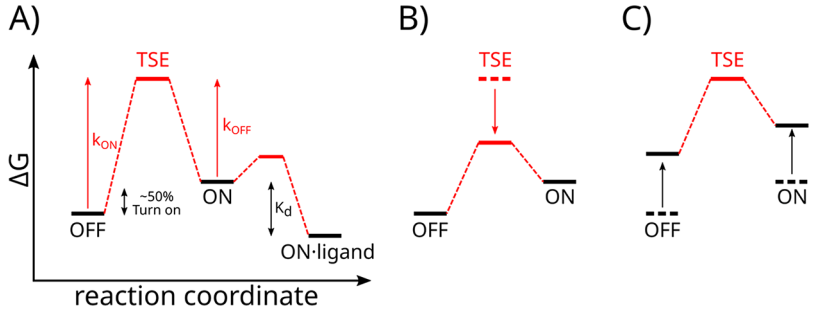


Figure 1. Manipulating barrier heights to accelerate switching rates. (A) Free energy diagram of a protein biosensor with stable OFF and ON states, showing the transition path between them (red). Optimizing the equilibrium properties of the switch (turn-on/turn-off ratio, limit of detection) can be achieved by introducing mutations that shift the relative thermodynamic stabilities of OFF and ON conformations and alter the affinity of the ON state for the target ligand. (B) As in a classic enzyme mechanism, conformational switching can be accelerated by stabilizing the TSE. (C) In practice, it is often more tractable to lower the TSE barrier by destabilizing the ON and OFF folds by introducing mutations that delete interactions that are present in the stable states and absent in the TSE.

change over a wide range of time scales. In other cases, the goal is to make the switch respond more slowly. For instance, decreasing the turn-off rate is useful for enhancing the sensitivity of biosensors because it enables the ON signal to accumulate and for activating optogenetic tools because it allows for a durable biological response that persists well after light is removed, with reduction of photodamage and photobleaching. For the purpose of this review we assume that faster kinetics/lower barrier heights are intended, although the same principles apply if one desires the opposite effect. The main goal is to be able to optimize response times to match that of a given biological process or practical application.

This frontier is a particularly challenging one, requiring the analysis of transient states that experiments typically cannot capture. While these transient states are ideally generated as part of complete, atomically detailed pathways of the switching process from molecular dynamics simulations, such simulations have not been feasible due to the long time scales of switching processes (>milliseconds). Of particular interest is therefore the synergistic use of experimental techniques and computational strategies that can enable the generation of detailed structures of transient states for the design of more responsive switchable proteins. We also comment on promising future directions.

We define a protein conformational switch as one in which input and output functionalities are integrated into a single molecule, often by means of fusing receptor and reporter domains in such a way as to facilitate allosteric, interdomain conformational changes. Switch designs can be classified into three broad categories. The first uses an input domain that has naturally evolved the ability to switch between two stable conformations in response to a stimulus. In the second design, the input domain is an existing protein that has but a single fold, with conformational change being achieved via a folding/unfolding reaction that is linked to ligand binding. The third category involves similar mechanisms but employs de novo design principles to generate sequences and structures that may not have existed previously. In each case, the designer is faced with the challenge of converting the binding interaction to an observable signal by means of coupling the input response to an output response. Below we discuss the kinetic barriers that are present in some examples of each category and the experimental approaches that have been used to identify and modulate these barriers.

EXPERIMENTAL APPROACHES FOR CHARACTERIZING TRANSITION STATES

Central to both experimental and computational approaches for tuning rates is to first identify and characterize TSEs between switch conformations. Depending on the type of switch, as discussed below, these transitions can vary in extent from whole-molecule folding/unfolding to rigid-body domain movement to localized structural rearrangements. The challenge facing experimental methods is to observe sparsely populated states that approximate the relevant TSEs. A protein engineering method known as φ-value analysis[5] has been used to map the TSEs of global folding/unfolding. This technique entails introducing a point mutation into a protein and measuring the extent to which it changes the equilibrium constant between native and unfolded states versus the rates of folding and/or unfolding. φ-analysis has been applied to many proteins, and some guiding principles have emerged.[6] NMR- and MS-based methods have been employed to probe more subtle conformational changes, often providing per-residue resolution and rates of interconversion. Recent examples include using ZZ-exchange NMR spectroscopy to measure site-specific folding rates of protein L9,[7] amide hydrogen/deuterium exchange to characterize conformational dynamics of dopamine[8] and XylE membrane-bound transporters,[9] and NMR relaxation dispersion to uncover hidden states and their rates of interconversion in glycotransferase fold switching[10] and dihydrofolate reductase enzymatic function.[11]

BINDING-INDUCED FOLDING SWITCHES

All native proteins can be made to unfold, and many proteins (including some that are disordered) recognize ligands with high affinity and specificity when they are folded. These features make binding-induced folding a generalizable platform for biosensor engineering. One such example is the alternate frame folding (AFF) design, which was used to convert the small calcium binding protein calbindin D_{9k} into the fluorescent calcium sensor, calbindin-AFF.[12] The AFF modification entailed duplicating the N-terminal EF-hand of calbindin (that contained a calcium-binding residue; cyan in Figure 2A) and fusing it to the protein's C-terminus (magenta). Joining the two polypeptides with a linker long enough to span the N-to-C distance of calbindin allowed calbindin-AFF to switch between two folding "frames", one of which corresponded to the original amino acid sequence (WT frame) and the other to that of a circular permutant (CP

https://doi.org/10.1021/acs.jpcb.1c04082
J. Phys. Chem. B 2021, 125, 9069−9077

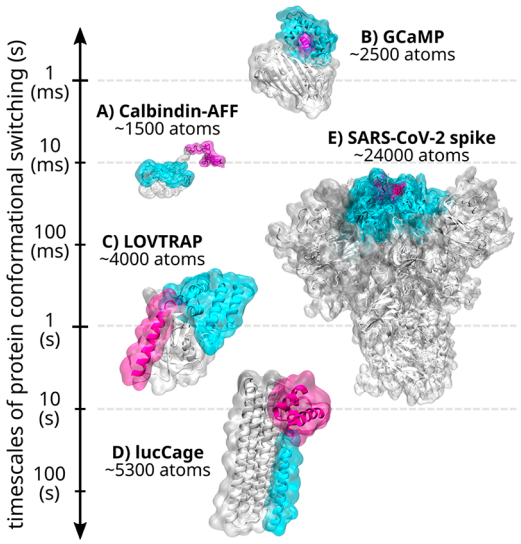


Figure 2. Protein conformational switches and their response times. (A) The calbindin-AFF construct (PDB ID for calbindin D_{9k}: 3ICB) switches via Ca^{2+}-driven unfolding/folding of two duplicate EF-hands (cyan and magenta) and their dissociation/docking with a shared region (gray). (B) The GCaMP calcium sensor (PDB ID: 3EK4) entails Ca^{2+}-induced binding of CaM (cyan) to the RS20 peptide (magenta), which protects the GFP (gray) chromophore from solvent and turns on fluorescence. (C) The LOVTRAP optogenetic construct (PDB ID: 5EFW) involves light-triggered dissociation of the Jα helix (magenta) from LOV2 (gray), resulting in dissociation of Zdark (cyan). (D) The lucCage biosensor (PDB ID: 7CBC) is composed of a cage (gray) and a latch (cyan), to which an analyte recognition domain (magenta) has been fused. Binding of the analyte together with a key (which resembles the latch; not shown) causes the latch to dissociate and expose a sequence in the latch that complements and activates a reporter enzyme. (E) The SARS-CoV-2 spike protein (PDB ID: 6VXX) involves opening of the receptor binding domain (cyan) from the core domain (gray), as gated by a glycan (magenta) attached to the N343 residue.

frame). The duplicate segments extend from the C-terminus and N-terminus of the WT and CP frames, respectively, as disordered peptides. Binding of calcium to one of the duplicate EF-hands induces it to fold and dock against the shared region of calbindin-AFF (gray), displacing and unfolding its counterpart. The switch was driven in either direction by mutating a calcium-binding residue in one or the other duplicate EF-hand, and the conformational change was reported by strategic placement of donor and quencher fluorophores. The turn-on and turn-off half times were in the 1−10 s range.

For AFF and other binding-induced folding switch designs, it is reasonable to anticipate that ON/OFF switching times may be accelerated by lowering the barriers to folding and unfolding. Typically, this is done by introducing mutations that raise the free energy of native or denatured states relative to that of the TSE by using the φ-value analysis approach. Nevertheless, identifying rate-enhancing mutations by experimental means remains largely a hit-or-miss prospect. Moreover, in the case of calbindin-AFF, the rate-limiting step appears to involve partial unfolding rather than global unfolding,[13] the former of which being more challenging to characterize by traditional φ-value analysis. Computational methods were invaluable to improving the response rate of calbindin-AFF (*vide infra*).

BIOSENSORS WITH PREEXISTING SWITCHABLE INPUT DOMAINS

In contrast to calbindin-AFF, the GCaMP family of genetically encoded calcium indicators (GECIs) employ an input domain (calmodulin, or CaM) that naturally evolved to undergo a dramatic conformational change upon calcium binding. In its calcium-free state, CaM's N-terminal EF-hand, C-terminal EF-hand, and connecting polypeptide adopt a compact, closed conformation. Binding of Ca^{2+} to the EF-hands induces a shift to an extended state that exposes the connecting helix for binding to many protein domains such as the RS20 peptide from myosin light chain kinase. To transduce this change to a fluorescent output, GCaMPs fuse RS20 (magenta in Figure 2B) and CaM (cyan) to the N- and C-termini of GFP (gray) that has been circularly permuted near its chromophore.[14] The interaction between CaM and RS20 protects the chromophore from solvent access, resulting in fluorescent turn-on.

GCaMPs and other GECIs have revolutionized studies of calcium signaling in vivo. To do so, it was necessary to shorten their response times to match those of rapid fluctuations in cellular calcium concentration (1−100 ms). The GCaMP response time is limited by its turn-off rate, which is determined not only by calcium release but also by the extended-to-closed conformational change of CaM that follows. Chemical intuition predicts that mutating residues in the CaM EF-hands (that weaken calcium binding) as well in RS20 (that weaken peptide binding) will accelerate the turn-off rate. Both predictions proved correct; turn-off rates were increased from 2.48 to 4.68 s^{-1},[15] 5.8 to 21 s^{-1},[16] and 4.62 to 99 s^{-1},[17] in various GCaMP GECIs. These results illustrate a central point of this review. Accelerated turn-off rates tended to correlate with higher K_d of the sensors, especially for the EF-hand mutants. This relationship arises because any mutation that weakens ligand binding affinity (RS20 can be considered a second ligand in the GCaMP switch) will likely raise the free energy of the ON state relative to that of the OFF state, thus changing the sensor's equilibrium properties (e.g., sensitivity). If one wishes to optimize response time without perturbing affinity, the mutation(s) should alter the free energies of the ground states relative to the TSE and not with respect to each other (Figure 1B). Rational selection of these mutation sites requires knowledge of the allosteric mechanism gained through experimental or computational means.

The class of photoactivatable proteins exemplified by the second light-oxygen-voltage-sensing domain 2 (LOV2) from *Avena sativa* phototropin 1 is another example of a bioswitch built from pre-existing allosteric domains. Blue light absorption triggers the ON state, in which a covalent bond forms between the flavin mononucleotide (FMN) chromophore and a conserved cysteine.[18] Cys adduct formation is coupled with the rotation of a conserved Gln with concomitant unfolding and dissociation of the N-terminal helix (A′α) and the C-terminal helix (Jα; magenta in Figure 2C) from the LOV2 core domain (gray).[19,20] When the blue light is removed, the photoadduct spontaneously breaks and LOV2 returns to its OFF state, with A′α and Jα folding and rebinding to the core domain. Covalent changes to the FMN chromophore occur on the microsecond time scale, and Jα unfolding proceeds on the millisecond time scale. The ON to OFF reversion, however, requires minutes to hours, making it the rate-limiting step in the photocycle. Random mutagenesis of 7 of the ∼20 amino acid sites that comprise the FMN binding pocket identified

https://doi.org/10.1021/acs.jpcb.1c04082
J. Phys. Chem. B 2021, 125, 9069−9077

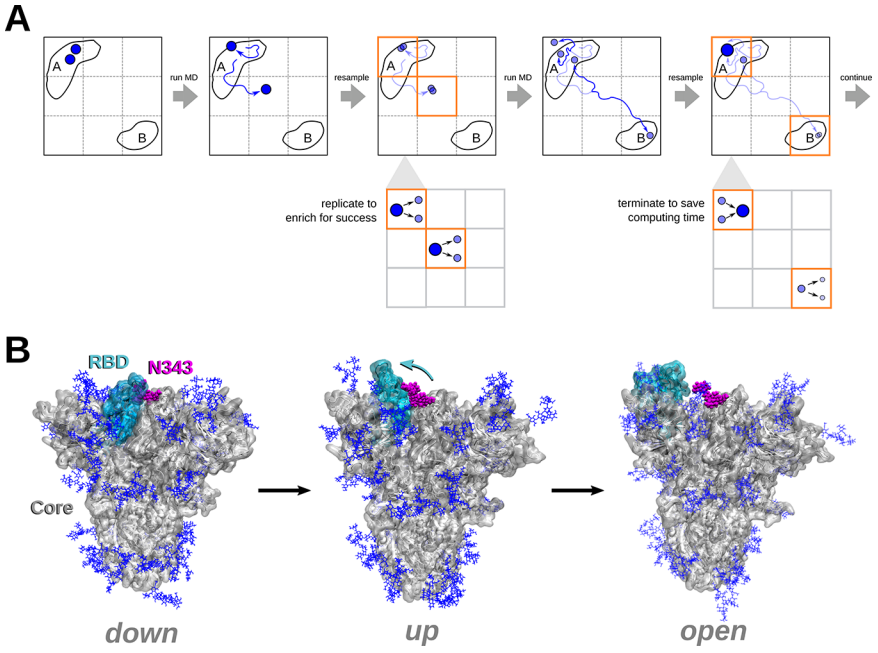


Figure 3. Weighted ensemble simulations of the opening of the SARS-CoV-2 spike protein. (A) Schematic of the weighted ensemble strategy. Trajectories (blue circles) are initiated from state A with equal statistical weights, propagating the dynamics in parallel (blue arrows) for fixed time intervals and applying a resampling procedure after each time interval to ensure equal coverage of configurational space (in this illustration, two trajectories per bin along a two-dimensional progress). The resampling procedure involves replicating trajectories that make transitions to less visited bins and occasionally terminating trajectories that have not made such transitions while rigorously tracking the trajectory weights (indicated by the sizes of the circles). The process of running dynamics and resampling is repeated until a desired number of trajectories have arrived in the target state B. (B) The SARS-CoV-2 spike activation process simulated using the WE strategy. The simulations involved the head region of the spike protein (gray) with full glycosylation (blue) and captured hundreds of switching pathways from the "down" state of the receptor binding domain (cyan) to the "up" and "open" states. Based on these pathways, the glycan attached to the N343 residue (magenta) functions as a gate that controls the switching process.

mutants that exhibit reversion rates from 21-fold faster to 78-fold slower than those of WT LOV2.[21] The mechanism(s) of rate enhancement remain unclear but may involve destabilization of the Cys-FMN adduct.

An example of LOV2 used as in input domain for a functional switch is the LOV2 Trap and Release of Protein (LOVTRAP) system.[22] LOVTRAP is a two-component switch consisting of LOV2 and Zdark, a 38-residue peptide that was evolved by mRNA display to bind the dark conformation of LOV2. The crystal structure of the dark-state complex revealed that Zdark (cyan in Figure 2C) binds to the LOV2 core domain as well as the tip of Jα. Light-induced unfolding of Jα causes Zdark to dissociate in under a second. As anticipated from earlier LOV2 studies, the dark-to-light conformational change limits the overall response time of LOVTRAP. Reassociation half-times were tuned from 10-fold faster to 26-fold slower (covering a range of 2 s to 9 min) relative to the WT LOV2 construct[22] by mutating two of the FMN-contacting residues previously described.[21,23] LOVTRAP has been used to introduce photocontrol to protein subcellular localization and protein−protein interactions. A protein of interest (POI) is fused to Zdark (or LOV2), and LOV2 (or Zdark) is sequestered to an organelle or anchored to a membrane. Light irradiation causes the POI to dissociate and diffuse to its preferred cellular location and interact with its natural binding partner.[24] In addition to its use in the modular, two-component Zdark system, reversible Jα unfolding has been employed to regulate functions of specific proteins by directly fusing LOV2 to nanobodies,[19] Src kinase,[25] Rac,[26] CamKII,[27] and others. A guide for how to engineer LOV2-based photoswitches, along with tables of the characterized kinetic mutants and potential applications, has been published recently.[28]

DE NOVO DESIGNED SWITCHES

While building complex allosteric pathways such as those encoded in the CaM and LOV2 sequences is presently out of reach, de novo design of novel protein scaffolds and binding interfaces has reached adolescence, if not maturity. De novo methods thus establish a route for creating binding domains that can be customized to recognize ligands of choice as well as new mechanisms for transducing input to output signals via coupled binding events. The latching orthogonal cage−key (LOCKR) family of protein switches, developed by Baker and colleagues, is composed of a six-helix bundle with the first five helices designated as the "cage" (gray in Figure 2D) and the sixth as the "latch" (cyan).[29] The "key" is an exogenously added helix that resembles the latch and competes with the latch for docking to the cage. The latch embeds a peptide that can bind a protein partner but is made cryptic by burial in the latch−cage interface. The switch is turned on by addition of the key, which displaces the latch and exposes the peptide for binding its partner. The identity of the peptide establishes the output signal; existing examples are a Bim sequence (programmed cell death) and a degron peptide (protein degradation).

LOCKR switches are activated by a single binding event but are capable of multiple output functionalities. The related lucCage design reverses this relationship to enable biosensors that bind different targets and produce a dedicated output

https://doi.org/10.1021/acs.jpcb.1c04082
J. Phys. Chem. B 2021, 125, 9069−9077

signal (e.g., luminescence).[30] To do so, the latch was modified to contain a domain at its C-terminus (magenta in Figure 2D) that was de novo engineered to possess shared affinity for the cage as well as to any one of a number of analytes to be detected. The activating peptide, also in the latch, was changed to a split luciferase fragment. The complementary luciferase fragment was fused to the key. The combination of analyte binding to the latch and key binding to the cage displaces the latch and allows the luciferase fragments to complement, turning on bioluminescence.

LOCKR and lucCage represent combinations of folds and stabilizing interactions that do not exist in nature. Moreover, the Rosetta-based computational methods used to design them only target the final, lowest-energy structure. They do not consider partially folded structures, pathways, or barriers. A fundamental question thus arises: how do switching rates of de novo designed proteins compare with rates of folding/unfolding and conformational changes of natural proteins? LOCKR and lucCage exhibit turn-on and turn-off times in the minutes-to-hours scale.[30] These times are similar to those of the protein fragment exchange (FREX)-based biosensors.[31] Introduced in 2014, FREX sensors established the analogous unlocking/exchange mechanism to generate output (FRET) but were made from the human fibronectin 3 binding scaffold. This limited comparison suggests that de novo switches already operate with rates in the biological regime even though their designs are based solely on thermodynamic and not kinetic principles. Baker speculates that the folding landscapes of de novo designed proteins tend to be smooth funnels, devoid of large energy barriers, because the design process successfully eliminates competing low-energy states with very different structures.[32] Nevertheless, there is always room for improvement, and response times of switches based on de novo designs and natural proteins alike can be optimized by using the computational approaches described below.

COMPUTATIONAL APPROACHES TO TUNING RATES

To our knowledge, only one computational study has reported the rational enhancement of kinetics for a protein conformational switch.[33] The goal of this study was to speed up the slow response time (hundreds of milliseconds) of the engineered protein-based calcium sensor, calbindin-AFF (Figure 2A), by at least an order of magnitude to detect fast physiological Ca^{2+} fluctuations. The computational strategy involved (i) a minimal, residue-level protein model (one bead per residue), (ii) a Gō-type potential[34] that was parametrized to reproduce the thermodynamic stability of each switch component, and (iii) the weighted ensemble (WE) path sampling strategy,[35] which can be orders of magnitude more efficient than standard simulations in generating pathways and rate constants for rare events (e.g., protein folding and protein binding) without introducing any external bias in the dynamics or altering the free energy landscape (Figure 3A).[36] The only prerequisites for this strategy are the structure and experimental folding free energy of the nonpermutant switch component. Despite the simplicity of the simulation model, this strategy identified previously untested mutations that decreased response time by as much as 32-fold (590 to 19 ms) via preferential destabilization of the ground states relative to the transition path ensemble (TPE), which is defined as all transient states in productive pathways, beginning where the trajectory last exited the initial state and ending where the trajectory first entered the target state. In particular, we focused on large, hydrophobic residues that form the most pairwise residue contacts in the initial ground state relative to the TPE-prime candidates for "underpacking" mutations that destabilize the ground state by removing hydrophobic interactions. Importantly, a negative control mutation was correctly predicted to have little effect on the kinetics despite being located near the other mutations. Furthermore, this study demonstrated that the efficiency of the WE strategy relative to standard simulations in estimating rate constants increases *exponentially* with the effective free energy barrier and can therefore be applied to switches of a similar size (less than a few hundred amino acids) with even slower response times (<100 s).[33]

Although the WE strategy has not been used with atomistic models to rationally manipulate the kinetics for an engineered protein-based conformational switch, the strategy has enabled the generation of rate constants and atomistic pathways for complex biological processes such as protein folding,[37] protein−protein binding,[38] and protein−ligand unbinding.[39] Furthermore, encouraging WE results have been obtained for the activation process of a particularly large natural switchable protein: the glycosylated SARS-CoV-2 spike protein,[40] which must open before binding the human ACE2 receptor to fuse and infect the human host cell. The system for this WE simulation consisted of the head region (residues 16−1140), explicit water molecules, and a physiological ionic strength (150 mM NaCl), totaling almost half a million atoms.

As part of an international team effort that was awarded the 2020 Gordon Bell Special Prize for HPC-Based COVID-19 Research, WE simulations yielded atomically detailed pathways for the opening of the spike receptor binding domain (cyan in Figure 2E) from glycan-shielded state (down) to exposed (up) and open states (Figure 3B).[41] The conformations of the open state align closely with the cryo-EM structure of the ACE2-bound spike protein.[42] While standard MD simulations would require hundreds of years to capture a single, atomically detailed pathway for the opening of the spike—a seconds time scale process[43]—the WE simulations generated hundreds of pathways for spike opening in 45 days by using 100 NVIDIA V100 GPUs in parallel on the TACC Longhorn supercomputer. These pathways reveal that a glycan attached to the N343 residue (magenta in Figure 2E) functions as a gate that controls the opening (switching) process of the spike protein. The functional importance of this glycan has been validated by biolayer interferometry experiments, which revealed a 56% reduction in binding to the ACE2 receptor when N343 is mutated to an alanine. Furthermore, the large-scale collective motions of the spike-opening process are consistent with those observed in two-dimensional cryogenic electron microscopy images of the spike protein.[44] The WE simulations set a new high-water mark for ensemble simulations of atomistic pathways, capturing seconds time scale motions for a massive protein system.

In another simulation study, which was completed on the Folding@home distributing computing resource, adaptive sampling also captured the open conformations of the spike protein,[45] including the ACE2-bound spike conformation that was sampled by the WE simulations.[44] Like the WE strategy, adaptive sampling is an enhanced sampling strategy that involves iteratively splitting (or replicating) trajectories that have progressed closer to the target state. Together, these results demonstrate the power of "splitting" strategies in sampling switching processes that are beyond the milliseconds

https://doi.org/10.1021/acs.jpcb.1c04082
J. Phys. Chem. B 2021, 125, 9069−9077

time scale and the value of applying such strategies with atomistic models—even when the estimation of rate constants remains a challenge. In contrast to many engineered protein conformational switches, it is *not* of interest to enhance the switch response time of the spike protein. Rather, the ultimate goal of these studies is to inform strategies for locking the protein in the OFF state, for example, by targeting structures of stable or transient states with small molecules as potential drug inhibitors of COVID-19.

FUTURE AREAS OF IMPROVEMENT FOR COMPUTATIONAL STRATEGIES

Promising avenues for improving the effectiveness of computational strategies in tuning the kinetics of protein conformational switches include (i) more accurate residue-level, coarse-grained simulation models (force fields) that can offer orders of magnitude speedup over all-atom force fields and (ii) more efficient enhanced sampling strategies to enable faster predictions of mutations that can enhance switching kinetics. Given the slow response times of many engineered protein-based switches, enhanced sampling strategies are essential for capturing switching pathways, even with the use of coarse-grained force fields.

An ongoing challenge of coarse-grained force fields is the ability to simulate protein folding transitions with realistic kinetics. Gō-type potentials[34] on their own have been useful from the perspective of protein engineering in terms of (i) their abilities to reproduce experimental stabilities of individual switch components by optimizing the primary adjustable parameter (the well-depth ε) and (ii) their abilities to capture the cooperativity of protein folding, yielding fragment stabilities that are consistent with experimental data.[33] However, such models yield artificially accelerated dynamics due to the neglect of stabilizing non-native interactions[46] and may not capture non-native, metastable intermediates. On the other hand, the latest generation of coarse-grained force fields that include non-native interactions such as the MARTINI 3[47] and SIRAH 2.0 force fields[48] have not yet matured to the point of being adequate on their own for simulating the folding transitions that can occur for certain protein conformational switches, requiring restraints to maintain secondary structures. To combine the best of both worlds, one might use a hybrid of a Gō-type potential and coarse-grained force field such as MARTINI 3 or SIRAH 2.0. In the very least, electrostatic interactions—both native and non-native—could be used with Gō-type potentials to provide a more realistic, rugged free energy landscape for more quantitative modeling of the protein switching process.

A major challenge of the WE strategy and many other enhanced sampling strategies is the identification of a progress coordinate for the process of interest (e.g., conformational switching). Recent deep learning approaches identify potential progress coordinates by encoding a high-dimensional set of conformational and dynamical features from a training set of trajectory data onto a low-dimensional representation of the features; the progress coordinate can then be decoded to obtain physically relevant details. Two such approaches are the Convolutional Variational Autoencoder (CVAE) method[49] and the Reweighted Autoencoded Variational Bayes for Enhanced Sampling (RAVE) method.[50] The CVAE method has been applied to protein folding, differentiating between various intermediates in the folding process of Fs-peptide,[49] and the RAVE method has been able to detect subtle loop fluctuations in the T4 lysozyme enzyme.[51] Both of these studies highlight the promise of such strategies for aiding the enhanced sampling of large conformational transitions of protein switches. In addition, such strategies could learn effective progress coordinates more efficiently by using complete pathways of the switching processes from WE simulations as training data.[41] These simulations could involve coarse-grained models even if the end goal is to simulate with all-atom models—as long as the coarse-grained simulations have captured the relevant slow motions of the process. Once an effective progress coordinate has been identified, an adaptive binning strategy such as the Minimal Adaptive Binning (MAB) strategy may be applied to automate the placement of bins along the progress coordinate during a simulation to more efficiently surmount "bottleneck" regions.[52]

Finally, several strategies have been developed for more efficient estimation of rate constants from simulations that have not yet reached a steady state. These strategies include the Rates from Event Duration (RED) scheme, which can estimate rate constants with up to 50% greater efficiency than the original scheme for WE simulations[35] by incorporating the probability distribution of sampled event durations (barrier crossing times).[53] Rate constants can also be estimated more efficiently by constructing a history-augmented Markov state model (haMSM) from completed simulations (e.g., weighted ensemble, adaptive sampling, and standard simulations).[54] In contrast to a standard MSM, an haMSM does not require the use of a long lag time (e.g., ∼100 ns) and can therefore provide pathway and kinetics observables for time scales that are both shorter and longer than the lag time.[55] To further accelerate convergence to a steady state, an haMSM could be constructed periodically during a WE simulation to iteratively reweight trajectories.[55] The combination of this on-the-fly reweighting with WE simulation could enable the estimation of rate constants for processes as slow as the seconds time scale, including the switching process of the SARS-CoV-2 spike.[44]

Computational strategies for tuning rates might be applied in two stages. In the first stage, a large set of switch constructs could be virtually screened by using coarse-grained simulations, qualitatively ranking the constructs based on the extent of switching (signal-to-noise) and kinetics (response time). In the second stage, the top one to three switch constructs from the first stage could be characterized by using all-atom simulations to quantitatively identify candidate residues for mutation to improve the response time of the switch. As mentioned above, both stages benefit greatly from the application of enhanced sampling strategies that provide rigorous kinetics (e.g., the WE strategy). To further improve on the efficiency of the computational strategy, deep learning/artificial intelligence strategies could be used to identify more effective progress coordinates for the enhanced sampling and to aid in the detailed analysis of how the protein conformational transitions can occur.

INTEGRATING EXPERIMENTAL AND COMPUTATIONAL APPROACHES

The power of synergistically combining experimental and computational strategies has been demonstrated for the engineered calbindin-AFF calcium sensor[33]—the only study (to our knowledge) to date that has been successful in rationally enhancing the response time of a protein switch. While time-resolved experiments can measure rate constants for the overall switching process and the thermodynamic

https://doi.org/10.1021/acs.jpcb.1c04082
J. Phys. Chem. B 2021, 125, 9069−9077

stabilities of each switch component (i.e., folding free energies), molecular simulations can provide complete pathways for the switching process, including structures of transient states, which are essential for predicting mutations that could enhance the kinetics. The only prerequisites for these simulations are the structures of the individual switch components and the experimental folding free energy of each component. The latter is used to parametrize the simulation model to yield the expected relative stabilities of the stable states. In the case of a AFF switch construct, only the structure and folding free energy of the parent protein were required for parametrization of the model. The other stable state is a circular permutant of the same protein which can be modeled based on the structure of the parent protein. To further reduce the amount of guesswork and effort required of experiments, these simulations could be used to virtually screen candidate mutations for enhanced response times. Importantly, both thermodynamic and kinetics experiments provide validation of the simulations and help inform the level of detail that is required of the simulation models.

CONCLUDING THOUGHTS

In closing, experimental and computational strategies have matured to the point where they can be synergistically combined to reduce the amount of guesswork required to engineer protein conformational switches with desired response times. In our own studies, experimentally determined protein structures and thermodynamic stabilities played critical roles in establishing computational simulations and calibrating them. Conversely, theoretical results inform experiments. For example, de novo approaches alone are seldom sufficient to generate functioning switches. Instead, they typically define structures and amino acid sequences that serve as the starting points for directed or random mutagenesis and library screening experiments. While the prediction of switch response times on time scales beyond milliseconds remain a challenge for atomistic simulations, such simulations of the seconds time scale switching process of the massive SARS-CoV-2 spike protein have demonstrated that the generation of complete pathways for the switching process is in its own right highly valuable, providing direct views of *how* the protein switches from the OFF to the ON state, including the structures of transient states for manipulating the switching kinetics. Given the ever-ongoing advances in computer software and hardware, the future is bright for quantitative predictions of switching kinetics.

AUTHOR INFORMATION

Corresponding Authors

Lillian T. Chong − *Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, United States;* orcid.org/0000-0002-0590-483X; Email: ltchong@pitt.edu

Stewart N. Loh − *Department of Biochemistry and Molecular Biology, State University of New York Upstate Medical University, Syracuse, New York 13210, United States;* Email: lohs@upstate.edu

Authors

Anthony T. Bogetti − *Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, United States;* orcid.org/0000-0003-0610-2879

Maria F. Presti − *Department of Biochemistry and Molecular Biology, State University of New York Upstate Medical University, Syracuse, New York 13210, United States;* orcid.org/0000-0003-1541-5770

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.1c04082

Author Contributions

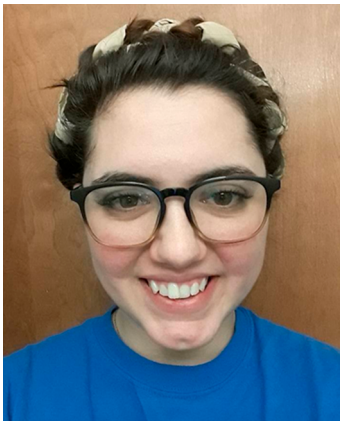
A.T.B. and M.F.P. contributed equally to this work.

Notes

The authors declare the following competing financial interest(s): L.T.C. is an Open Science Fellow with Silicon Therapeutics.

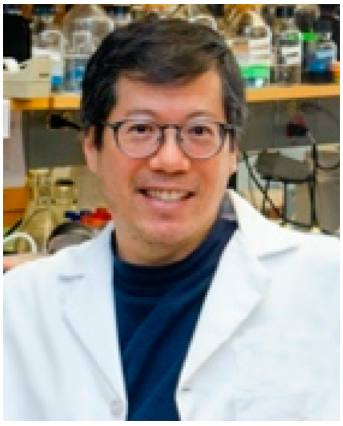
Biographies

Anthony T. Bogetti is a graduate student at the University of Pittsburgh working toward a Ph.D. in Chemistry with Prof. Lillian Chong. He received his B.S. in Chemistry (2017) from Messiah University. Anthony is working on applications of the weighted ensemble path sampling strategy to the simulation of alternate frame folding protein conformational switches. He also works hard to improve the underlying methodology, usability, and accessibility of the weighted ensemble strategy.



Maria F. Presti is a graduate student at SUNY Upstate Medical University working toward a Ph.D. in Biochemistry and Molecular Biology with Prof. Stewart Loh. She received her B.A. in Chemistry and Biology (2016) from Bard College at Simon's Rock. Maria is working on creating an adaptable biosensor scaffold with monobody alternate frame folds and has a general interest in protein engineering, biosensor development, and scientific outreach.

https://doi.org/10.1021/acs.jpcb.1c04082
J. Phys. Chem. B 2021, 125, 9069−9077



Stewart N. Loh received his Ph.D. in Biochemistry from the University of Wisconsin—Madison in 1993 in the laboratory of John L. Markley and was a Damon Runyon-Walter Winchell Postdoctoral Fellow in the Biochemistry Department at Stanford University from 1993 to 1996. Dr. Loh began studying protein folding mechanisms in the laboratory of Dr. Robert "Buzz" Baldwin at Stanford and has continued these studies since joining the faculty at SUNY Upstate Medical University in 1996. Prof. Loh's primary research focus is to develop mechanisms for converting ordinary proteins into conformational switches for use in biology, biotechnology, and medicine. His laboratory also studies folding and misfolding of the p53 tumor suppressor and how to restore function to tumorigenic p53 mutants by using small molecules.

Lillian T. Chong is an Associate Professor in the Department of Chemistry at the University of Pittsburgh. She earned a B.S. in Chemistry at the Massachusetts Institute of Technology and a Ph.D. in Biophysics with Prof. Peter Kollman from the University of California in San Francisco. She completed her postdoctoral work with Prof. Vijay Pande at Stanford University and Dr. William Swope at the IBM Almaden Research Center. Prof. Chong's research involves the development and application of molecular simulation approaches to model a variety of biophysical processes. Her laboratory also applies these approaches to the rational design of protein conformational switches.

■ ACKNOWLEDGMENTS

This work was supported by NIH 1R01GM115805-01 to L.T.C., NIH Grant GM115762 to S.N.L., and the University of Pittsburgh to A.T.B. (Dietrich School of Arts and Sciences Graduate Fellowship). Computational resources were provided by the University of Pittsburgh's Center for Research Computing. We thank Harsimranjit Sekhon, Rommie Amaro, Terra Sztain, and Lorenzo Casalino for helpful discussions.

■ REFERENCES

(1) Adamson, H.; Jeuken, L. J. C. Engineering Protein Switches for Rapid Diagnostic Tests. *ACS Sens.* **2020**, *5*, 3001−3012.
(2) Farahani, P. E.; Reed, E. H.; Underhill, E. J.; Aoki, K.; Toettcher, J. E. Signaling, Deconstructed: Using Optogenetics to Dissect and Direct Information Flow in Biological Systems. *Annu. Rev. Biomed. Eng.* **2021**, *23*, 61−87.
(3) Pan, X.; Kortemme, T. Recent Advances in de Novo Protein Design: Principles, Methods, and Applications. *J. Biol. Chem.* **2021**, *296*, 100558.
(4) Baker, D. What Has de Novo Protein Design Taught Us about Protein Folding and Biophysics? *Protein Sci.* **2019**, *28*, 678−683.
(5) Fersht, A. R.; Matouschek, A.; Serrano, L. The Folding of an Enzyme. I. Theory of Protein Engineering Analysis of Stability and Pathway of Protein Folding. *J. Mol. Biol.* **1992**, *224*, 771−782.
(6) Naganathan, A. N.; Muñoz, V. Insights into Protein Folding Mechanisms from Large Scale Analysis of Mutational Effects. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 8611.
(7) Zhang, Y.; Kitazawa, S.; Peran, I.; Stenzoski, N.; McCallum, S. A.; Raleigh, D. P.; Royer, C. A. High Pressure ZZ-Exchange NMR Reveals Key Features of Protein Folding Transition States. *J. Am. Chem. Soc.* **2016**, *138* (46), 15260−15266.
(8) Nielsen, A. K.; Möller, I. R.; Wang, Y.; Rasmussen, S. G. F.; Lindorff-Larsen, K.; Rand, K. D.; Loland, C. J. Substrate-Induced Conformational Dynamics of the Dopamine Transporter. *Nat. Commun.* **2019**, *10*, 2714.
(9) Jia, R.; Martens, C.; Shekhar, M.; Pant, S.; Pellowe, G. A.; Lau, A. M.; Findlay, H. E.; Harris, N. J.; Tajkhorshid, E.; Booth, P. J.; Politis, A. Hydrogen-Deuterium Exchange Mass Spectrometry Captures Distinct Dynamics upon Substrate and Inhibitor Binding to a Transporter. *Nat. Commun.* **2020**, *11* (1), 6162.
(10) Liebau, J.; Tersa, M.; Trastoy, B.; Patrick, J.; Rodrigo-Unzueta, A.; Corzana, F.; Sparrman, T.; Guerin, M. E.; Mäler, L. Unveiling the Activation Dynamics of a Fold-Switch Bacterial Glycosyltransferase by 19F NMR. *J. Biol. Chem.* **2020**, *295* (29), 9868−9878.
(11) Fenwick, R. B.; Oyen, D.; van den Bedem, H.; Dyson, H. J.; Wright, P. E. Modeling of Hidden Structures Using Sparse Chemical Shift Data from NMR Relaxation Dispersion. *Biophys. J.* **2021**, *120* (2), 296−305.
(12) Stratton, M. M.; Mitrea, D. M.; Loh, S. N. A Ca2+-Sensing Molecular Switch Based on Alternate Frame Protein Folding. *ACS Chem. Biol.* **2008**, *3*, 723−732.
(13) Stratton, M. M.; Loh, S. N. On the Mechanism of Protein Fold-Switching by a Molecular Sensor. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 3260−3269.
(14) Nakai, J.; Ohkura, M.; Imoto, K. A High Signal-to-Noise Ca2+ Probe Composed of a Single Green Fluorescent Protein. *Nat. Biotechnol.* **2001**, *19*, 137−141.
(15) Inoue, M.; Takeuchi, A.; Manita, S.; Horigane, S.-I.; Sakamoto, M.; Kawakami, R.; Yamaguchi, K.; Otomo, K.; Yokoyama, H.; Kim, R.; Yokoyama, T.; et al. Rational Engineering of XCaMPs, a Multicolor GECI Suite for In Vivo Imaging of Complex Brain Circuit Dynamics. *Cell* **2019**, *177*, 1346−1360.
(16) Kerruth, S.; Coates, C.; Dürst, C. D.; Oertner, T. G.; Török, K. The Kinetic Mechanisms of Fast-Decay Red-Fluorescent Genetically Encoded Calcium Indicators. *J. Biol. Chem.* **2019**, *294*, 3934−3946.
(17) Sun, X. R.; Badura, A.; Pacheco, D. A.; Lynch, L. A.; Schneider, E. R.; Taylor, M. P.; Hogue, I. B.; Enquist, L. W.; Murthy, M.; Wang, S. S.-H. Fast GCaMPs for Improved Tracking of Neuronal Activity. *Nat. Commun.* **2013**, *4*, 2170.
(18) Halavaty, A. S.; Moffat, K. N- and C-Terminal Flanking Regions Modulate Light-Induced Signal Transduction in the LOV2 Domain of the Blue Light Sensor Phototropin 1 from Avena Sativa. *Biochemistry* **2007**, *46*, 14001−14009.
(19) Iuliano, J. N.; Collado, J. T.; Gil, A. A.; Ravindran, P. T.; Lukacs, A.; Shin, S.; Woroniecka, H. A.; Adamczyk, K.; Aramini, J. M.; Edupuganti, U. R.; et al. Unraveling the Mechanism of a LOV Domain Optogenetic Sensor: A Glutamine Lever Induces Unfolding of the Jα Helix. *ACS Chem. Biol.* **2020**, *15*, 2752−2765.

https://doi.org/10.1021/acs.jpcb.1c04082
J. Phys. Chem. B 2021, 125, 9069−9077

(20) Pudasaini, A.; El-Arab, K. K.; Zoltowski, B. D. LOV-Based Optogenetic Devices: Light-Driven Modules to Impart Photoregulated Control of Cellular Signaling. *Front. Mol. Biosci.* **2015**, *2*, 18.
(21) Kawano, F.; Aono, Y.; Suzuki, H.; Sato, M. Fluorescence Imaging-Based High-Throughput Screening of Fast- and Slow-Cycling LOV Proteins. *PLoS One* **2013**, *8*, e82693.
(22) Wang, H.; Vilela, M.; Winkler, A.; Tarnawski, M.; Schlichting, I.; Yumerefendi, H.; Kuhlman, B.; Liu, R.; Danuser, G.; Hahn, K. M. LOVTRAP, An Optogenetic System for Photo-Induced Protein Dissociation. *Nat. Methods* **2016**, *13*, 755−758.
(23) Zoltowski, B. D.; Vaccaro, B.; Crane, B. R. Mechanism-Based Tuning of a LOV Domain Photoreceptor. *Nat. Chem. Biol.* **2009**, *5*, 827−834.
(24) Tischer, D. K.; Weiner, O. D. Light-based tuning of ligand half-life supports kinetic proofreading model of T cell signaling. *eLife* **2019**, *8*, No. e42498.
(25) Dagliyan, O.; Tarnawski, M.; Chu, P.-H.; Shirvanyants, D.; Schlichting, I.; Dokholyan, N. V.; Hahn, K. M. Engineering Extrinsic Disorder to Control Protein Activity in Living Cells. *Science* **2016**, *354*, 1441−1444.
(26) Wang, X.; He, L.; Wu, Y. I.; Hahn, K. M.; Montell, D. J. Light-Mediated Activation Reveals a Key Role for Rac in Collective Guidance of Cell Movement in Vivo. *Nat. Cell Biol.* **2010**, *12*, 591−597.
(27) Shibata, A. C. E.; Ueda, H. H.; Eto, K.; Onda, M.; Sato, A.; Ohba, T.; Nabekura, J.; Murakoshi, H. Photoactivatable CaMKII Induces Synaptic Plasticity in Single Synapses. *Nat. Commun.* **2021**, *12*, 751.
(28) Zimmerman, S. P.; Kuhlman, B.; Yumerefendi, H. Engineering and Application of LOV2-Based Photoswitches. *Methods Enzymol.* **2016**, *580*, 169−190.
(29) Langan, R. A.; Boyken, S. E.; Ng, A. H.; Samson, J. A.; Dods, G.; Westbrook, A. M.; Nguyen, T. H.; Lajoie, M. J.; Chen, Z.; Berger, S.; et al. De Novo Design of Bioactive Protein Switches. *Nature* **2019**, *572*, 205−210.
(30) Quijano-Rubio, A.; Yeh, H.-W.; Park, J.; Lee, H.; Langan, R. A.; Boyken, S. E.; Lajoie, M. J.; Cao, L.; Chow, C. M.; Miranda, M. C.; et al. De Novo Design of Modular and Tunable Protein Biosensors. *Nature* **2021**, *591*, 482−487.
(31) Zheng, H.; Bi, J.; Krendel, M.; Loh, S. N. Converting a Binding Protein into a Biosensing Conformational Switch Using Protein Fragment Exchange. *Biochemistry* **2014**, *53*, 5505−5514.
(32) Baker, D. What Has de Novo Protein Design Taught Us about Protein Folding and Biophysics? *Protein Sci.* **2019**, *28*, 678−683.
(33) DeGrave, A. J.; Ha, J.-H.; Loh, S. N.; Chong, L. T. Large Enhancement of Response Times of a Protein Conformational Switch by Computational Design. *Nat. Commun.* **2018**, *9*, 1013.
(34) Go, N. Theoretical Studies of Protein Folding. *Annu. Rev. Biophys. Bioeng.* **1983**, *12*, 183−210.
(35) Huber, G. A.; Kim, S. Weighted-Ensemble Brownian Dynamics Simulations for Protein Association Reactions. *Biophys. J.* **1996**, *70*, 97−110.
(36) Zuckerman, D. M.; Chong, L. T. Weighted Ensemble Simulation: Review of Methodology, Applications, and Software. *Annu. Rev. Biophys.* **2017**, *46*, 43−57.
(37) Adhikari, U.; Mostofian, B.; Copperman, J.; Subramanian, S. R.; Petersen, A. A.; Zuckerman, D. M. Computational Estimation of Microsecond to Second Atomistic Folding Times. *J. Am. Chem. Soc.* **2019**, *141*, 6519−6526.
(38) Saglam, A. S.; Chong, L. T. Protein−Protein Binding Pathways and Calculations of Rate Constants Using Fully-Continuous, Explicit-Solvent Simulations. *Chem. Sci.* **2019**, *10*, 2360−2372.
(39) Dixon, T.; Uyar, A.; Ferguson-Miller, S.; Dickson, A. Membrane-Mediated Ligand Unbinding of the PK-11195 Ligand from TSPO. *Biophys. J.* **2021**, *120*, 158−167.
(40) Casalino, L.; Gaieb, Z.; Goldsmith, J. A.; Hjorth, C. K.; Dommer, A. C.; Harbison, A. M.; Fogarty, C. A.; Barros, E. P.; Taylor, B. C.; McLellan, J. S.; et al. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* **2020**, *6*, 1722−1734.
(41) Casalino, L.; Dommer, A. C.; Gaieb, Z.; Barros, E. P.; Sztain, T.; Ahn, S.-H.; Trifan, A.; Brace, A.; Bogetti, A. T.; Clyde, A. AI-Driven Multiscale Simulations Illuminate Mechanisms of SARS-CoV-2 Spike Dynamics. *International Journal of High Performance Computing Applications* **2021**, 1−20.
(42) Benton, D. J.; Wrobel, A. G.; Xu, P.; Roustan, C.; Martin, S. R.; Rosenthal, P. B.; Skehel, J. J.; Gamblin, S. J. Receptor Binding and Priming of the Spike Protein of SARS-CoV-2 for Membrane Fusion. *Nature* **2020**, *588*, 327−330.
(43) Lu, M.; Uchil, P. D.; Li, W.; Zheng, D.; Terry, D. S.; Gorman, J.; Shi, W.; Zhang, B.; Zhou, T.; Ding, S.; et al. Real-Time Conformational Dynamics of SARS-CoV-2 Spikes on Virus Particles. *Cell Host Microbe* **2020**, *28*, 880−891.
(44) Sztain, T.; Ahn, S.-H.; Bogetti, A. T.; Casalino, L.; Goldsmith, J. A.; McCool, R. S.; Kearns, F. L.; McCammon, J. A.; McLellan, J. S.; Chong, L. T.; Amaro, R. E. A Glycan Gate Controls Opening of the SARS-CoV-2 Spike Protein. *Nat. Chem.*, in press, **2021**; DOI: 10.1038/s41557-021-00758-3.
(45) Zimmerman, M. I.; Porter, J. R.; Ward, M. D.; Singh, S.; Vithani, N.; Meller, A.; Mallimadugula, U. L.; Kuhn, C. E.; Borowsky, J. H.; Wiewiora, R. P.; et al. SARS-CoV-2 simulations go exascale to predict dramatic spike opening and cryptic pockets across the proteome. *Nat. Chem.* **2021**, *13*, 651−659.
(46) Clementi, C.; Nymeyer, H.; Onuchic, J. N. Topological and Energetic Factors: What Determinesthe Structural Details of the Transition State Ensembleand "En-Route" Intermediates for Protein Folding? An Investigation for Small Globular Proteins. *J. Mol. Biol.* **2000**, *298*, 937−953.
(47) Souza, P. C. T.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M. H.; Wassenaar, T. A.; et al. Martini 3: A General Purpose Force Field for Coarse-Grained Molecular Dynamics. *Nat. Methods* **2021**, *18*, 382−388.
(48) Machado, M. R.; Barrera, E. E.; Klein, F.; Sóñora, M.; Silva, S.; Pantano, S. The SIRAH 2.0 Force Field: Altius, Fortius, Citius. *J. Chem. Theory Comput.* **2019**, *15*, 2719−2733.
(49) Bhowmik, D.; Gao, S.; Young, M. T.; Ramanathan, A. Deep Clustering of Protein Folding Simulations. *BMC Bioinf.* **2018**, *19*, 484.
(50) Ribeiro, J. M. L.; Bravo, P.; Wang, Y.; Tiwary, P. Reweighted Autoencoded Variational Bayes for Enhanced Sampling (RAVE). *J. Chem. Phys.* **2018**, *149*, 072301.
(51) Smith, Z.; Ravindra, P.; Wang, Y.; Cooley, R.; Tiwary, P. Discovering Protein Conformational Flexibility through Artificial-Intelligence-Aided Molecular Dynamics. *J. Phys. Chem. B* **2020**, *124*, 8221−8229.
(52) Torrillo, P. A.; Bogetti, A. T.; Chong, L. T. A Minimal, Adaptive Binning Scheme for Weighted Ensemble Simulations. *J. Phys. Chem. A* **2021**, *125*, 1642−1649.
(53) DeGrave, A. J.; Bogetti, A. T.; Chong, L. T. The RED Scheme: Rate-Constant Estimation from Pre-Steady State Weighted Ensemble Simulations. *J. Chem. Phys.* **2021**, *154*, 114111.
(54) Copperman, J.; Zuckerman, D. M. Accelerated Estimation of Long-Timescale Kinetics from Weighted Ensemble Simulation via Non-Markovian "Microbin" Analysis. *J. Chem. Theory Comput.* **2020**, *16*, 6763−6775.
(55) Suárez, E.; Wiewiora, R. P.; Wehmeyer, C.; Noé, F.; Chodera, J. D.; Zuckerman, D. M. What Markov State Models Can and Cannot Do: Correlation versus Path-Based Observables in Protein Folding Models. *J. Chem. Theory Comput.* **2021**, *17*, 3119−3133.

https://doi.org/10.1021/acs.jpcb.1c04082
J. Phys. Chem. B 2021, 125, 9069−9077