

Folding@home: Achievements from over 20 years of citizen science herald the exascale era

Vincent A. Voelz,¹ Vijay S. Pande,² and Gregory R. Bowman^{3,*}

¹Department of Chemistry, Temple University, Philadelphia, Pennsylvania; ²Andreessen Horowitz, Menlo Park, California; and ³Departments of Biochemistry & Biophysics and of Bioengineering, University of Pennsylvania, Philadelphia, Pennsylvania

ABSTRACT Simulations of biomolecules have enormous potential to inform our understanding of biology but require extremely demanding calculations. For over 20 years, the Folding@home distributed computing project has pioneered a massively parallel approach to biomolecular simulation, harnessing the resources of citizen scientists across the globe. Here, we summarize the scientific and technical advances this perspective has enabled. As the project's name implies, the early years of Folding@home focused on driving advances in our understanding of protein folding by developing statistical methods for capturing long-timescale processes and facilitating insight into complex dynamical processes. Success laid a foundation for broadening the scope of Folding@home to address other functionally relevant conformational changes, such as receptor signaling, enzyme dynamics, and ligand binding. Continued algorithmic advances, hardware developments such as graphics processing unit (GPU)-based computing, and the growing scale of Folding@home have enabled the project to focus on new areas where massively parallel sampling can be impactful. While previous work sought to expand toward larger proteins with slower conformational changes, new work focuses on large-scale comparative studies of different protein sequences and chemical compounds to better understand biology and inform the development of small-molecule drugs. Progress on these fronts enabled the community to pivot quickly in response to the COVID-19 pandemic, expanding to become the world's first exascale computer and deploying this massive resource to provide insight into the inner workings of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus and aid the development of new antivirals. This success provides a glimpse of what is to come as exascale supercomputers come online and as Folding@home continues its work.

INTRODUCTION

Atomically detailed computer simulations of protein dynamics have the potential to provide insight into the mechanisms of biological processes that would be impossible to obtain by experiment alone. Molecular simulation can advance our understanding of biology and has already become an indispensable tool in drug design and protein engineering. However, these simulations are extremely computationally demanding, prompting myriad efforts to develop clever algorithms and new hardware to accelerate these calculations.

For over 20 years, the Folding@home distributed computing project has tackled this challenge by pooling the computer power of volunteers around the world to build a planetary-scale virtual supercomputer (1). By enabling anyone with a computer and an Internet connection to become a citizen scientist by helping run simulations on

their personal computer(s), the project has achieved numerous computing records (e.g., first petascale and exascale computing platform (2)) and simulation records (e.g., first simulations of ms-timescale processes (3)). Moreover, it has enabled scientific advances in topics ranging from our understanding of fundamental processes like protein folding (4–6) to new opportunities to treat diseases like viral infections, cancer, and neurodegeneration.

Folding@home has been a community effort from the beginning and shows the great power of engaging the public in the scientific process. Many volunteers participate out of scientific curiosity or a personal connection to one of the diseases being addressed. Other volunteers are gamers and computing enthusiasts, attracted by the technical challenges and gamification of participation, which gives users the chance to earn points proportional to their contributions. To engage users, the scientific team invests significant effort in social media and outreach to explain the relevant scientific principles, the questions addressed, and the new insights obtained. A community-driven ecosystem has developed around the project, providing everything from technical

Submitted September 2, 2022, and accepted for publication March 16, 2023.

*Correspondence: grbowman@seas.upenn.edu

Editor: Meyer Jackson.

<https://doi.org/10.1016/j.bpj.2023.03.028>

© 2023 Biophysical Society.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



support to scientific discussions. Many volunteers have gained a deeper appreciation for (and understanding of) science through their participation, with some going so far as to pursue science, technology, engineering, and math careers.

Recently, work on COVID-19 drove unprecedented growth in the project, creating a computing resource orders of magnitude larger than the largest supercomputers and enabling rapid progress on the global health threat that the pandemic presented (2). This development, driven by a self-organizing community of Folding@home participants, exemplifies the collective ability of citizen scientists to help tackle critical scientific problems.

Here, we reflect on what Folding@home has accomplished and the implications for the future of computational biophysics as exascale computing becomes more accessible. Such machines will enable researchers to go big in every possible dimension. Much will be learned from simulating larger systems. However, we propose that there is also much to learn from simulating many different conditions in parallel and comparing the results. For example, comparing the dynamics of multiple variants will increase our understanding of protein function, and comparing the binding of many compounds will accelerate the discovery of new drugs. We focus on proteins in this review but note that Folding@home is equally applicable to other molecular systems and has been applied fruitfully to molecules like RNA (7,8), lipids (9), and carbon nanotubes (10).

The inspiration for Folding@home

Folding@home was originally conceived at Stanford University by Vijay Pande to elucidate fundamental mechanisms of protein folding, the process by which proteins self-assemble into functioning molecular machines (1). At the time, there was great excitement at the prospect that computer simulations could provide an atomically detailed picture of how proteins fold that could not be achieved via any conceived experimental technique. However, reaching the relevant timescales even for very small model systems was far beyond reach of any existing supercomputer, much less commodity hardware.

The Napster platform for sharing music (as .mp3 files) provided a source of inspiration. Storing a large library of music at a central location and distributing it over the internet would have been a costly endeavor. However, Napster's peer-to-peer approach to file sharing allowed a large community of users to come together to create a powerful file-sharing service. Could the protein folding problem also be broken down into independent pieces that could be performed by a large, distributed community?

This inspiration spurred the Pande lab to develop an algorithm that broke the protein folding problem down into a large number of completely independent simulations (11). The algorithm made the simplifying assumption that protein folding can be modeled as a two-state process with a single

unfolded state separated by a large free energy barrier from a single folded state. A single simulation started from the folded state would make repeated attempts to cross the barrier into the folded state. Given enough simulation time, it would make many transitions back and forth between the folded and unfolded states, and one could infer the folding rate and the probabilities of the two states from the simulation. The key insight was that these attempts to cross the barrier between the folded and unfolded states did not need to be made serially. One could instead run many independent simulations starting from the unfolded state, each of which would make independent attempts to transition to the folded state. Quantities such as the folding rate could then be inferred from the relative number of simulations that successfully reached the folded state compared with the total number of simulations. Roughly speaking, capturing a ms-timescale folding process did not require a single multi- μ s-long simulation, one simply needed microseconds of total simulations time divided among different simulations. Of course, the model would work well for two-state systems and break down for more complex ones (Fig. 1). However, two-state systems were a great place to start given that they are some of the smallest and fastest folding proteins where simulations and experiments can first meet.

In the year 2000, the Pande lab announced the Folding@home distributed computing project to enable anyone with a computer and an Internet connection to contribute their personal computing power to run large numbers of completely independent simulations of protein folding, which would be aggregated and used to gain insight by the scientific team. Google took note and added a button to the toolbar they distributed for searching the web that allowed users to volunteer their computers to contribute to Folding@home. Many people opted to participate in Folding@home by letting the software use their computing power when their machines were otherwise idle. Folding@home quickly began setting computing records. For example, it holds the Guinness world record for being the first petascale computer, capable of performing a million billion operations per second. More recently, Folding@home became the first exascale computer, capable of performing a billion billion operations per second (2).

Insights into protein folding

Folding@home enabled a number of early successes in capturing the atomic details of the folding of small model systems (4). For example, Zagrovic captured the folding of a β -hairpin (12). Snow et al. then used Folding@home to capture the folding of a small mini protein called BBA5 (13). They predicted the folding rate, and good agreement with the prediction was found in corresponding laser temperature jump experiments. In parallel, the Pande lab showed that simulations were sufficiently accurate to fold the villin headpiece (14), a small fast folding protein that

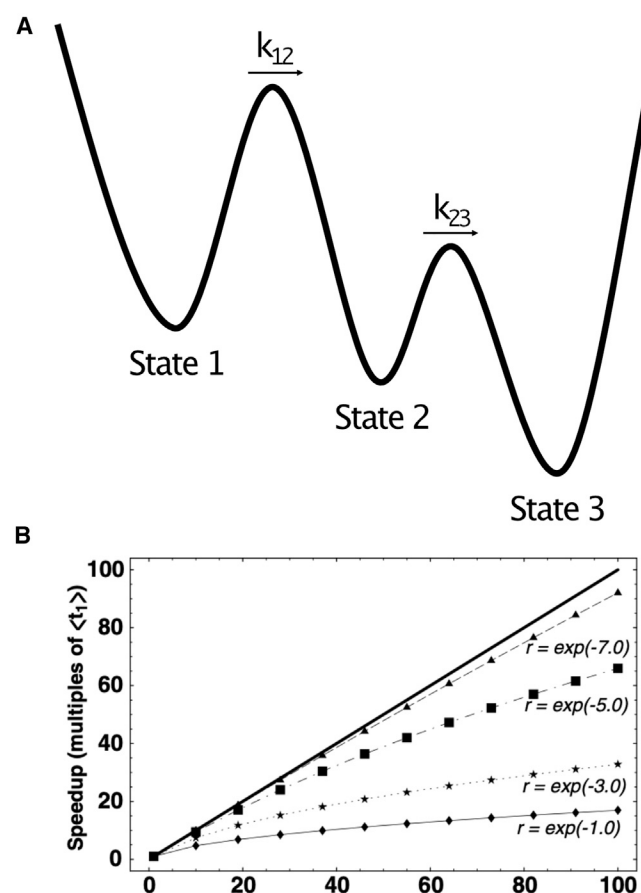


FIGURE 1 Running M simulations of length t_1 can provide as much information as running one simulation of length $M \times t_1$, providing an enormous speedup if one runs parallel simulations over a single long simulation. (A) A simple three-state landscape where k_{12} is the rate of transitioning from state 1 to state 2 and k_{23} is the rate of transitioning from state 2 to state 3. (B) Speedup versus number of simulations for this simple landscape. The key conclusion is that the more two state a system is (i.e., the more one barrier is much larger than the other), the closer one comes to linear scaling, where running M simulations of length t_1 is equivalent to running one simulation of length $M \times t_1$. Shown are plots for a range of $r = k_{12}/k_{23}$, where k_{12} is the rate constant of the slow barrier crossing, and k_{23} is the rate constant of the fast barrier crossing. Reproduced from Shirts and Pande (11).

served as one of the primary systems where scientists could connect simulations and experiments in the coming years.

Many of the early applications of Folding@home found that even the simplest systems had complex folding dynamics with multiple intermediate states, either those known experimentally or with transient intermediates with lifetimes much longer than a single simulation, breaking the applicability of the original two-state approximation regime. Together with the desire to tackle even more complex systems, this insight prompted work on the development of new methods for moving from a two-state framework for protein folding to a multistate view. In 2004, Singhal et al. published the Pande lab's first papers on Markov state models (MSMs) (15). The core idea was to build a

network model of a protein's conformational ensemble, with states (or nodes) corresponding to free energy minima where a protein tends to dwell and links representing the probability of hopping between pairs of states in a fixed time interval, called the lag time of the model. This paradigm quickly provided new insights into the folding of small proteins, like villin (16–18).

One of the major insights stemming from work on protein folding is that the free energy landscapes of many proteins have a hub-like topology, with many parallel paths leading to the folded state (5). This observation was first made for the 35-residue villin headpiece (19). Given that this protein is so small and fast folding (μ s-timescale folding), an open question was how general this behavior was. Voelz et al. then captured the first ms-timescale folding process and found the same hub-like topology (Fig. 2) (3). Furthermore, they observed that different folding pathways resembled different models of protein folding that had been proposed based on experiments performed on different systems. They suggested that mutations likely alter the relative probabilities of such pathways, changing the experimentally observed folding mechanism. Similar observations were later made for other proteins (20,21).

Misfolding and neurodegeneration

While Folding@home's first priority was to understand protein folding, understanding the role of dynamics in protein function and dysfunction was always a larger goal. One of the first applications of Folding@home in this direction was to protein misfolding diseases, like Alzheimer's and Huntington's. Disordered proteins that are difficult to study using standard structural techniques are often implicated in these diseases. In principle, simulations are just as applicable to these systems as to folded proteins. Of course, adequately sampling the relevant degrees of freedom is challenging for such heterogeneous systems. Fortunately, the same principles that have enabled Folding@home to model the heterogeneous unfolded states of proteins and their folding transitions have enabled progress on disordered proteins and misfolding. For example, Kelley et al. proposed a mechanism for how N17 of the Huntingtin protein initializes dimerization and nucleates aggregation (22). Small populations of β -sheet-rich structures were also found when studying the folding of helical proteins, suggesting that such structures may be a common feature that provides an opportunity for aggregation (20).

Simulations performed on Folding@home also provided a number of insights into A β and its role in Alzheimer's disease (AD). Lin et al. proposed that the 42-residue version of A β is more pathogenic because it stabilizes a β -hairpin that may nucleate oligomerization more than the shorter 40-residue version of A β (23). Based on this insight, Novick et al. designed variants of A β and small molecules that inhibited aggregation (24). Simulations were also used to

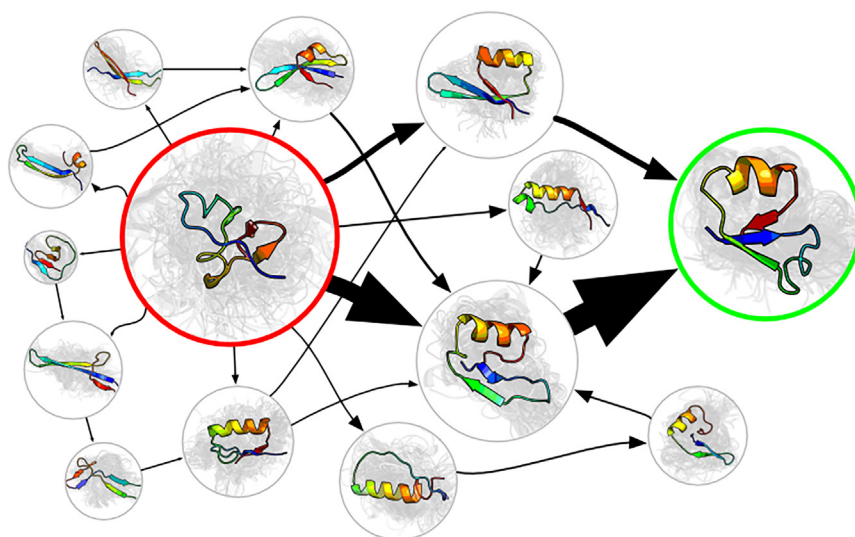


FIGURE 2 The 10 highest-flux folding pathways for the 39-residue protein NTL9, which folds on a ms timescale. State sizes are proportional to the population, arrow widths are proportional to the flux, the colored ribbon shows a representative structure, and the gray structures convey the extent of structural diversity. Reproduced from Voelz et al. (3). To see this figure in color, go online.

understand how familial mutations associated with AD change A β 's structural preferences and, ultimately, AD risk (25).

More recently, Folding@home simulations helped provide unprecedented insight into apolipoprotein E (ApoE) (26), which is the strongest genetic risk factor for AD (27). Like A β and other disordered proteins, ApoE has been difficult to study due to its propensity to form heterogeneous oligomers at very low concentrations. Simulations are complicated by the fact that ApoE has both well-folded and disordered regions. However, running milliseconds of simulations on Folding@home has proved capable of accounting for this spectrum of different degrees of heterogeneity, providing atomically detailed models that are in excellent agreement with single-molecule experiments and provide a basis for understanding how variants modulate AD risk and for designing new therapeutics.

Conformational change

In the last decade, Folding@home's focus has largely shifted from protein folding to other conformational changes of importance for protein function, dysfunction, and the design of new drugs and proteins (28–30).

One of the first focal points was understanding the relevance of fundamental concepts like induced fit and conformational selection. Induced fit posits that protein conformational changes are caused by interactions with binding partners, whereas conformational selection posits that proteins are constantly undergoing spontaneous conformational changes and that binding partners shift the relative probabilities of the different conformations that are available. Addressing this point is of practical importance because it determines what must be simulated. If conformational change dominates, then much can be learned from

simulations of isolated proteins. In contrast, if induced fit dominates, the simulations must explicitly include a protein's binding partner(s) to observe functionally relevant conformational changes. One can get a qualitative sense of the importance of these two mechanisms by simulating ligand-free proteins and asking if they sample conformations resembling ligand-bound structures. A more quantitative evaluation can be performed by simulating the binding process and comparing bound and unbound ensembles.

Much evidence supports the importance of conformational selection, making simulations of individual proteins highly informative in many cases. For example, analysis of simulations of WW domains suggested that conformational selection is at play in their binding interactions (31). These simulations did not include any of the WW domain's binding partners, so they could not directly assess the relative importance of conformational selection and induced fit. Simulations of the lysine-, arginine-, and ornithine-binding (LAO) protein and one of its substrates, arginine, directly addressed the relative roles of conformational selection and induced fit, finding a significant role for conformational selection (Fig. 3) (32). Specifically, the authors found that binding occurs via a three-state mechanism in which the protein exists in an equilibrium between an open and partially closed state. Ligand interacts favorably with the partially closed state and can then induce total closure, which is extremely rare in the absence of ligand. Later studies of the binding of the p53 TAD peptide and its protein receptor MDM2 also suggested a mixture of conformational selection and induced fit (33). Importantly, this work recognized that the dominant mechanism is dependent on the concentrations of the species and found that induced fit dominates across a wide range of concentrations in this particular case. A study of cyclic peptide binding to MDM2 found that the idea of conformational selection

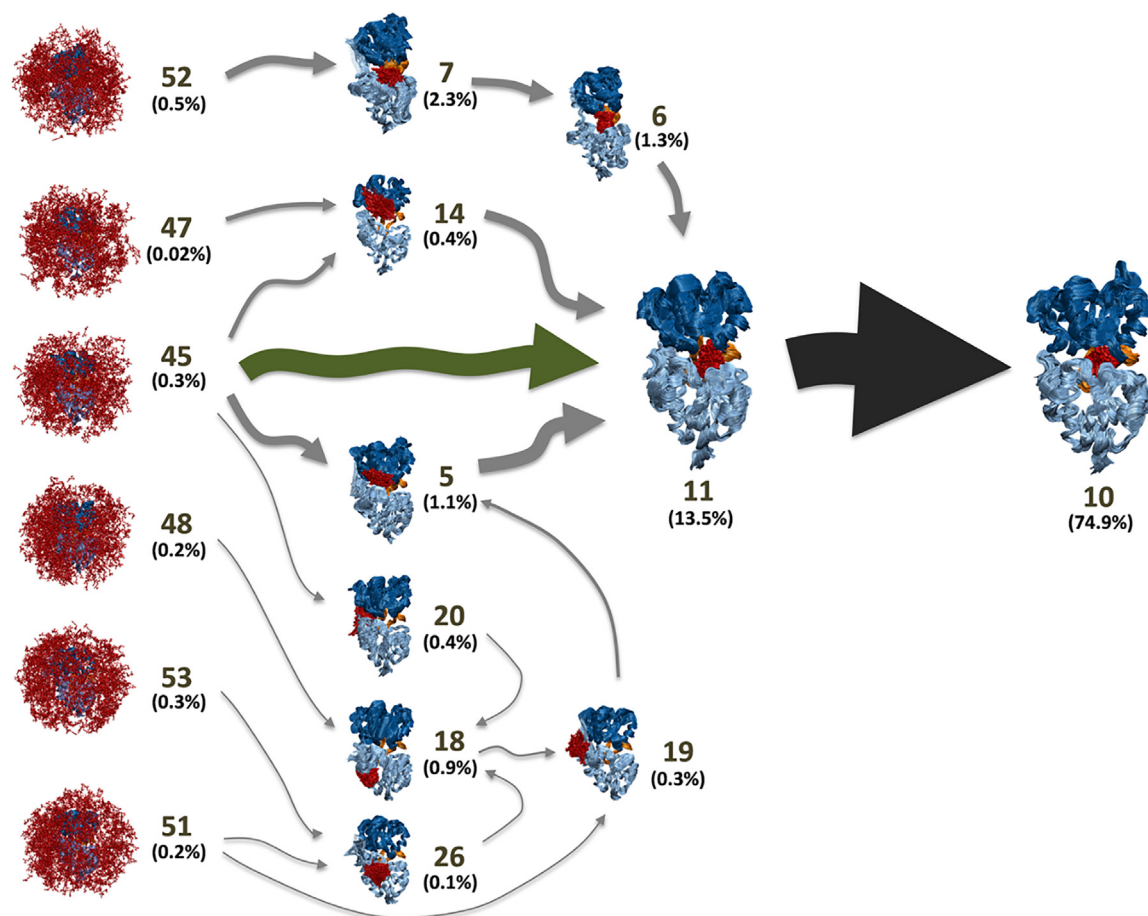


FIGURE 3 The 10 highest flux pathways from the unbound states of the lysine-, arginine-, and ornithine (LAO)-binding protein to the arginine-bound state. The arrow widths are proportional to the flux, the two lobes of the lysine-, arginine-, and ornithine-binding protein are shown in dark/light blue, and arginine is shown as red sticks. State numbers and their equilibrium populations are also shown. The conformational selection and induced fit pathways from the unbound states to the encounter complex state are shown in green and gray arrows, respectively. Reproduced from Silva et al. (32). To see this figure in color, go online.

applies just as well to foldable ligands (34). A mix of conformational selection and induced fit was also observed in the binding of a ligand to β -lactamase (35). Together, these studies suggest that conformational selection plays a sufficiently significant role that much can be learned from simulations of isolated proteins without explicitly including their binding partners in the simulation.

The value of simulating isolated proteins has borne out in many subsequent studies. For example, Sun et al. inferred how G protein-coupled receptor binding to G proteins allosterically triggers nucleotide release from a site over 30 Å away from simulations of G proteins (36). Porter et al. later used similar logic to identify conformational dynamics in the active sites of isolated myosin motor proteins that are predictive of the lifetime of the actin-bound states of myosins (37). Functionally relevant dynamics have also been observed in simulations of kinases (38,39), G protein-coupled receptors (40), enzymes like SETD8 (41), DNA binding (42), and transcription (43,44).

Methods and tools for the broader scientific community

Besides the scientific insights, some of Folding@home's greatest legacies are the many algorithms and software that have been developed and shared with the broader scientific community.

MSM methods have been one of the most important contributions (45–47). As already mentioned, the Pande lab began developing and applying ideas on how to construct and analyze these models in the mid-2000s (15,16,48), in synergy with concurrent developments in the rest of the field (49,50). These algorithms were soon incorporated into a software package, called MSMBuilder (51–53), that was adopted by many theorists for analyzing their simulation data. Later versions of the code drew heavily from another software package developed by the Folding@home community, called MDTraj (54), which provides many useful functions for operating on and analyzing simulation trajectories. The enspara software (55) also provides extra functionality,

especially for dealing with large datasets like those frequently generated on Folding@home and tools for inferring allostery and the effects of mutations (56,57). Methods have also been introduced for assessing the quality of different MSMs (58–61), defining states for complex systems (62), and accounting for memory effects (63,64). Analysis methods built on the MSM framework facilitate a wide variety of research, from understanding allosteric networks (56,65,66) to understanding the effects of mutations and drugs (57), predicting the effect of such perturbations (67), and reconciling simulations and experiments (68).

Building off the MSM framework, the Folding@home community has developed a variety of adaptive sampling algorithms for capturing rare events with far less simulation data than brute force simulations would require. The basic idea is to iterate between running simulations, building an MSM, and then using that MSM to decide where to run the next batch of simulations. The first adaptive sampling simulations were geared toward capturing protein folding and aimed to reduce the uncertainty in the slowest process observed in the simulations (69,70). It was found that adaptive sampling could capture slow processes with far less simulation data than brute force simulations by counteracting simulations' tendency to redundantly repeat the same high-probability events (71). While these methods were effective at reducing uncertainty in a model, they were less adept at discovering new conformational states. Adaptive seeding methods helped alleviate this issue by using other enhanced sampling methods to quickly explore conformational space, followed by additional simulations to ensure that adequate data were gathered from across conformational space to build a good MSM (72). Starting simulations from poorly sampled states also proved to be a useful way to promote the discovery of new conformational states (73). However, it was later found that simply choosing poorly sampled states as starting points for new simulations could also lead to pathological outcomes, like spending enormous compute resources exploring high-energy states that were of little relevance compared with higher probability regions of conformational space (74). Zimmerman et al. introduced the idea of balancing exploration-exploitation trade-offs in a goal-oriented adaptive sampling algorithm called FAST (75). This algorithm and related methods like REAP (76) have proved extremely valuable. The core idea is to balance between preferentially simulating states that optimize some progress variable (e.g., maximizing a distance or minimizing a root-mean-square deviation to a target structure) and broad exploration of conformational space. In doing so, the algorithms focus sampling on the relevant regions of conformational space while simultaneously reducing redundant sampling and avoiding dead ends. FAST simulations of systems like 263-residue β -lactamase enzymes capture slow processes with orders of magnitude less simulation time than conventional simulations (74,75). Furthermore, FAST simulations

of the 3600-residue severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike run on Azure captured enormous conformational changes (2), while an equivalent simulation time using conventional simulations on supercomputers or Anton2 only captured small fluctuations around the starting conformation (77). More discussion of the spike can be found in the section on COVID. Here, we simply emphasize that capturing the conformational changes seen with FAST simulations costing about \$70,000 on the cloud would have cost well over \$3 million with conventional simulations.

The Folding@home community has also contributed software for running individual simulations faster. For example, the team created the first petascale supercomputer by working with Sony to deploy simulation software on the Playstation (78). They also contributed code that could make use of multicore CPUs (17). Moreover, the Pande lab was one of the early groups to recognize that graphics processing units that were developed to quickly update every pixel on a computer monitor could be repurposed to quickly update the position and velocity of every atom in a protein (79). This led to the development of the widely used OpenMM software (80).

Finally, Folding@home has made numerous contributions to the evaluation and creation of force fields. Force fields are the parameters that describe interatomic interactions in a simulation. Evaluating their performance is difficult because it requires running enough simulations to explore the relevant conformational space and a means to make quantitative comparisons to experiments. The parallelism of Folding@home makes it well suited for testing many force fields. Multiple papers have developed ways of predicting NMR observables from simulations and comparing the performance of simulations conducted with multiple force fields to experiments (60,81,82). Other methods have been developed for modeling spectroscopic observables (21,26,83,84). Learning from these comparisons also led the development of new force fields (85).

Guiding experiments

An increasing focus of Folding@home, especially since Dr. Greg Bowman became director in 2018, has been using simulations and MSMs to guide experiments. Making true predictions and then confirming them in subsequent experimental tests is the holy grail of computational biophysics, requiring accurate force fields, sufficient sampling, extraction of useful insights from complex structural ensembles, and a mix of qualitative and quantitative means to predict experimental outcomes.

Early efforts focused on understanding how mutations alter that stability and activity of β -lactamase enzymes (29). For example, the M182T mutation has long been known to stabilize TEM β -lactamase, but the mechanism remained unclear, as different crystal structures of the variant

suggested different mechanisms of stabilization. Zimmerman et al. found that M182T stabilizes a helix at a key domain boundary by capping the helix, ultimately stabilizing both domains of the protein (86). To test their insight, Zimmerman et al. then proposed a number of other mutations to modulate β -lactamase's stability, built MSMs to make quantitative predictions, and then experimentally tested their results. Interestingly, biochemical intuition suggested that an M182N mutation should also stabilize β -lactamase by capping the same helix, but the MSM for this variant predicted the mutation to be neutral due to a competing destabilizing interaction. This and other predictions were confirmed experimentally, showing the power of MSMs to outperform human biochemical intuition. Related work used the same principles to understand how mutations allosterically alter the enzymes activity, inform the design of new variants to modulate activity, and perform experimental tests that confirmed the models' predictions (87,88).

Folding@home has also been making important contributions to our ability to identify and drug cryptic pockets. These pockets are absent in known crystal structures but form due to protein dynamics. They present a number of new opportunities, from providing a means to target proteins otherwise thought to lack druggable pockets to ways to allosterically enhance desirable activities, a feat that is impossible with traditional drug design given its focus on inhibiting targets by sterically blocking interactions. MSMs of β -lactamase demonstrated that simulations can capture known cryptic pockets (35). While induced fit can play a role in cryptic pocket opening, conformational selection plays a sufficiently important role that pockets can often be seen in simulations of ligand-free proteins. This finding means that simulations can be used to discover pockets without foreknowledge of ligands that bind them, and information from the simulations can then be used to design ligands. Subsequently, these methods were used to predict novel cryptic pockets in multiple β -lactamases and to design experimental tests of these pockets, which were ultimately confirmed to exist and exert allosteric control over enzyme activity (66,89). Knoverek et al. then showed that these pockets are functionally relevant, as increasing pocket opening can actually facilitate hydrolysis of some substrates (88).

Simulations on Folding@home are also helping to improve drug discovery beyond elucidating cryptic pockets. For example, Hart et al. proposed an approach to incorporate protein conformational heterogeneity into rational drug design, called Boltzmann docking, and found that it predicts affinities better than docking against single-crystal structures (87). The algorithm works by docking compounds against representative conformations of each state of an MSM and then ranking them based on a population-weighted average docking score. In addition to naturally considering a set of structures rather than having to pick a

single structure from a large simulated ensemble, the method also naturally favors compounds that bind to higher-probability conformations and therefore have to pay a lower penalty to stabilize their target conformation. Hart et al. then used this approach to design and experimentally confirm novel β -lactamase inhibitors that target a cryptic pocket found by simulation (90). Meller et al. later derived a more physically rigorous version of this algorithm, called MSM-docking, and used it to guide experimental predictions of binding affinities and experimental tests (91).

Now, these methods are being applied broadly to a number of biologically important proteins. For example, Cruz et al. predicted a cryptic pocket in an Ebola protein that was thought to lack druggable pockets based on known crystal structures and then experimentally confirmed the existence of this pocket and its allosteric control over a key RNA-binding interaction (Fig. 4) (92). Meller et al. used these tools to understand how sequence variations in myosin motors control the probability of pocket opening and thereby modulate the affinity for the allosteric inhibitor blebbistatin (91). To test this approach, they predicted the affinity of blebbistatin for a new motor and experimentally confirmed their prediction. MSMs are also being used to

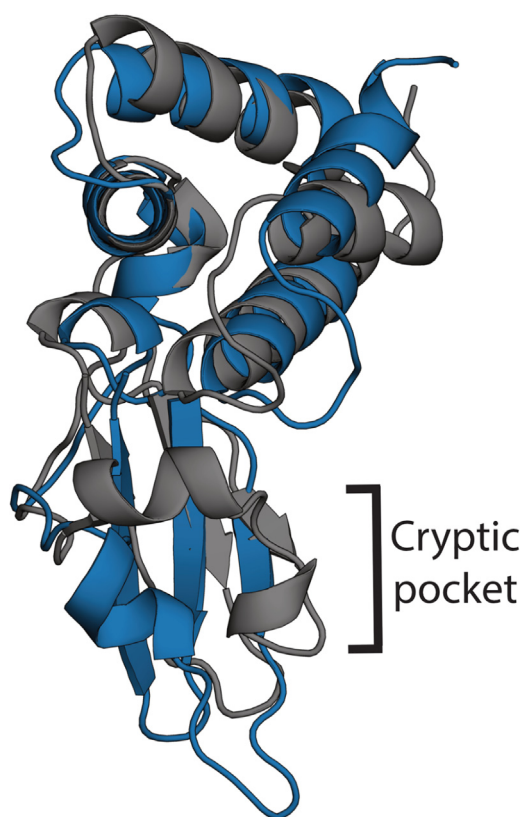


FIGURE 4 A structure with a cryptic pocket in Ebola's VP35 protein (blue) overlaid with the crystal structure (gray). The existence of this pocket and its allosteric control over RNA binding were confirmed using chemical labeling experiments. Reproduced from Cruz et al. (92). To see this figure in color, go online.

understand how mutations modulate the structural preferences of the ApoE protein and, ultimately, people's risk of developing AD (26).

COVID-19

Folding@home's strengths and past achievements positioned it to pivot quickly to address the COVID-19 pandemic. In February 2020, just before lockdowns started in the United States, the Chodera lab launched the first Folding@home simulations of SARS-CoV-2 viral proteins based on emerging structural data (93). The community's response was literally overwhelming: over the next month, Folding@home's user base grew from ~30,000 active users to over one million (2), causing server-side infrastructure to stagger under the load. This crisis was ameliorated with an outpouring of generosity from numerous tech organizations, who offered help scaling up Folding@home's infrastructure in the cloud. At its peak, the community's aggregate compute power was estimated to be at least 5-fold greater than the Summit supercomputer, which was the world's fastest traditional supercomputer at the time.

A key goal during this time was to contribute to the development of new therapeutics by exploiting Folding@home's ability to computationally test many possible drugs in parallel. Toward this end, Folding@home joined in in an international collaboration—the COVID Moonshot (<https://covid.postera.ai/covid>)—to develop a patent-free small-molecule protease inhibitor that would serve as an inexpensive therapy, as well as a prophylactic for high-risk individuals (94). Following high-throughput crystallization studies by the UK's Diamond Light Source that identified multiple molecules that bind the SARS-CoV-2 main protease (95), the COVID Moonshot project was formed to accelerate the development of potent inhibitors for clinical trials. Folding@home performed absolute and relative free energy calculations (96) on a massive scale to prioritize new molecules for synthesis and subsequent experimental tests. In principle, these calculations—which sample protein and ligand motions—are far more accurate than computational docking but are enormously expensive, generally limiting these calculations to a few dozen at a time. Using the parallelism of Folding@home, the team quickly screened tens of thousands of potential inhibitors. Many iterations of computational design and testing followed, leading to the development of novel chemical scaffolds that are now progressing toward clinical trials (97,98).

Folding@home's parallelism also enabled the scientific team to search through the viral proteome for new potential drug targets. Indeed, cryptic pockets were found across much of the SARS-CoV-2 proteome. A noteworthy example is a cryptic pocket in the protein Nsp16 (Fig. 5) (99). This pocket is shared across extant coronaviruses but is absent in human homologs of the protein. Therefore, drugs that target this site could inhibit all coronaviruses without im-

pacting similar human proteins and causing unwanted side effects.

Protein variants from different coronaviruses were also compared to understand what factors make the SARS-CoV-2 virus so infectious and point toward new ways to thwart the virus. For example, spike proteins from multiple coronaviruses were simulated to understand how differences in their conformational preferences alter the viruses' ability to enter cells and evade host immunity. MSMs built from these simulations revealed that the spike undergoes a dramatic opening motion that is far larger than was expected from cryoelectron microscopy data and that controls a trade-off between cell entry and immune evasion (Fig. 6) (2). Viruses whose spike proteins spend more time in the open state are better at cell entry, but they are also more susceptible to recognition by our immune systems. One factor that enabled SARS-CoV-2 to become a pandemic is that its spike is more closed, making it better at immune evasion. However, the virus maintains its ability to enter cells by acquiring mutations that increase the affinity between an open spike and its binding partner on host cells, angiotensin-converting enzyme 2 (ACE2). The result is that SARS-CoV-2 is better at evading our immune systems than the original SARS virus and just as good at cell entry. Subsequent viral variants have optimized these factors further.

Other simulations interrogated how existing nucleoside triphosphate analogs (like remdesivir) target viral RNA replication (100), the function of the SARS-CoV-2 envelope protein (an ion channel and membrane protein crucial to viral assembly) (101), and the design of ACE2 decoys that could block the spike from engaging host ACE2 (102).

The pandemic also exemplifies the ways that Folding@home has impact beyond its direct contributions to science. For example, many volunteers reported mental health benefits from the opportunity to take action at a time when they otherwise felt helpless. Others continue to take a keener interest in science, with outcomes ranging from continued involvement in Folding@home to pursuing science, technology, engineering, and math careers. The pandemic also accelerated trends in open science. All COVID-19 data and results generated by Folding@home are publicly available through the COVID-19 Structure and Therapeutics Hub developed jointly with the Molecular Sciences Software Institute (<http://covid.molssi.org>) to empower other researchers, and this practice is being continued as the scientific team returns to working on other problems.

Outlook

Folding@home has enabled numerous scientific insights by steadily providing the ability to simulate longer timescales, larger systems, and more numerous variants. Work on SARS-CoV-2 exemplifies the ways that many different proteins can be studied in parallel to find the best drug targets,

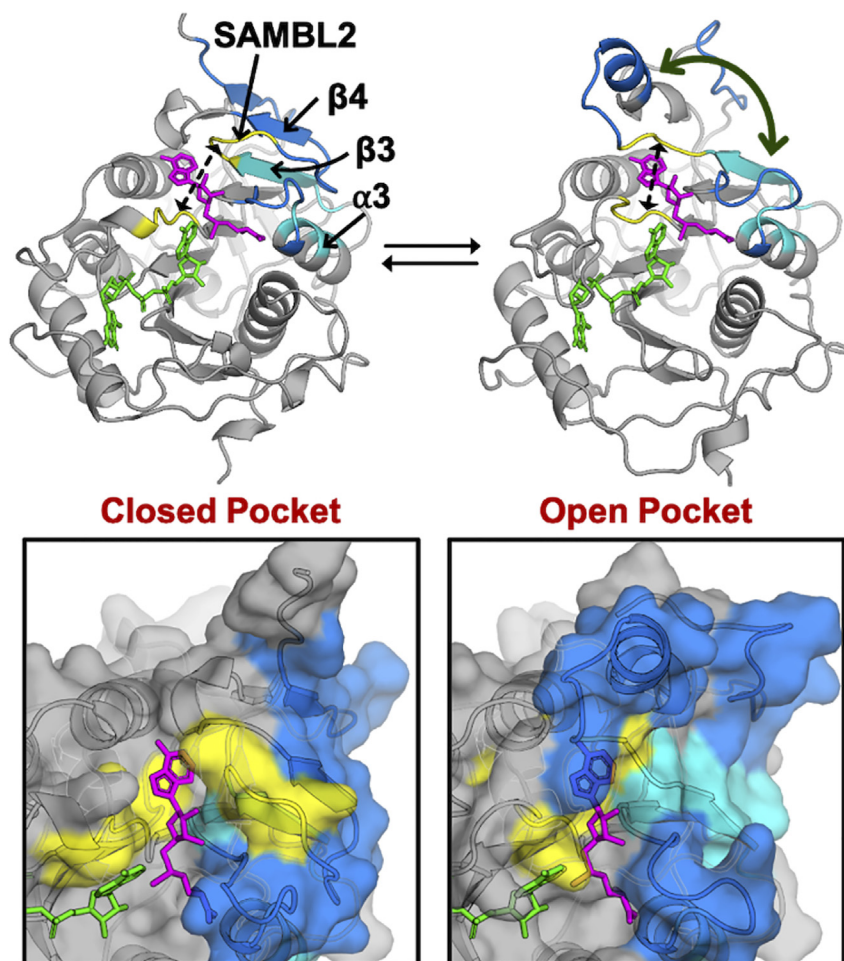


FIGURE 5 Structural states with the Nsp16 cryptic pocket closed and open, showing how pocket opening is correlated with collapse of the active site's S-adenosyl-L-methionine (SAM)-binding pocket. The insets show surface views of the closed and open pockets. Residues exposed upon pocket opening are shown in cyan, and the regions undergoing the opening motion are shown in blue. SAM is in magenta sticks, and the RNA substrate is in green sticks. Collapse of the SAM-binding pocket is measured as the distance between two loops labeled SAMBL2 and gate loop 2, shown in yellow. Reproduced from Vithani et al. (99). To see this figure in color, go online.

that many variants of a protein can be compared to understand how they function, and that many small molecules can be screened to find the best potential drugs. We are already seeing these use patterns continue into the future. The artificial intelligence-driven structure-enabled antiviral platform is working to automate the discovery of oral antivirals for pandemic preparedness. Comparative studies of large systems like myosin motors are providing new insight

into sequence-ensemble-function relationships (37,91). Ongoing work screening the folding and binding properties of peptides and peptidomimetics on a massively parallel scale (34,103,104) may help identify new therapeutics. And the same platform is being used to study many other systems.

Folding@home has given researchers compute capability a decade ahead of their peers, leading to the creation of

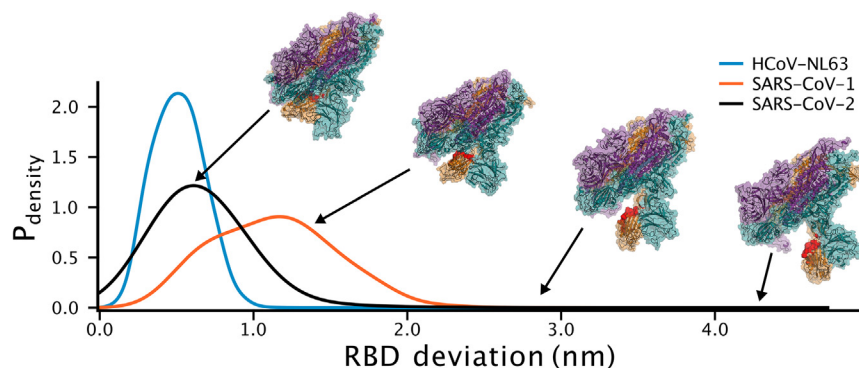


FIGURE 6 The probability distribution of spike opening for three spike homologs. Opening is quantified in terms of how far the center of mass of a receptor-binding domain deviates from its position in the closed (or down) state. The cryptic epitope for the antibody CR3022 (red) and the ACE2-binding interface are both exposed in open structures but buried in closed structures. As a result, more open spikes are better at cell entry but more susceptible to host immunity. Reproduced from Zimmerman et al. (2). To see this figure in color, go online.

novel computing paradigms (e.g., graphics-processing unit computing with OpenMM) and algorithms for today's cloud-based compute (e.g., MSMs). Looking ahead to the exascale era and beyond, Folding@home is poised to be an essential tool for research, serving as a general platform for addressing pressing problems in both basic science and human disease while remaining ready to rapidly respond to emergent health threats. Citizen scientists, unite!

To participate, visit <https://foldingathome.org/start-folding/>.

AUTHOR CONTRIBUTIONS

V.A.V., V.S.P., and G.R.B. wrote the paper.

ACKNOWLEDGMENTS

Thanks to the entire Folding@home community for all their contributions. G.R.B. was funded by NSF MCB 2218156, NIH R01GM124007, NIH AG067194, and a Packard Fellowship. V.A.V. is supported by NIH R01GM123296. Thanks to John Chodera for helpful comments.

DECLARATION OF INTERESTS

G.R.B. is a cofounder and board member of Decrypt Biomedicine. V.S.P. is the Managing Partner of a16z BioHealth and is deeply involved in a rapidly evolving set of companies.

REFERENCES

- Shirts, M., and V. S. Pande. 2000. COMPUTING: screen savers of the world unite! *Science*. 290:1903–1904.
- Zimmerman, M. I., J. R. Porter, ..., G. R. Bowman. 2021. SARS-CoV-2 simulations go exascale to predict dramatic spike opening and cryptic pockets across the proteome. *Nat. Chem.* 13:651–659.
- Voelz, V. A., G. R. Bowman, ..., V. S. Pande. 2010. Molecular simulation of ab initio protein folding for a millisecond folder NTL9 (1–39). *J. Am. Chem. Soc.* 132:1526–1528.
- Snow, C. D., E. J. Sorin, ..., V. S. Pande. 2005. How well can simulation predict protein folding kinetics and thermodynamics? *Annu. Rev. Biophys. Biomol. Struct.* 34:43–69.
- Bowman, G. R., V. A. Voelz, and V. S. Pande. 2011. Taming the complexity of protein folding. *Curr. Opin. Struct. Biol.* 21:4–11.
- Lane, T. J., D. Shukla, ..., V. S. Pande. 2013. To milliseconds and beyond: challenges in the simulation of protein folding. *Curr. Opin. Struct. Biol.* 23:58–65.
- Sorin, E. J., M. A. Engelhardt, ..., V. S. Pande. 2002. RNA simulations: probing hairpin unfolding and the dynamics of a GNRA tetraloop. *J. Mol. Biol.* 317:493–506.
- Bowman, G., X. Huang, ..., V. S. Pande. 2008. Structural insight into RNA hairpin folding intermediates. *J. Am. Chem. Soc.* 130:9676–9678.
- Kasson, P. M., and V. S. Pande. 2007. Control of membrane fusion mechanism by lipid composition: predictions from ensemble molecular dynamics. *PLoS Comput. Biol.* 3:e220.
- Sorin, E. J., and V. S. Pande. 2006. Nanotube confinement denatures protein helices. *J. Am. Chem. Soc.* 128:6316–6317.
- Shirts, M. R., and V. S. Pande. 2001. Mathematical analysis of coupled parallel simulations. *Phys. Rev. Lett.* 86:4983–4987.
- Zagrovic, B., E. J. Sorin, and V. Pande. 2001. beta-hairpin folding simulations in atomistic detail using an implicit solvent model. *J. Mol. Biol.* 313:151–169.
- Snow, C. D., H. Nguyen, ..., M. Gruebele. 2002. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature*. 420:102–106.
- Zagrovic, B., C. D. Snow, ..., V. S. Pande. 2002. Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing. *J. Mol. Biol.* 323:927–937.
- Singhal, N., C. Snow, and V. S. Pande. 2004. Using path sampling to build better Markovian state models: predicting the folding rate and mechanism of a tryptophan zipper beta hairpin. *J. Chem. Phys.* 121:415–425.
- Jayachandran, G., V. Vishal, and V. S. Pande. 2006. Using massively parallel simulation and Markovian models to study protein folding: examining the dynamics of the villin headpiece. *J. Chem. Phys.* 124:164902.
- Ensign, D. L., P. M. Kasson, and V. S. Pande. 2007. Heterogeneity even at the speed limit of folding: large-scale molecular dynamics study of a fast-folding variant of the villin headpiece. *J. Mol. Biol.* 374:806–816.
- Bowman, G. R., K. A. Beauchamp, ..., V. S. Pande. 2009. Progress and challenges in the automated construction of Markov state models for full protein systems. *J. Chem. Phys.* 131:124101.
- BOWMAN, G., and V. Pande. 2010. Protein folded states are kinetic hubs. *Proc. Natl. Acad. Sci. USA*. 107:10890–10895.
- Bowman, G. R., V. A. Voelz, and V. S. Pande. 2011. Atomistic folding simulations of the five-helix bundle protein λ (6–85). *J. Am. Chem. Soc.* 133:664–667.
- Voelz, V. A., M. Jäger, ..., V. S. Pande. 2012. Slow unfolded-state structuring in Acyl-CoA binding protein folding revealed by simulation and experiment. *J. Am. Chem. Soc.* 134:12565–12577.
- Kelley, N. W., X. Huang, ..., V. S. Pande. 2009. The predicted structure of the headpiece of the Huntingtin protein and its implications on Huntingtin aggregation. *J. Mol. Biol.* 388:919–927.
- Lin, Y.-S., G. R. Bowman, ..., V. S. Pande. 2012. Investigating how peptide length and a pathogenic mutation modify the structural ensemble of amyloid beta monomer. *Biophys. J.* 102:315–324.
- Novick, P. A., D. H. Lopes, ..., V. S. Pande. 2012. Design of beta-amyloid aggregation inhibitors from a predicted structural motif. *J. Med. Chem.* 55:3002–3010.
- Lin, Y. S., and V. S. Pande. 2012. Effects of familial mutations on the monomer structure of Abeta(4)(2). *Biophys. J.* 103:L47–L49.
- Stuchell-Brereton, M. D., M. I. Zimmerman, ..., A. Soranno. 2023. Apolipoprotein E4 has extensive conformational heterogeneity in lipid-free and lipid-bound forms. *Proc. Natl. Acad. Sci. USA*. 120:e2215371120.
- Raber, J., Y. Huang, and J. W. Ashford. 2004. ApoE genotype accounts for the vast majority of AD risk and AD pathology. *Neurobiol. Aging*. 25:641–650.
- Shukla, D., C. X. Hernández, ..., V. S. Pande. 2015. Markov state models provide insights into dynamic modulation of protein function. *Acc. Chem. Res.* 48:414–422.
- Knoverek, C. R., G. K. Amarasinghe, and G. R. Bowman. 2019. Advanced methods for Accessing protein shape-shifting present new therapeutic opportunities. *Trends Biochem. Sci.* 44:351–364.
- Wang, X., I. C. Unarta, ..., X. Huang. 2021. Elucidating molecular mechanisms of functional conformational changes of proteins via Markov state models. *Curr. Opin. Struct. Biol.* 67:69–77.
- Lane, T. J., G. R. Bowman, ..., V. S. Pande. 2011. Markov state model reveals folding and functional dynamics in ultra-long MD trajectories. *J. Am. Chem. Soc.* 133:18413–18419.
- Silva, D.-A., G. R. Bowman, ..., X. Huang. 2011. A role for both conformational selection and induced fit in ligand binding by the Lao protein. *PLoS Comput. Biol.* 7:e1002054.
- Zhou, G., G. A. Pantelopulos, ..., V. A. Voelz. 2017. Bridging microscopic and macroscopic mechanisms of p53-MDM2 binding with kinetic network models. *Biophys. J.* 113:785–793.

34. Ge, Y., S. Zhang, ..., V. A. Voelz. 2021. Solution-state preorganization of cyclic beta-hairpin ligands determines binding mechanism and affinities for MDM2. *J. Chem. Inf. Model.* 61:2353–2367.
35. Bowman, G. R., and P. L. Geissler. 2012. Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *Proc. Natl. Acad. Sci. USA.* 109:11681–11686.
36. Sun, X., S. Singh, ..., G. R. Bowman. 2018. Simulation of spontaneous G protein activation reveals a new intermediate driving GDP unbinding. *Elife.* 7:e38465.
37. Porter, J. R., A. Meller, ..., G. R. Bowman. 2020. Conformational distributions of isolated myosin motor domains encode their mechanochemical properties. *Elife.* 9:e55132.
38. Shukla, D., Y. Meng, ..., V. S. Pande. 2014. Activation pathway of Src kinase reveals intermediate states as targets for drug design. *Nat. Commun.* 5:3397.
39. Sultan, M. M., G. Kiss, and V. S. Pande. 2018. Towards simple kinetic models of functional dynamics for a kinase subfamily. *Nat. Chem.* 10:903–909.
40. Kohlhoff, K. J., D. Shukla, ..., V. S. Pande. 2014. Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways. *Nat. Chem.* 6:15–21.
41. Chen, S., R. P. Wiewiora, ..., M. Luo. 2019. The dynamic conformational landscape of the protein methyltransferase SETD8. *Elife.* 8:e45403.
42. Novack, D., L. Qian, ..., R. H. G. Baxter. 2022. Oncogenic mutations in the DNA-binding domain of FOXO1 that disrupt folding: quantitative insights from experiments and molecular simulations. *Biochemistry.* 61:1669–1682.
43. Unarta, I. C., L. Zhu, ..., X. Huang. 2018. Molecular mechanisms of RNA polymerase II transcription elongation elucidated by kinetic network models. *Curr. Opin. Struct. Biol.* 49:54–62.
44. Silva, D.-A., D. R. Weiss, ..., X. Huang. 2014. Millisecond dynamics of RNA polymerase II translocation at atomic resolution. *Proc. Natl. Acad. Sci. USA.* 111:7665–7670.
45. Pande, V. S., K. Beauchamp, and G. R. Bowman. 2010. Everything you wanted to know about Markov State Models but were afraid to ask. *Methods.* 52:99–105.
46. G. R. Bowman, V. S. Pande, and F. Noé, eds 2014. *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation* Springer.
47. Husic, B. E., and V. S. Pande. 2018. Markov state models: from an Art to a science. *J. Am. Chem. Soc.* 140:2386–2396.
48. Chodera, J. D., N. Singhal, ..., W. C. Swope. 2007. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Chem. Phys.* 126:155101.
49. Noé, F., and S. Fischer. 2008. Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr. Opin. Struct. Biol.* 18:154–162.
50. Swope, W. C., J. W. Pitera, and F. Suits. 2004. Describing protein folding kinetics by molecular dynamics simulations. 1. Theory. *J. Phys. Chem. B.* 108:6571–6581.
51. Bowman, G. R., X. Huang, and V. S. Pande. 2009. Using generalized ensemble simulations and Markov state models to identify conformational states. *Methods.* 49:197–201.
52. Beauchamp, K. A., G. R. Bowman, ..., V. S. Pande. 2011. MSMBuilder2: modeling conformational dynamics at the picosecond to millisecond scale. *J. Chem. Theory Comput.* 7:3412–3419.
53. Harrigan, M. P., M. M. Sultan, ..., V. S. Pande. 2017. MSMBuilder: statistical models for biomolecular dynamics. *Biophys. J.* 112:10–15.
54. McGibbon, R. T., K. A. Beauchamp, ..., V. S. Pande. 2015. MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.* 109:1528–1532.
55. Porter, J. R., M. I. Zimmerman, and G. R. Bowman. 2019. Enspara: modeling molecular ensembles with scalable data structures and parallel computing. *J. Chem. Phys.* 150:044108.
56. Singh, S., and G. R. Bowman. 2017. Quantifying allosteric communication via both concerted structural changes and conformational disorder with CARDS. *J. Chem. Theory Comput.* 13:1509–1517.
57. Ward, M. D., M. I. Zimmerman, ..., G. R. Bowman. 2021. Deep learning the structural determinants of protein biochemical properties by comparing structural ensembles with DiffNets. *Nat. Commun.* 12:3023.
58. Bacallado, S., and J. Chodera. 2009. Bayesian comparison of Markov models of molecular dynamics with detailed balance constraint. *J. Chem. Phys.* 131:045106.
59. Bowman, G. R. 2012. Improved coarse-graining of Markov state models via explicit consideration of statistical uncertainty. *J. Chem. Phys.* 137:134111.
60. Bowman, G. R., and P. L. Geissler. 2014. Extensive conformational heterogeneity within protein cores. *J. Phys. Chem. B.* 118:6417–6423.
61. Bowman, G. R., L. Meng, and X. Huang. 2013. Quantitative comparison of alternative methods for coarse-graining biological networks. *J. Chem. Phys.* 139:121905–121909.
62. Liu, S., L. Zhu, ..., X. Huang. 2017. Adaptive partitioning by local density-peaks: an efficient density-based clustering algorithm for analyzing molecular dynamics trajectories. *J. Comput. Chem.* 38:152–160.
63. Cao, S., A. Montoya-Castillo, ..., X. Huang. 2020. On the advantages of exploiting memory in Markov state models for biomolecular dynamics. *J. Chem. Phys.* 153:014105.
64. Gu, H., W. Wang, ..., X. Huang. 2022. RPnet: a reverse-projection-based neural network for coarse-graining metastable conformational states for protein dynamics. *Phys. Chem. Chem. Phys.* 24:1462–1474.
65. Brandman, R., Y. Brandman, and V. S. Pande. 2012. A-site residues move independently from P-site residues in all-atom molecular dynamics simulations of the 70S bacterial ribosome. *PLoS One.* 7:e29377.
66. Porter, J. R., K. E. Moeder, ..., G. R. Bowman. 2019. Cooperative changes in solvent exposure identify cryptic pockets, switches, and allosteric coupling. *Biophys. J.* 116:818–830.
67. Voelz, V. A., G. Zhou, and H. Wan. 2016. A maximum-caliber approach to predicting perturbed folding kinetics due to mutations. *J. Chem. Theory Comput.* 12:5768–5776.
68. Voelz, V. A., Y. Ge, and R. M. Raddi. 2021. Reconciling simulations and experiments with BiCEPs: a review. *Front. Mol. Biosci.* 8:661520.
69. Singhal, N., and V. S. Pande. 2005. Error analysis and efficient sampling in Markovian state models for molecular dynamics. *J. Chem. Phys.* 123:204909.
70. Hinrichs, N. S., and V. S. Pande. 2007. Calculation of the distribution of eigenvalues and eigenvectors in Markovian state models for molecular dynamics. *J. Chem. Phys.* 126:244101.
71. Bowman, G. R., D. L. Ensign, and V. S. Pande. 2010. Enhanced modeling via network theory: adaptive sampling of Markov state models. *J. Chem. Theory Comput.* 6:787–794.
72. Huang, X., G. R. Bowman, ..., V. S. Pande. 2009. Rapid equilibrium sampling initiated from nonequilibrium data. *Proc. Natl. Acad. Sci. USA.* 106:19765–19769.
73. Weber, J. K., and V. S. Pande. 2011. Characterization and rapid sampling of protein folding Markov state model topologies. *J. Chem. Theory Comput.* 7:3405–3411.
74. Zimmerman, M. I., J. R. Porter, ..., G. R. Bowman. 2018. Choice of adaptive sampling strategy impacts state discovery, transition probabilities, and the Apparent mechanism of conformational changes. *J. Chem. Theory Comput.* 14:5459–5475.
75. Zimmerman, M. I., and G. R. Bowman. 2015. FAST conformational searches by balancing exploration/exploitation trade-offs. *J. Chem. Theory Comput.* 11:5747–5757.
76. Shamsi, Z., K. J. Cheng, and D. Shukla. 2018. Reinforcement learning based adaptive sampling: REAPing rewards by exploring protein conformational landscapes. *J. Phys. Chem. B.* 122:8386–8395.

77. Casalino, L., Z. Gaieb, ..., R. E. Amaro. 2020. Beyond shielding: the roles of glycans in the SARS-CoV-2 spike protein. *ACS Cent. Sci.* 6:1722–1734.
78. Luttmann, E., and D. Ensign. 2009. Accelerating molecular dynamic simulation on the cell processor and Playstation 3. *J. Comput. Chem.* 30:268–274.
79. Ensign, D., C. Bruns, and V. Pande. 2009. Accelerating molecular dynamic simulation on graphics processing units. *J. Comput. Chem.* 30:864–872.
80. Eastman, P., M. S. Friedrichs, ..., V. S. Pande. 2013. OpenMM 4: a reusable, extensible, hardware independent library for high performance molecular simulation. *J. Chem. Theory Comput.* 9:461–469.
81. Beauchamp, K. A., Y.-S. Lin, ..., V. S. Pande. 2012. Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements. *J. Chem. Theory Comput.* 8:1409–1414.
82. Bowman, G. R. 2016. Accurately modeling nanosecond protein dynamics requires at least microseconds of simulation. *J. Comput. Chem.* 37:558–566.
83. Beauchamp, K., and D. Ensign. 2011. Quantitative comparison of villin headpiece subdomain simulations and triplet–triplet energy transfer experiments. *Proc. Natl. Acad. Sci. USA.* 108:12734–12739.
84. Z Cui, R., D. A. Silva, ..., X. Huang. 2012. Bridging the gap between optical spectroscopic experiments and computer simulations for fast protein folding dynamics. *Curr. Phys. Chem.* 2:45–58.
85. Wang, L. P., T. J. Martinez, and V. S. Pande. 2014. Building force fields: an automatic, systematic, and reproducible approach. *J. Phys. Chem. Lett.* 5:1885–1891.
86. Zimmerman, M. I., K. M. Hart, ..., G. R. Bowman. 2017. Prediction of new stabilizing mutations based on mechanistic insights from Markov state models. *ACS Cent. Sci.* 3:1311–1321.
87. Hart, K. M., C. M. W. Ho, ..., G. R. Bowman. 2016. Modelling proteins' hidden conformations to predict antibiotic resistance. *Nat. Commun.* 7:12965.
88. Knovere, C. R., U. L. Mallimadugula, ..., G. R. Bowman. 2021. Opening of a cryptic pocket in beta-lactamase increases penicillinase activity. *Proc. Natl. Acad. Sci. USA.* 118:e2106473118.
89. Bowman, G. R., E. R. Bolin, ..., S. Marqusee. 2015. Discovery of multiple hidden allosteric sites by combining Markov state models and experiments. *Proc. Natl. Acad. Sci. USA.* 112:2734–2739.
90. Hart, K. M., K. E. Moeder, ..., G. R. Bowman. 2017. Designing small molecules to target cryptic pockets yields both positive and negative allosteric modulators. *PLoS One.* 12:e0178678.
91. Meller, A., J. M. Lotthammer, ..., G. R. Bowman. 2023. Drug specificity and affinity are encoded in the probability of cryptic pocket opening in myosin motor domains. *Elife.* 12:e83602.
92. Cruz, M. A., T. E. Frederick, ..., G. R. Bowman. 2022. A cryptic pocket in Ebola VP35 allosterically controls RNA binding. *Nat. Commun.* 13:2269.
93. Jin, Z., X. Du, ..., H. Yang. 2020. Structure of M(pro) from SARS-CoV-2 and discovery of its inhibitors. *Nature.* 582:289–293.
94. Chodera, J., A. A. Lee, ..., F. von Delft. 2020. Crowdsourcing drug discovery for pandemics. *Nat. Chem.* 12:581. <https://doi.org/10.1038/s41557-020-0496-2>.
95. Alice Douangamath, D. F., G. Paul, ..., J. Brandaõ-Neto. 2020. Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. Preprint at bioRxiv. <https://doi.org/10.1101/2020.05.27.118117>.
96. Rizzi, A., T. Jensen, ..., J. D. Chodera. 2020. The SAMPL6 SAMPLing challenge: assessing the reliability and efficiency of binding free energy calculations. *J. Comput. Aided Mol. Des.* 34:601–633.
97. Morris, A., W. McCorkindale, ..., A. A. Lee. 2021. Discovery of SARS-CoV-2 main protease inhibitors using a synthesis-directed de novo design model. *Chem. Commun.* 57:5909–5912.
98. Hagit Achdout, A. A., E. Bar-David, ..., A. von Delft. 2021. Open science discovery of oral non-covalent SARS-CoV-2 main protease inhibitor therapeutics. Preprint at bioRxiv. <https://doi.org/10.1101/2020.10.29.339317>.
99. Vithani, N., M. D. Ward, ..., G. R. Bowman. 2021. SARS-CoV-2 Nsp16 activation mechanism and a cryptic pocket with pan-coronavirus antiviral potential. *Biophys. J.* 120:2880–2889.
100. Choy, K. T., A. Y. L. Wong, ..., H. L. Yen. 2020. Remdesivir, lopinavir, emetine, and homoharringtonine inhibit SARS-CoV-2 replication in vitro. *Antiviral Res.* 178:104786.
101. Mehregan, A., S. Pérez-Conesa, ..., L. Delemotte. 2022. Probing effects of the SARS-CoV-2 E protein on membrane curvature and intracellular calcium. *Biochim. Biophys. Acta. Biomembr.* 1864:183994.
102. Zhang, L., S. Dutta, ..., A. B. Malik. 2022. Engineered ACE2 decoy mitigates lung injury and death induced by SARS-CoV-2 variants. *Nat. Chem. Biol.* 18:342–351.
103. Razavi, A. M., W. M. Wuest, and V. A. Voelz. 2014. Computational screening and selection of cyclic peptide hairpin mimetics by molecular simulation and kinetic network models. *J. Chem. Inf. Model.* 54:1425–1432.
104. Ge, Y., B. L. Kier, ..., V. A. Voelz. 2017. Computational and experimental evaluation of designed beta-cap hairpins using molecular simulations and kinetic network models. *J. Chem. Inf. Model.* 57:1609–1620.