Exploring and Learning the Universe of Protein Allostery Using Artificial Intelligence Augmented Biophysical and Computational Approaches

Steve Agajanian,

Keck Center for Science and Engineering, Graduate Program in Computational and Data Sciences, Schmid College of Science and Technology, Chapman University, Orange, California 92866, United States

Mohammed Alshahrani,

Keck Center for Science and Engineering, Graduate Program in Computational and Data Sciences, Schmid College of Science and Technology, Chapman University, Orange, California 92866, United States

Fang Bai,

Shanghai Institute for Advanced Immunochemical Studies, School of Life Science and Technology and Information Science and Technology, Shanghai Tech University, Shanghai 201210, China

Peng Tao,

Department of Chemistry, Center for Research Computing, Center for Drug Discovery, Design, and Delivery (CD4), Southern Methodist University, Dallas, Texas 75205, United States

Gennady M. Verkhivker

Keck Center for Science and Engineering, Graduate Program in Computational and Data Sciences, Schmid College of Science and Technology, Chapman University, Orange, California 92866, United States; Department of Biomedical and Pharmaceutical Sciences, Chapman University School of Pharmacy, Irvine, California 92618, United States

Abstract

Allosteric mechanisms are commonly employed regulatory tools used by proteins to orchestrate complex biochemical processes and control communications in cells. The quantitative understanding and characterization of allosteric molecular events are among major challenges in modern biology and require integration of innovative computational experimental approaches to obtain atomistic-level knowledge of the allosteric states, interactions, and dynamic conformational landscapes. The growing body of computational and experimental studies empowered by emerging artificial intelligence (AI) technologies has opened up new paradigms for exploring and learning the universe of protein allostery from first principles. In this review we analyze

Corresponding Author Gennady M. Verkhivker – Keck Center for Science and Engineering, Graduate Program in Computational and Data Sciences, Schmid College of Science and Technology, Chapman University, Orange, California 92866, United States; Department of Biomedical and Pharmaceutical Sciences, Chapman University School of Pharmacy, Irvine, California 92618, United States; verkhivk@chapman.edu; Fax: 714-532-6048.

recent developments in high-throughput deep mutational scanning of allosteric protein functions; applications and latest adaptations of Alpha-fold structural prediction methods for studies of protein dynamics and allostery; new frontiers in integrating machine learning and enhanced sampling techniques for characterization of allostery; and recent advances in structural biology approaches for studies of allosteric systems. We also highlight recent computational and experimental studies of the SARS-CoV-2 spike (S) proteins revealing an important and often hidden role of allosteric regulation driving functional conformational changes, binding interactions with the host receptor, and mutational escape mechanisms of S proteins which are critical for viral infection. We conclude with a summary and outlook of future directions suggesting that AI-augmented biophysical and computer simulation approaches are beginning to transform studies of protein allostery toward systematic characterization of allosteric landscapes, hidden allosteric states, and mechanisms which may bring about a new revolution in molecular biology and drug discovery.

Graphical Abstract
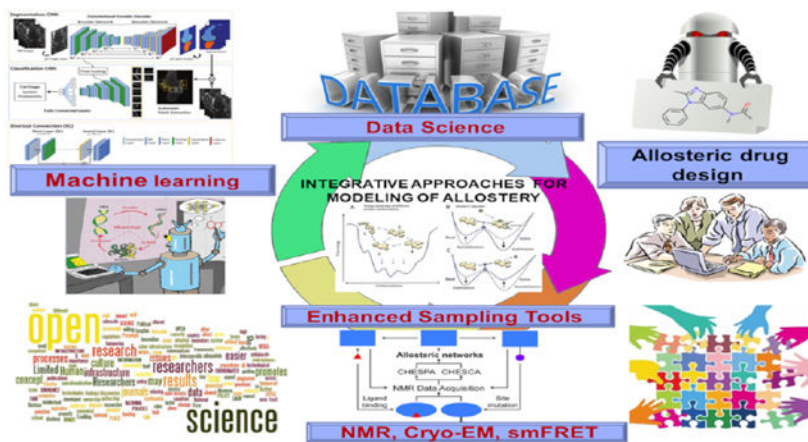
Keywords

allosteric mechanisms; artificial intelligence; protein allostery; first principles; high-throughput deep mutational scanning; allosteric drug design; machine learning; structural prediction methods; SARS-CoV-2

INTRODUCTION

Allosteric mechanisms are often used to control the activity of enzymes, ion channels, and other proteins, and are essential for the regulation of metabolic pathways, signal transduction, and other cellular processes. Allosteric regulation can also be used to regulate gene expression, as well as the production of hormones, neurotransmitters, and other molecules.[1-5] In general, allosteric interactions involve the binding of a ligand to a protein at a site other than its active site, causing a cascade of conformational changes and/or dynamic rearrangements in the system that affect the protein's activity as a result. Despite significant research efforts and continuous progress in understanding the diversity and complexity of allosteric molecular events, the interplay and balance of thermodynamic

and kinetic factors underlying molecular mechanisms of protein allostery are often difficult to monitor and characterize due to the dynamic nature of these processes and presence of short-lived hidden allosteric states involved in the regulation. Allosterically regulated proteins employ diverse molecular mechanisms to propagate various perturbations such as ligand binding or mutations, but the allosteric phenomenon is believed to be primarily driven by a thermodynamic principle that binding of an effector ligand stabilizes the active state over the inactive state and removing the effector ligand reverses this effect.[6-8] A conformational change in one state can affect the stability of other states, leading to altered binding and/or catalytic activities. The dynamic equilibrium between the inactive and active allosteric protein states can be often affected and selectively modulated through activated mutations, by post-translational modifications, and via binding with allosteric modulators and proteins. To understand the underlying principles of allosteric regulation, it is important to characterize the thermodynamic, structural, and dynamical properties of the different conformational states, as well as their interconversion pathways. While mechanistic studies of allosteric regulation are often focused on thermodynamic characterization of the functional states and their equilibrium, there has been an increasing realization of the critical role of the intrinsic protein dynamics in driving allosteric events through redistributions of dynamically modulated functional motions rather than population shifts involving appreciable structural transformations.[9,10] These allosteric models have explained the functional interplay between allosteric effects and conformational dynamics in a variety of dynamic protein systems. Hierarchical approaches combined multiscale equilibrium and nonequilibrium simulations with biophysical experiments to characterize remodeling of the free energy landscapes, detect allosteric functional states, and dissect signal transmission mechanisms.[11-19] It was proposed that population-shift based structural allostery and dynamically driven allostery that are often discussed as limiting scenarios of allosteric mechanisms and long-range communication can coexist and operate synchronously to adapt the protein free-energy landscape to incoming signals.

Probing and understanding the effect of perturbations lies at the core of many fundamental challenges and technologies of modern biology including allosteric phenomenon.[20] Perturbation-based physical approaches[21-24] emphasize the importance of simulated forces for probing of protein dynamics and prediction of phenotypic responses in complex biological systems. Combined with biophysical simulations and dynamic network models of proteins, these approaches can provide insightful mechanistic details of the underlying molecular mechanisms, quantify the protein response to various perturbations, and guide the identification of allosteric interactions, regulatory sites, and long-range communications.[25-27] To describe the multipartite organization and dynamic nature of biological systems regulated by allosteric regulatory events, the information-based theory of signal propagation[28-30] and dynamic network flow models that operate through a stochastic walk on the dynamics of the network[31-33] have been developed revealing details of multiscale dynamic relationships and the network community structure associated with functionally relevant protein changes.

Stochastic Markov state models (MSMs) have emerged as a robust and physically rigorous framework for characterization of hidden allosteric states, detection of cryptic allosteric pockets, and describing the kinetics of transitions between functional states during allosteric

events.[34-38] Combined with molecular dynamics (MD) simulations, MSM approaches can provide detailed network connectivity maps of states on the free energy landscape and estimate the effect of allosteric perturbations on the conformational equilibrium and kinetics of allosteric transitions.

Another challenge in quantitatively characterizing allosteric proteins is understanding the underlying mechanisms by which they respond to external signals. Allosteric proteins interact with their environment in complex ways, and the precise details of these interactions are often not well-understood. In order to accurately measure and characterize the behavior of allosteric proteins, it is important to gain a better understanding of the underlying mechanisms of their behavior. Current techniques for studying allosteric proteins are often limited in their ability to capture the full range of dynamic behavior exhibited by allosteric proteins. The development of new and improved tools for studying allosteric proteins is a key challenge in quantitatively characterizing these dynamic systems. The interdisciplinary structural biology strategies that exploited synergies between X-ray high-throughput crystallography, cryo-electron microscopy (cryo-EM), nuclear magnetic resonance spectroscopy (NMR), biophysical approaches, and multiscale computational methods are beginning to show a considerable potential in addressing some of these challenges and uncovering the invisible dynamic aspects of allosteric protein functions at the atomistic level.

This review is focused on a critical analysis of the latest developments in the field, marked by the emergence of innovative computational and experimental approaches that can dissect important principles of allosteric regulation and advance atomistic characterization of allosteric states, interactions and mechanisms from a unified perspective. In the next chapters we discuss recent developments in deep mutational scanning and mapping of allosteric energy landscapes; applications of Alpha-fold structural prediction methods for studies of protein dynamics and allostery; new developments in integrating enhanced sampling techniques and machine learning (ML) for characterization of dynamics and allostery; and recent advances in the experimental structural biology and biophysical approaches for studies of allosteric systems and regulatory mechanisms. We also highlight recent computational and experimental studies of SARS-CoV-2 spike proteins revealing complex dynamics and allosteric mechanisms underlying functional activities and virus transmission as well as integrative studies that discovered and validated previously unknown allosteric cryptic sites and allosteric modulators. We conclude with the outlook and future directions presenting our perspective on future developments in the field and speculate what methods and sources of information may be leveraged in the future to develop a unified framework for modeling of protein dynamics and allostery.

DEEP MUTATIONAL SCANNING AND ALLOSTERY: HIGH-THROUGHPUT BIOCHEMICAL TOOLS PARTNER WITH SIMULATIONS AND AI FOR MAPPING OF ALLOSTERIC LANDSCAPES AND REGULATORY HOTSPOTS

To probe principles of allostery, molecular mechanisms of allosteric regulation must be investigated for protein systems where allosteric signatures are intimately linked with

phenotypic responses that can be identified in biophysical studies. The recent biochemical studies extensively exploited advances in deep mutational scanning (DMS) methodology to map allosteric energy landscapes, investigate the molecular nature of allostery at the residue level, and identify the allosteric hotspots or residues critical for allosteric signaling.[39-43] The DMS approach has been a powerful tool for examining allosteric effects by systematically measuring the impact of mutational perturbations on various phenotypes using high-throughput experiments.[41-43]

A general approach for quantifying mutational effects for multiple molecular phenotypes using multidimensional DMS enabled a comprehensive characterization of allosteric mutations in protein domains and produced comprehensive atlases of allosteric communications, distinguishing the effects of mutations on allostery, binding, and protein stability.[41] By using innovative implementations of protein-fragment complementation assays, this pioneering study allowed for a detailed characterization of the biophysical effects of mutations by quantifying multiple molecular phenotypes in multiple genetic backgrounds and fitting the data into thermodynamic models using neural networks. Another study reported a large-scale analysis of the genotype-phenotype landscape for the lac repressor from *Escherichia coli* LacI enabling a quantitative map of the effect of amino acid substitutions on LacI allostery.[42] This study showed that in general allosteric phenotypes can be quantitatively predicted using additive approximations and neural network-based models. However, allosteric effects may also operate via less-conventional mechanisms that can synchronize and amplify combinations of silent amino acid substitutions to induce allosteric changes. This investigation reinforced the notion that allostery is a distributed biophysical phenomenon governed primarily by the ensemble-defined remodeling of the energy landscape and the thermodynamic free energy balance with additive contributions from many residues and interactions.[42] To examine whether allosteric mutations are abundant, structurally localized, or distributed in nature, an elegant saturation mutagenesis study of a synthetic allosteric system in which dihydrofolate reductase (DHFR) is regulated by a blue-light sensitive LOV2 domain was conducted.[43] By assessing the impact of 1548 viable DHFR single mutations on allostery, this study established that fewer than 5% of mutations could exhibit a statistically significant influence on allostery, and that allostery-disrupting mutations were proximal to the insertion site, while allostery-enhancing mutations appeared to be structurally distributed and enriched on the protein surface.[43] Importantly, this DMS profiling study revealed that engineering of mutations in weakly conserved and structurally distributed sites of the protein could lead to diverse evolutionary strategies for optimization and manipulation of allosteric regulation. Moreover, these fascinating experimental insights into allosteric mechanisms disclosed various weaknesses of computational approaches that may often overemphasize the role of structurally stable allosteric hotspots, while allostery may be in fact rescued and enhanced through distributed cooperative effects of a considerable number of weakly conserved flexible sites.

DMS analysis of the molecular chaperone Hsp90 encoded 14,160 amino acid variants and quantified growth effects under standard conditions and under various stress conditions.[44] The results showed that different environments could impose unique functional demands on the Hsp90, where function-beneficial mutations occupied the protein surface and were often localized near interfaces with the binding partners. Moreover, mutations that disrupt

binding to certain clients can lead to the reprioritization of others, providing a roadmap for rational rewiring of cellular networks.[44] Interestingly, this comprehensive DMS mapping of the Hsp90 fitness maps revealed patterns that were generally consistent with the computational analysis of allosteric changes of the molecular chaperone, suggesting that mutations affecting client binding can be intimately involved in modulation of the Hsp90 allosteric communications.[45,46]

The ease and proliferation of DMS tools in modern biochemical studies enabled a systematic characterization and comparison of allosteric hotspots across multiple homologous proteins allowing for in-depth analysis of allosteric effects in protein families (Figure 1). Using computational protein design, single-residue saturation mutagenesis and random mutagenesis, along with multiplex assembly, DMS was employed to build a more comprehensive catalog of Lac repressor allosteric variants comparable in specificity and induction to wild-type LacI with its inducer.[47] An insightful review of DMS and high-throughput mutational methods emphasized the transformative role and advantages of these emerging technologies for understanding of the allosteric phenomenon and their unique ability to comprehensively map the functional landscape at the resolution of individual residues.[48] Furthermore, DMS can be used for profiling double mutants that disrupt or restore normal allosteric functions. Lastly, this analysis highlighted the large scale and data-rich nature of the DMS output that is perfectly suited for data mining and predicting residues that are exclusively important for allostery using ML models.[48] By integrating computational design, high-throughput screening along with structural and biophysical analysis of an allosteric transcription factor, the recent study showed that epistatic interactions can shape up the protein fitness landscape and allosteric functions, leading to new binding specificity.[49]

DMS experiments were combined with molecular dynamics (MD) simulations and network analysis into a function-centric approach[50] that examined the underlying functional landscape of a bacterial transcription factor showing how disrupted allosteric switches can be restored through functional plasticity and redundancy of flexible positions, suggesting the role of diverse and broad ensembles of mutational communication pathways in propagating allosteric phenotypic effects (Figure 1). This seminal study revealed that residues critical for allosteric signaling are often weakly conserved leading to multiple solutions to the thermodynamic principle of cooperativity, in contrast to the view of a finely tuned allosteric residue network maintained under evolutionary selection. In a subsequent study, DMS of four homologous bacterial allosteric transcription factors produced a large pool of data that was leveraged by deep learning (DL) to build a robust predictor of allosteric hotspots revealing that regulatory sites mediating allostery are widely distributed on the protein rather than being restricted to specific pathways linking the allosteric and active sites.[51] Moreover, a model trained on one protein can predict hotspots in a homologue, demonstrating that global structural and dynamic properties are typically strong predictors of allosteric importance for a given residue than local and physicochemical properties. Engineering of allosteric functions and regulation via limited number of key mutations were demonstrated by the analysis of the malate (MalDH) and lactate dehydrogenase (LDH) superfamily in which a few key mutations induced a reorganization of the conformational landscape rendering the emergence of allostery in LDH proteins. which we targeted for investigation

by site-directed mutagenesis.[52] The recent advances in DMS tools and rapid emergence of multiplexed (pooled) screens producing a large number of mutational perturbations and measurements in proteins using a single-pot experiment represent a considerable breakthrough in revealing allosteric functional landscapes while supplying ML tools with invaluable data sets to manipulate allosteric functions and engineer novel allosteric proteins.

ALPHAFOLD, PROTEIN ENSEMBLES AND ALLOSTERY: PAVING THE WAY FOR THE NEXT AI REVOLUTION IN MOLECULAR BIOLOGY

Among the emerging trends in studies of protein structure and dynamics is the growing realization and rapidly expanding efforts to develop a new generation of ML approaches that leverage the wealth of experimental and simulation data for autonomous assessment of dynamic events and regulatory mechanisms. The remarkable success of advanced ML methods in protein structure modeling is exemplified by achievements of AlphaFold2 (AF2) that leverages covariation and representations of amino acid contacts on graph neural networks to yield a robust DL framework that trains on the sequences of homologous proteins to predict a single accurate structure for all sequences.[53-55] The AlphaFold database, hosted at EMBL-EBI (https://alphafold.ebi.ac.uk/), provides free access to more than 200 million protein structure predictions—a remarkable advancement in structural biology that was inconceivable even several years ago.[56] A number of insightful reviews highlighted the key shortcomings and limitations of the AF2 technology in resolving the looming computational biology challenges as the predicted structural models remain static and are unable to directly describe functionally relevant dynamic changes in protein systems and allosteric signaling mechanisms.[57-59] Nussinov and colleagues emphasized that for understanding of the regulatory mechanisms the AF2 predicted structures need to be accompanied by their representative ensembles and relative populations that are essential for quantifying allosteric phenomena—a formidable and ambitious task that is now knocking on the door to test the limits of artificial intelligence (AI) technologies.[59]

In the current review we highlight some of the most recent "post-Alphafold2" developments that leveraged achievements in structure prediction to develop new modeling frameworks that attempt to extend beyond predicting a single protein structure and toward accurately capturing protein dynamics and regulatory mechanisms from first principles. Several latest studies outlined a simple and yet plausible strategy that leveraged a multiparameter complexity of the AF2 methodology to predict different functional conformations using a benchmark set of topologically diverse transporters and GPCR proteins—a first step toward adapting the powerful ML apparatus for modeling of protein ensembles and populations.[60,61] By varying different parameters of the predictor such as the number of models generated, the number of known structures of protein homologues as templates, and by counterintuitively reducing the depth of the input multiple sequence alignments by stochastic subsampling, this study reported a robust generation of multiple functional conformations required for protein activities and regulation[61] The results of this study indicated that AF2 parameters can be manipulated in a specified manner to accurately model multiple functional conformations for transporters and GPCRs whose structures were not used in the training set. A more general approach leverages AF2 to model

alternative functional conformations and is benchmarked on canonical examples of protein flexibility, showing promise in recapitulating the conformational landscape of membrane proteins.[62] In this approach, the initial AF2-predicted models are scanned to identify interaction surfaces within the structure, followed by modifying MSA profiles using in silico alanine mutagenesis and forcing the attention neural networks within the AF2 engine to uncover new residue contacts and promoting more heterogeneous coevolutionary couplings of protein residues to produce alternative protein conformations. Integration of double electron–electron resonance (DEER) spectroscopy and ensembles of multiple structural models obtained by the modified AF2 showed a good agreement with the experimental conformational dynamics.[63] Despite a significant and somewhat unexpected success of these AF2 adaptations, a methodological "tweaking" of the AF2 architecture may be sensitive to the protein families and evolutionary patterns among homologies sequences, indicating a need for the development of dynamics-centric neural networks and more systematic probing and adaptation of the AF2 architectures for explicit exploration of conformational ensembles.

Consistent with these arguments, several most recent studies indicated that remarkable structure prediction capabilities of AF2 cannot be readily expanded to learn and predict the conformational landscapes and allosteric conformational changes that drive protein functions and regulation. Using a curated collection of unbound (apo) and bound (holo) structures from the database of Conformational Diversity in the Native State of proteins (CoDNaS) it was found that AF2 predictions are biased toward a single conformer and cannot capture conformational diversity present in the apo and holo pairs with the same precision that can be estimated for a single representative conformation of a given protein.[64] Interestingly, AF2 predictions single out the holo protein form in 70% of the studied cases, but are unable to reproduce conformational diversity through assessment of the top predicted conformer models, suggesting that AF2 neural networks cannot simultaneously predict the protein structure and the conformational ensemble. In another study, the performance of AF2 was tested on a set of 98 fold-switching proteins with at least two distinct tertiary structures, revealing that 94% of predictions captured only one of the experimentally determined conformation but often failed to capture the other functional states among top predicted conformers.[65] By extracting AF2 predictions for the wild-type and single protein mutants, the predicted AF2 metrics were correlated with the experimental protein stability changes for 976 mutations in 90 proteins from the Thermo-Mut Database, showing weak or no correlation with the experimental changes of protein stability.[66] At the same time, AF2 predictions of ligand binding sites, protein disordered regions, and protein–protein interactions are superior to the existing tools even though AF2 networks were not initially trained on structures of protein–protein complexes.[67-69] Hence, while AF2 tools have excelled in predicting static structures of proteins, it remains unclear how these neural networks should be tweaked to predict conformational ensembles and identify allosteric states including low-populated dynamic functional conformations involved in allostery. The three-track network architectures developed by Baker and colleagues that incorporate and manipulate neural networks to transform and integrate sequence information with the 2D distance maps and 3D structure throughout the training have been equally powerful for protein structure prediction and provide architectural flexibility that could be potentially

adapted for prediction of dynamics and allosteric states.[70] Recent illuminating AI-driven protein design studies showed that deep networks trained to predict native protein structures from their sequences can be inverted to design new proteins, and this conceptual strategy could be potentially reformulated and applied to model conformational ensembles of the same protein.[71-73] These approaches developed by the Baker lab including constrained hallucination to optimize sequences for structures containing the desired functional site[71] and an inpainting approach that designs a viable protein scaffold around a given functional site[72] may provide a platform for "dialing in" sequence-inferred variability of the predicted structures as a proxy for rational modeling of protein ensembles.

NMR spectroscopy technologies offer new powerful means to characterize protein dynamics and detect hidden allosteric conformations[74-77] which in combination with AF2 modeling could be a promising direction for accurate prediction of conformational dynamics, allostery, and detection of rare states, but these enhancements of deep learning networks require the enhanced databases of NMR data for training.[78] An alternative direction is a systematic exploration of AI systems and ML approaches to capitalize on the wealth of computational and experimental information about protein dynamics and conformational ensembles. These approaches have a potential to become a unifying data-centric platform for synthesizing advances in theory and experimental technologies, leading to the development of robust and efficient computational models and expert systems for prediction of allosteric effects in protein systems.

EXPANDING THE HORIZONS OF EXPERIMENT-GUIDED MOLECULAR SIMULATIONS FOR STUDIES OF PROTEIN ALLOSTERY WITH NETWORK AND AI MODELS: COMING TO RESCUE YOU OR REPLACE YOU?

The recently emerging directions in molecular simulations of biomolecular systems and characterization of conformational dynamics reflect the arrival of the new era of dynamic structural biology exemplified by the increasing cooperation of cryo-EM and single molecule FRET techniques with the enhanced simulation approaches and AI/ML models. The development of enhanced sampling techniques for exploration of conformational landscapes with the aid of neural networks and DL architectures provided a significant impetus for studies of protein allostery and allosteric mechanisms. ML approaches were employed to facilitate exploration of conformational landscapes using MD simulations via optimal selection of reaction coordinates,[79-83] enhanced conformational sampling by reinforcement learning,[84,85] goal-oriented active learning,[86] and more recently by autonomous generation of the equilibrium ensembles using Boltzmann neural network generators.[87] A recent review summarized the fundamentals of generative ML applications for exploring the free energy surfaces and kinetics of proteins.[88] Here, we highlight several recent impactful developments in facilitating "autonomous" enhanced sampling of protein systems in which ML is deployed to learn the representations and distributions of biasing potentials as well as physics-based thermodynamic and kinetic constraints to drive a more efficient exploration of the conformational landscapes and prediction of functionally relevant dynamic states. By combining DL and biased MD simulations, physically meaningful collective variables can be determined resolving the bottlenecks that

often hinder the reliable characterization of conformational transitions and rare events.[89] ML together with the variationally enhanced sampling method allowed for learning and optimization of sampling-biasing potentials that can be represented in the form of neural networks.[90] Using the principle of variational inference implemented through deep neural networks and a predictive information bottleneck concept, a recently introduced framework leverages short MD simulations to estimate the reaction coordinates and perform iterative biased simulations that can subsequently enhance exploration of conformational landscapes and reliable inference of the associated thermodynamic and kinetic characteristics of the system.[91,92] Moreover, AI-based State Predictive Information Bottleneck (SPIB) approach can reliably learn a reaction coordinate via a deep neural network even from short and under-sampled trajectories.[93] Further developments of these concepts produced a path sampling approach that integrates generic thermodynamic or kinetic constraints into long short-term memory (LSTM) networks to accurately learn time series such as MD trajectories for systems from different application domains.[94] Going forward, the developments of these integrative biophysical approaches that leverage AI and ML tools to represent physics-based thermodynamic and kinetic drivers of efficient sampling in the form of neural networks would have significant implications for "autonomous" mapping of conformational landscapes, monitoring of allosteric changes, and detection of functional allosteric states.

MSM approaches are powerful tools for exploring long-time dynamic changes underlying the function of many allosterically regulated proteins, allowing for detailed network maps of functional states on the conformational landscape and quantitative analysis of the effect of perturbations on the thermodynamics and kinetics of allosteric transitions. However, the application of MSMs to characterize functional conformational changes in highly dynamic protein systems remains challenging due to heterogeneity of localized structural changes involved in allosteric transformations. As a result, a robust selection of structural features that can describe the slowest dynamics of allosteric conformational changes is an important bottleneck of the MSM approaches.[95] The automatic selection of physically meaningful and efficient reaction coordinates using ML approaches allows MSM tools to identify functionally relevant states which is the key to proper interpretation of allosteric regulation mechanisms.[95] The powerful synergy and complementarity of MSM approaches and by employing ML-augmented tools for detection of functionally relevant regions on the conformational landscapes and identification of structurally important multidimensional reaction coordinates, AI models can facilitate a more rapid advancement in the MSM methodologies with broader applications in studying functional conformational changes of proteins. In particular, ML models came to the rescue by streamlining this analysis and allowing for automatic selection of the essential features that can explain conformational changes and the distribution of metastable states.[96,97] The ML approach that identifies feature importance via an iterative exclusion principle can uncover versatile reaction coordinates that account for the dynamics of the slow degrees of freedom and allows for efficient sampling of the conformational landscapes and detection of hidden intermediate states of the system.[97]

A variational approach to the Markov process neural network (VAMPNets) provides a framework for predictions of molecular kinetics using neural networks by combining the steps of featurization, dimensionality reduction, discretization, coarse-grained kinetic

modeling, and generation of states into a single end-to-end learning system.[98] Combining VAMPNet and graph-level dynamics with neural networks provided an end-to-end framework termed GraphVAMPNet to efficiently learn high resolution metastable states from the long-time scale MD trajectories.[99] This approach also employed an attention learning mechanism to find the important residues for classification of conformational ensembles into different metastable states. A Gaussian mixture variational autoencoder (GMVAE) can learn a reduced representation of the free energy landscape of protein folding with highly separated clusters that correspond to the metastable states during folding.[100] Using quasi-MSM (qMSM) based on the Generalized Master Equation framework, only a handful of functionally relevant metastable states can be obtained from short MD simulations to facilitate the interpretation of regulatory mechanisms associated with specific local and global conformational changes.[101] The key ensemble properties of biological systems can be learned from MD simulations and described by easily interpretable metrics using a range of different ML methods including principal component analysis (PCA), random forests (RFs), and three types of neural networks (NNs): autoencoders (AEs), restricted Boltzmann machines (RBMs), and multilayer perceptrons (MLPs).[102] This versatile framework enables efficient learning of the key molecular features driving various biomolecular processes such as allosteric conformational rearrangements of the soluble protein calmodulin, the effect of ligand binding to a GPCR, and the allosteric coupling of an ion channel VSD to a transmembrane potential.

Several computational studies employed combinations of enhanced simulation schemes and various ML models to directly infer molecular determinants of allosteric changes and ligand-induced ensemble changes in proteins. A ML-based method (Linear Discriminant Analysis) was applied to reveal differences between the apo and allosteric inhibitor-bound ensembles in an automated way.[103] Another ML method was developed for the direct conformational ensemble comparison and understanding of temporal relationships during allosteric stimulation of hemagglutinin-neuraminidase.[104] Zhou and colleagues examined allosteric mechanism of Vivid (VVD) protein as one light, oxygen, or voltage (LOV) domain using an enhanced allosteric community model based on ML models.[105] Variational autoencoders (VAEs) have been successfully employed to explore the conformational space and allosteric transitions in adenosine kinase (Figure 2), showing that the learned latent space can be used to generate unsampled protein conformations and initiate additional MD simulations to sample a transition from the closed to the open states and explored hidden allosteric states.[106] An autoencoder-based detection method for characterization of ligand-induced dynamic allostery used a comparison of time fluctuations of the protein structures in the form of distance matrices obtained from MD simulations.[107]

In this simple and elegant approach, the autoencoder was first trained based on the time fluctuations of protein residues in the apo form and used to inspect data in both the apo and holo forms, showing that the ligand-induced allosteric changes in dynamics can be identified and attributed to specific reorganization of cooperative fluctuation motions among residue pairs on a long-time scale. A neural relational inference model based on a graph neural network used an autoencoder architecture to explore the latent embedding of an allosteric system and learn the long-range interactions and communications between distant sites in the ligand-induced allosteric regulation of Pin1, conformational transition of SOD1 protein

and the activation of MEK1 by oncogenic mutations.[108] By requiring a dimensionality reduction algorithm to predict the biochemical differences between protein variants instead of assuming whether large structural changes are more important than local changes, a new ML approach termed DiffNets uses a self-supervised autoencoder to learn features of the conformational ensembles that are relevant to dissect the biochemical differences between protein systems.[109]

MD simulations have been widely applied with smFRET experiments to provide atomistic insights into the dynamic behavior of biomolecules. A ML-based approach was proposed which links MD simulations and single-molecule experiments by constructing the initial MSM from a raw set of simulation data and a learning step in which hidden Markov modeling is performed to optimize the initial MSM using smFRET measurement data.[110] MD simulations can also be combined with the information provided by smFRET experiments to steer the simulation from one conformational state to the other using accelerated or enhanced sampling techniques.[111]

Combined with biophysical approaches and multiscale computational methods, NMR studies have been instrumental in uncovering the invisible aspects of protein "life" including mapping of allosteric landscapes for protein domains.[112-116] Using a combination of triple-resonance NMR and computational network analysis, the allosteric effects of specific kinase mutations and communication paths between regulatory elements and catalytic sites can be characterized.[117]

NMR chemical shift covariance (CHESCA) and projection (CHESPA) analyses[118-121] can identify residue interaction networks that show correlated changes in chemical shifts due to allosteric perturbations caused by ligand binding or mutations designed to modulate allosteric conformational equilibria. Using statistical comparative analyses of the NMR chemical shift variations elicited by the selected perturbations, the CHESCA approach characterizes perturbation-specific chemical shift patterns serving as distinctive signatures of allosteric mechanisms. NMR studies of PKA and PKG kinases revealed a wide range of noncanonical allosteric effectors ranging from post-translational modifications to disease-related mutations that can define diverse mechanisms of constitute activation.[122] The two newly proposed CHESCA-based methodologies, called temperature CHESCA (T-CHESCA) and CLASS-CHESCA, can prioritize predicted allosteric sites and identify the core allosteric residues.[123] These NMR CHESCA adaptations are based on the invariance of core inter-residue correlations to changes in the chemical shifts of the active and inactive conformations interconverting in fast exchange. Integration of NMR spectroscopy and surface plasmon resonance revealed dynamic communication networks of residues linking the ligand-binding site to the activation interface in the glucocorticoid receptor and identified a specific motif acting as a ligand- and coregulator-dependent allosteric switch governing transcriptional activation.[124] A recently introduced NMR-guided directed evolution approach highlighted a new role of NMR in the selection process of mutational libraries as this approach can identify locations of the allosteric hotspots and mutations that can minimize nonessential protein dynamics to achieve high catalytic efficiency without a priori structural information.[125]

Solution NMR experiments and Gaussian-accelerated molecular dynamics (GaMD) simulations examined the structural and dynamic determinants of allosteric signaling within the CRISPR-Cas9 HNH nuclease, advancing our understanding of the allosteric pathway of activation.[126] A further integration of NMR with multimicrosecond molecular dynamics (MD) simulations and graph-based network modeling probed the effects of mutations on the structure and allosteric communication within the CRISPR-Cas9 system, showing that mutations responsible for increasing the specificity of Cas9 alter the allosteric structure of the catalytic HNH domain.[127] Attempts are still underway to develop fundamental, if not ubiquitous, theory of protein allostery. The recent study hypothesized that higher-order cooperativities among multiple binding events rather than pairwise cooperativities are needed to decipher protein allostery.[128] This graph-based method extends the idea of allosteric regulation to systems with many distinct conformational degrees of freedom and provides a conceptual framework for considering complex allosteric systems with multiple distinct conformations as versatile apparatus functioning to integrate information from ligand binding.

ALLOSTERIC REGULATION MODELS AND MACHINE LEARNING IN STRUCTURAL BIOLOGY AND FUNCTIONAL STUDIES OF THE SARS-CoV-2 SPIKE PROTEINS AND ESCAPE MECHANISMS

Here, we discuss recent advances in integrative structural biology of SARS-CoV-2 spike proteins, which highlight an important and often hidden role of allosteric regulation driving functional conformational changes, binding interactions with the host receptor and mutational escape mechanisms of S proteins which are critical for viral infection. The latest developments in structural and computational studies of SARS-CoV-2 S proteins also underscore the value of AI-based approaches to unveil otherwise cryptic allostery states, druggable allosteric sites, and regulatory mechanisms. The rapidly growing body of structural and functional studies established that the mechanism of SARS-CoV-2 infection that involves conformational transitions between distinct functional forms and activation of the viral spike (S) glycoprotein trimer which consists of an amino (N)-terminal S1 subunit and carboxyl (C)-terminal S2 subunit where S1 participates in the interactions with the angiotensin-converting enzyme 2 (ACE2) host receptor using the receptor-binding domain (RBD).[129,130] Conformational transitions between the closed S state with RBDs in the "down" conformation and the receptor-bound open state in which RBDs can adopt an "up" conformation were characterized using biophysical experiments suggesting that mechanisms of conformational selection and receptor-induced structural adaptation can often involve allosteric stabilization and regulation.[129,130] Recent experimental and computational studies suggested that dynamic biological functions of the SARS-CoV-2 S proteins and mutational escape mechanisms can be rationalized and predicted by examining critical molecular events related to viral infection and dissemination through the lens of protein allostery and the allosteric regulatory landscape of the SARS-CoV-2 S protein. Conformational dynamics of SARS-CoV-2 S protein in the absence or presence of ligands visualized using smFRET imaging assays showed that ACE2 binding is controlled by the conformational landscape of the RBD via population-shift mechanism, in which

ACE2 captures the intrinsically accessible up RBD conformation rather than inducing a conformational change.[131] Moreover, smFRET data demonstrated that that antibodies that target diverse epitopes of the S protein located away from the RBD can allosterically modulate the RBD functional dynamics and shift the thermodynamic equilibrium toward the open S form that promotes ACE2 binding. Conformational dynamics of SARS-CoV-2 trimeric S glycoprotein in complex with ACE2 revealed by cryo-EM experiments further confirmed that binding can modulate the conformational landscape of the S trimer and induce continuous swing motions between allosteric states.[132] This cryo-EM investigation proposed a mechanism of conformational transitions of the SARS-CoV-2 S trimer acting as the dynamic allosteric fusion machine from the ground prefusion state to the postfusion state, in which ACE2 binding shifts the conformational landscape toward the open RBD state and promoting a cascade of allosteric responses of the fusion machine facilitating transitions toward the postfusion state.[132] The energy landscape of the SARS-CoV-2 S proteins and complexes with antibodies revealed extensive conformational heterogeneity in which changes between unbound protein and complexes with antibodies are often reminiscent of apo-to-holo switching using the preexisting conformational equilibrium. The intrinsic flexibility of the SARS-Cov-2 S proteins examined by enhanced sampling simulations agreed with FRET cryo-EM experiments, unveiling a multitude of functional allosteric conformations with druggable cryptic pockets (Figure 3).[133,134]

These experimental and computational studies supported an emerging paradigm that allosterically regulated dynamics of the S protein may provide a versatile mechanism for efficient virus transmission and enable diversity of escape mutation-induced allosteric responses that counteract the effects of antibodies. Using Folding@home distributed computing project adaptive sampling simulations of the viral proteome captured dramatic opening of the apo Spike complex, far beyond that seen experimentally, and predicts the existence of "cryptic" epitopes and hidden allosteric pockets.[135] Using the cryo-EM Metainference (EMMI) method that can accurately model conformational ensembles by combining simulations with cryo-EM data, the intermediate states in the opening pathway of SARS-CoV-2 S protein were identified signaling a potentially druggable cryptic allosteric site located in the vicinity of the RBD recognition site.[136] These extensive simulation studies have provided a strong evidence of conformational heterogeneity of the S protein capable of adopting a multitude of functional conformations and unveiling previously unknown cryptic pockets during allosteric transitions between the open and closed forms. Our recent studies combined multiscale simulations of conformational landscapes with coevolutionary analysis and network-based modeling of the SARS-CoV-2 proteins to examine allosteric mechanisms of the SARS-CoV-2 S proteins.[137-140] These studies suggested that coevolution, conformational dynamics, and allostery conspire to drive cooperative binding interactions and signal transmission of the SARS-CoV-2 S protein with ACE2 enzyme. These studies provided compelling evidence that the SARS-CoV-2 S protein can function as a functionally adaptable allosterically regulated machine that exploits plasticity of allosteric centers to fine-tune responses to antibody binding, where the experimentally confirmed regulatory hotspots correspond to the global mediating centers of the allosteric interaction networks (Figure 3). By examining conformational landscapes and the residue interaction networks in the SARS-CoV-2 Omicron spike protein structures, we have shown that the Omicron

mutational sites are dynamically coupled and form a central engine of the allosterically regulated spike machinery that regulates the balance between conformational plasticity, protein stability, and functional adaptability.[141] MD simulations demonstrated an allosteric crosstalk within the RBD in the apo- and the ACE2 receptor-bound states.[142] Allosteric interactions between SARS-CoV-2 spike mutational sites were also confirmed in extensive MD simulations, suggesting that the interplay of spatially proximal local interactions and long-range communications between sites of escape mutations can represent an evolutionary strategy employed by the virus to modulate virulence of emerging SARS-CoV-2 variants.[143]

Elucidating this relationship between local interactions and their global effects is essential to understanding evolution of allosteric proteins that can be manifested as epistatic nonadditive changes in biophysical properties at the level of biological function.[49] The effect of nonadditive, epistatic relationships among S-RBD mutations was assessed by comparing the effects of all single mutants at the RBD-ACE2 interfaces for the Omicron variants, showing that structural constraints can curtail the virus evolution and put limits on antibody escape.[144] A systematic analysis of the epistatic effects in the S-RBD proteins using DMS analysis of all amino acid mutations in the SARS-CoV-2 S variants showed nonadditive contributions of physically proximal mutational sites as well as long-range couplings between sites of escape mutations.[145]

The functional and systems biology studies reinforced the notion that the Omicron mutations may have emerged as an evolutionary product of balancing multiple fitness requirements, including the immune escape, productive binding with the host receptor, conformational plasticity, and allosteric communications.[146,147] The reversed allosteric communication approach is based on the premise that allosteric signaling in proteins is bidirectional and can propagate from an allosteric to orthosteric site and vice versa, thus providing means for detecting cryptic allosteric sites.[148,149] An integrated computational and experimental strategy exploited the reversed allosteric communication concepts to combine MD simulations with MSM for characterization of binding shifts in the protein ensembles and identification of cryptic allosteric sites.[150] A network-based adaptation of the reversed allosteric communication approach was proposed to identify allosteric hotspots and infer this analysis to characterize the distribution of allosteric binding pockets (Figure 3) in the SARS-CoV-2 Spike Omicron BA.1, BA.1.1, BA.2, and BA.3 variant complexes.[151] Integrative computational and experimental studies detailed allosteric communications in an S protein trimer and validated the allosteric site located between SD1 and SD2 subdomains of the S protein (Figure 3).[152] By screening commercial compound databases, several hits were selected and validated at both the molecular level and cellular level for their binding strength and antivirus activities (Figure 4).

We also reported the discovery of potential small molecules targeting the SARS-CoV-2 S protein by combining in silico technologies with in vitro experimental methods. Using mass spectrometry (MS) and surface plasmon resonance (SPR) methods our studies have discovered and validated five natural products as potential modulators of the S activity.[153] Using a combination of in silico and biochemical tools, *N*-acetylneuraminic acid (Neu5Ac), a type of predominant sialic acid found in human cells, was tested as a molecular probe of the S protein and validated as an allosteric modulator.[154] A similar dual strategy

of molecular docking and SPR screening of compound libraries interrogated 57,641 compounds and identified 17 binders of ACE2 and 6 potent blockers of the RBD that compete with the RBD-ACE2 interactions in an SPR-based competition assay.[155] Although identification and validation of allosteric modulators of the SARS-CoV-2 S proteins remain to be challenging tasks, exploiting allosteric regulatory mechanisms and allosteric binding sites in SARS-CoV-2 proteins has potential to discover viable broad-spectrum therapeutic agents with utility for drug resistance.

AI expert systems and ML approaches showed a considerable promise to reveal functions of SARS-CoV-2 spike (S) proteins particularly predicting patterns of evolving mutations and mutational escape mechanisms. Deep mutational learning (DML), a machine-learning-guided protein engineering technology, was developed to investigate the enormous sequence space of combinatorial mutations and accurately predict the impact of these mutations on ACE2 binding and antibody escape.[156] This method integrates yeast display screening of RBD mutational libraries with deep sequencing into an ML approach that can predict antibody robustness to a large variety of SARS-CoV-2 variants, thus serving as a guide for selection of effective therapeutics for virus infection.[156] A comprehensive ML-based investigative framework for analysis of S protein mutations was developed and applied to 4296 Omicron viral genomes, revealing a core haplotype of 28 polymutants in the S protein and a separate core haplotype of 17 polymutants in nonspike genes.[157] A multitask ML framework that harnesses systematic mutation screens in the RBD of the S protein for predicting SARS-CoV- 2 antibody escape was recently unveiled.[15] This ML model analyzes data on escape from multiple antibodies simultaneously, creating a latent representation of mutations that is effective in predicting the escape potential and binding properties of the virus.[158] ML models have been actively deployed to facilitate physics-based predictions of the S proteins with ACE2 and antibodies, revealing the impact of RBD mutations and suggesting novel sets of mutations that strongly modulate binding and escape properties of the virus.[159,160]

CONCLUDING REMARKS AND FUTURE PERSPECTIVES

Despite the growing evidence that many complex protein systems and regulatory assemblies function as dynamic and versatile allosteric machines, the understanding and characterization of the protein allostery universe even for a single system which includes hidden allosteric protein states, allosteric interactions, and communication pathways are still surprisingly limited. Although many theories and models have been developed in attempts to rigorously describe this phenomenon, the highly dynamic, complex, and diverse nature of allosteric events and mechanisms continues to pose new challenges to the field testing the limitations of existing technologies and making the quest for a universal theory of allostery an important priority of computational and structural biology. Among emerging directions in the field are computational methods for the identification and mapping of allosteric networks, as well as novel experimental approaches to study allosteric mechanisms, including time-resolved and single-molecule studies; approaches to engineering allosteric regulation to enhance function and facilitate the design of sensors and drugs, the design of synthetic chemical networks that use allostery in feedback mechanisms, directed evolution of allostery, nonequilibrium simulation methods for modeling of allosteric

ensembles and pathways. The latest advances in structural characterization of allosteric molecular events and hidden functional states important for allosteric function using cryo-EM, NMR, smFRET spectroscopy have highlighted the growing need for data-centric integrative biophysics approaches. By developing an open science infrastructure for ML studies of allosteric regulation and validating computational approaches using integrative studies of allosteric mechanisms, the scientific community can expand the toolkit of approaches and chemical probes for dissecting and interrogation allosteric mechanisms in many therapeutically important proteins. The development of community-accessible tools that uniquely leverage the existing experimental and simulation knowledge base to enable interrogation of the allosteric functions can provide a much needed impetus to further experimental technologies and enable steady progress.

Data Availability Statement

Crystal structures were obtained and downloaded from the Protein Data Bank (http://www.rcsb.org), accession numbers 6X2C (the cryo-EM structure of SARS-CoV-2 S-protein in the closed 3RBD-down state), 6X2A (the cryo-EM structure of SARS-CoV-2 S-protein in the open 1RBD-up state), and 6X2B (the cryo-EM structure of SARS-CoV-2 S-protein in the open 2RBD-up state). All simulations were performed using the NAMD 2.13 package that was obtained from Web site https://www.ks.uiuc.edu/Development/Download/. All simulations were performed using the all-atom additive CHARMM36 protein force field that can be obtained from http://mackerell.umaryland.edu/charmm_ff.shtml. The residue interaction network files were obtained for all structures using the Residue Interaction Network Generator (RING) program RING v2.0.1 freely available at http://old.protein.bio.unipd.it/ring/. The computations of network parameters were done using NAPS program available at https://bioinf.iiit.ac.in/NAPS/index.php and Cytoscape 3.8.2 environment available at https://cytoscape.org/download.html. The rendering of protein structures was done with interactive visualization program UCSF ChimeraX package (https://www.rbvi.ucsf.edu/chimerax/) and Pymol (https://pymol.org/2/) . The software tools used in this study, including SciPy (https://www.scipy.org), and Pandas (https://pandas.pydata.org) are freely available at their Web sites. All the data obtained in this work (including simulation trajectories, topology and parameter files, dynamic residue interaction networks, and analysis), all the software tools, and the in-house scripts are freely available in the GitHub sites https://github.com/smu-tao-group/protein-VAE; https://github.com/smu-tao-group/PASSer2.0 ; https://github.com/kassabry/Perturbation_Experiment.

ACKNOWLEDGMENTS

REFERENCES

(1). Monod J; Wyman J; Changeux JP On the nature of allosteric transitions: a plausible model. J. Mol. Biol 1965, 12, 88–118. [PubMed: 14343300]

(2). Koshland DE Jr. Conformational changes: how small is hig enough? Nat. Med 1998, 4 (10), 1112–1114. [PubMed: 9771734]

(3). Changeux JP Allostery and the Monod-Wyman-Changeux model after 50 years. Anna. Rev. Biophys 2012, 41, 103–133. [PubMed: 22224598]

(4). Changeux JP; Edelstein SJ Allosteric mechanisms of signal transduction. Science 2005, 308 (5727), 1424–1428. [PubMed: 15933191]

(5). Popovych N; Sun S; Ehright RH; Kalodimos CG Dynamically driven protein allostery. Nat. Struct Mol. Biol 2006, 13 (9), 831–838. [PubMed: 16906160]

(6). Hilser VJ; Wrahl JO; Motlagh HN Structural and energetic basis of allostery. Anna. Rev. Biophys 2012, 41, 585–609. [PubMed: 22577828]

(7). Wrahl JO; Gu J; Liu T; Schrank TP; Whitten ST; Hilser VJ The role of protein conformational fluctuations in allostery, function, and evolution. Biophys Chem. 2011, 159 (1), 129–141. [PubMed: 21684672]

(8). Motlagh HN; Wrahl JO; Li J; Hilser VJ The ensemble nature of allostery. Nature 2014, 508 (7496), 331–339. [PubMed: 24740064]

(9). Tzeng SR; Kalodimos CG The role of slow and fast protein motions in allosteric interactions. Biophys Rev. 2015, 7 (2), 251–255. [PubMed: 28510174]

(10). Imelio JA; Trajtenberg F; Buschiazzo A Allostery and protein plasticity: the keystones for bacterial signaling and regulation. Biophys Rev. 2021, 13 (6), 943–953. [PubMed: 35059019]

(11). Buchenberg S; Sittel F; Stock G Time-resolved observation of protein allosteric communication. Proc. Natl. Acad. Sci. U.S.A 2017, 114 (33), E6804–E6811. [PubMed: 28760989]

(12). Stock G; Hamm P A non-equilibrium approach to allosteric communication. Philos. Trans. R Soc. London B Biol. Sci 2018, 373 (1749), 20170187. [PubMed: 29735740]

(13). Bozovic O; Zanobini C; Gulzar A; Jankovic B; Buhrke D; Post M; Wolf S; Stock G; Hamm P Real-time observation of ligand-induced allosteric transitions in a PDZ domain. Proc. Natl. Acad. Sci. U.S.A 2020, 117 (42), 26031–26039. [PubMed: 33020277]

(14). Wolf S; Sohmen B; Hellenkamp B; Thurn J; Stock G; Hugel T Hierarchical dynamics in allostery following ATP hydrolysis monitored by single molecule FRET measurements and MD simulations. Chem. Sci 2021, 12 (9), 3350–3359. [PubMed: 34164105]

(15). Oliveira ASF; Edsall CJ; Woods CJ; Bates P; Nunez GV; Wonnacott S; Bermudez I; Ciccotti G; Gallagher T; Sessions RB; Mulholland AJ A General Mechanism for Signal Propagation in the Nicotinic Acetylcholine Receptor Family. J. Am. Chem. Soc 2019, 141 (51), 19953–19958. [PubMed: 31805762]

(16). Oliveira ASF; Shoemark DK; Campello HR; Wonnacott S; Gallagher T; Sessions RB; Mulholland AJ Identification of the Initial Steps in Signal Transduction in the $\alpha 4\beta 2$ Nicotinic Receptor: Insights from Equilibrium and Nonequilibrium Simulations. Structure 2019, 27 (7), 1171–1183. [PubMed: 31130483]

(17). Abreu B; Lopes EF; Oliveira ASF; Soares CM F508del disturbs the dynamics of the nucleotide binding domains of CFTR before and after ATP hydrolysis. Proteins 2020, 88 (1), 113–126. [PubMed: 31298435]

(18). Galdadas I; Qu S; Oliveira ASF; Olehnovics E; Mack AR; Mojica MF; Agarwal PK; Tooke CL; Gervasio FL; Spencer J; Bonomo RA; Mulholland AJ; Haider S Allosteric communication in class A β-lactamases occurs via cooperative coupling of loop dynamics. Elife 2021, 10, No. e66567. [PubMed: 33755013]

(19). Oliveira ASF; Ciccotti G; Haider S; Mulholland AJ Dynamical nonequilibrium molecular dynamics reveals the structural basis for allostery and signal propagation in biomolecular systems. Eur. Phys. J. B 2021, 94 (7), 144. [PubMed: 34720710]

(20). Weinkam P; Pons J; Sali A Structure-based model of allostery predicts coupling between distant sites. Proc. Natl. Acad. Sci. U.S.A 2012, 109 (13), 4875–4880. [PubMed: 22403063]

(21). Molinelli EJ; Korkut A; Wang W; Miller ML; Gauthier NP; Jing X; Kaushik P; He Q; Mills G; Solit DB; Pratilas CA; Weigt M; Braunstein A; Pagnani A; Zecchina R; Sander C Perturbation biology: inferring signaling networks in cellular systems. PLoS Comput. Biol 2013, 9 (12), No. e1003290. [PubMed: 24367245]

(22). Hakhverdyan Z; Molloy KR; Keegan S; Herricks T; Lepore DM; Munson M; Subbotin RI; Fenyö D; Aitchison JD; Fernandez-Martinez J; Chait BT; Rout MP Dissecting the Structural Dynamics of the Nuclear Pore Complex. Mol. Cell 2021, 81 (1), 153–165. [PubMed: 33333016]

(23). Brotzakis ZF; Vendruscolo M; Bolhuis PG A method of incorporating rate constants as kinetic constraints in molecular dynamics simulations. Proc. Natl. Acad. Sci. U.S.A 2021, 118 (2), No. e2012423118. [PubMed: 33376207]

(24). Kutlu Y; Ben-Tal N; Haliloglu T Global Dynamics Renders Protein Sites with High Functional Response. J. Phys. Chem. B 2021, 125 (18), 4734–4745. [PubMed: 33914546]

(25). Atilgan C; Atilgan AR Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein. PLoS Comput. Biol 2009, 5, No. e1000544. [PubMed: 19851447]

(26). Atilgan C; Gerek ZN; Ozkan SB; Atilgan AR Manipulation of conformational change in proteins by single-residue perturbations. Biophys. J 2010, 99, 933–943. [PubMed: 20682272]

(27). Abdizadeh H; Atilgan C Predicting long term cooperativity and specific modulators of receptor interactions in human transferrin from dynamics within a single microstate. Phys. Chem. Phys 2016, 18, 7916–7926.

(28). Chennubhotla C; Bahar I Markov propagation of allosteric effects in biomolecular systems: application to GroEL-GroES. Mol. Syst. Biol 2006, 2, 36. [PubMed: 16820777]

(29). Chennubhotla C; Bahar I Signal propagation in proteins and relation to equilibrium fluctuations. PLoS Comput. Biol 2007, 3 (9), 1716–1726. [PubMed: 17892319]

(30). Chennubhotla C; Yang Z; Bahar I Coupling between global dynamics and signal transduction pathways: a mechanism of allostery for chaperonin GroEL. Mol. Biosyst 2008, 4 (4), 287–292. [PubMed: 18354781]

(31). Rosvall M; Bergstrom CT Maps of random walks on complex networks reveal community structure. Proc. Natl. Acad. Sci. U.S.A 2008, 105 (4), 1118–1123. [PubMed: 18216267]

(32). Rosvall M; Esquivel AV; Lancichinetti A; West JD; Lambiotte R Memory in network flows and its effects on spreading dynamics and community detection. Nat. Commun 2014, 5, 4630. [PubMed: 25109694]

(33). Delvenne JC; Lambiotte R; Rocha LE Diffusion on networked systems is a question of time or structure. Nat. Commun 2015, 6, 7366. [PubMed: 26054307]

(34). Shukla D; Hernandez CX; Weber JK; Pande VS Markov state models provide insights into dynamic modulation of protein function. Acc. Chem. Res 2015, 48 (2), 414–422. [PubMed: 25625937]

(35). Wu H; Paul F; Wehmeyer C; Noe F Multiensemble Markov models of molecular thermodynamics and kinetics. Proc. Natl. Acad. Sci. U.S.A 2016, 113 (23), E3221–E3230. [PubMed: 27226302]

(36). Sengupta U; Strodel B Markov models for the elucidation of allosteric regulation. Philos. Trans. R Soc. London B Biol. Sci 2018, 373 (1749), 20170178. [PubMed: 29735732]

(37). Bowman GR; Bolin ER; Hart KM; Maguire BC; Marquee S Discovery of multiple hidden allosteric sites by combining Markov state models and experiments. Proc. Natl. Acad. Sci. U.S.A 2015, 112 (9), 2734–2739. [PubMed: 25730859]

(38). Hart KM; Ho CM; Dutta S; Gross ML; Bowman GR Modelling proteins' hidden conformations to predict antibiotic resistance. Nat. Commun 2016, 7, 12965. [PubMed: 27708258]

(39). Fowler DM; Fields S Deep mutational scanning: a new style of protein science. Nat. Methods 2014, 11 (8), 801–807. [PubMed: 25075907]

(40). Starr TN; Greaney AJ; Hilton SK; Ellis D; Crawford KHD; Dingens AS; Navarro MJ; Bowen JE; Tortorici MA; Walls AC; King NP; Veesler D; Bloom JD Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. Cell 2020, 182 (5), 1295–1310. [PubMed: 32841599]

(41). Faure AJ; Domingo J; Schmiedel JM; Hidalgo-Carcedo C; Diss G; Lehner B Mapping the energetic and allosteric landscapes of protein binding domains. Nature 2022, 604 (7904), 175–183. [PubMed: 35388192]

(42). Tack DS; Tonner PD; Pressman A; Olson ND; Levy SF; Romantseva EF; Alperovich N; Vasilyeva O; Ross D The genotype-phenotype landscape of an allosteric protein. Mol. Syst. Biol 2021, 17 (3), No. e10179. [PubMed: 33784029]

(43). McCormick JW; Russo MA; Thompson S; Blevins A; Reynolds KA Structurally distributed surface sites tune allosteric regulation. Elife 2021, 10, No. e68346. [PubMed: 34132193]

(44). Flynn JM; Rossouw A; Cote-Hammarlof P; Fragata I; Mavor D; Hollins C 3rd; Bank C; Bolon DN Comprehensive fitness maps of Hsp90 show widespread environmental dependence. Elife 2020, 9, No. e53810. [PubMed: 32129763]

(45). Verkhivker GM Exploring Mechanisms of Allosteric Regulation and Communication Switching in the Multiprotein Regulatory Complexes of the Hsp90 Chaperone with Cochaperones and Client Proteins: Atomistic Insights from Integrative Biophysical Modeling and Network Analysis of Conformational Landscapes. J. Mol. Biol 2022, 434 (17), 167506. [PubMed: 35202628]

(46). Verkhivker GM Conformational Dynamics and Mechanisms of Client Protein Integration into the Hsp90 Chaperone Controlled by Allosteric Interactions of Regulatory Switches: Perturbation-Based Network Approach for Mutational Profiling of the Hsp90 Binding and Allostery. J. Phys. Chem. B 2022, 126 (29), 5421–5442.

(47). Taylor ND; Garruss AS; Moretti R; Chan S; Arbing MA; Cascio D; Rogers JK; Isaacs FJ; Kosuri S; Baker D; Fields S; Church GM; Raman S Engineering an allosteric transcription factor to respond to new ligands. Nat. Methods 2016, 13 (2), 177–183. [PubMed: 26689263]

(48). Raman S. Systems Approaches to Understanding and Designing Allosteric Proteins. Biochemistry 2018, 57 (4), 376–382. [PubMed: 29235352]

(49). Nishikawa KK; Hoppe N; Smith R; Bingman C; Raman S Epistasis shapes the fitness landscape of an allosteric specificity switch. Nat. Commun 2021, 12 (l), 5562. [PubMed: 34548494]

(50). Leander M; Yuan Y; Meger A; Cui Q; Raman S Functional plasticity and evolutionary adaptation of allosteric regulation. Proc. Natl. Acad. Sci. U.S.A 2020, 117 (41), 25445–25454. [PubMed: 32999067]

(51). Leander M; Liu Z; Cui Q; Raman S Deep mutational scanning and machine learning reveal structural and molecular rules governing allosteric hotspots in homologous proteins. Elife 2022, 11, No. e79932. [PubMed: 36226916]

(52). Iorio A; Brochier-Armanet C; Mas C; Sterpone F; Madern D Protein Conformational Space at the Edge of Allostery: Turning a Nonallosteric Malate Dehydrogenase into an "Allosterized" Enzyme Using Evolution-Guided Punctual Mutations. Mol. Biol. Evol 2022, 39 (9), msac186. [PubMed: 36056899]

(53). Jumper J; Evans R; Pritzel A; Green T; Figurnov M; Ronneberger O; Tunyasuvunakool K; Bates R; Žídek A; Potapenko A; Bridgland A; Meyer C; Kohl SAA; Ballard AJ; Cowie A; Romera-Paredes B; Nikolov S; Jain R; Adler J; Back T; Petersen S; Reiman D; Clancy E; Zielinski M; Steinegger M; Pacholska M; Berghammer T; Bodenstein S; Silver D; Vinyals O; Senior AW; Kavukcuoglu K; Kohli P; Hassabis D Highly Accurate Protein Structure Prediction with AlphaFold. Nature 2021, 596 (7873), 583–589. [PubMed: 34265844]

(54). Tunyasuvunakool K; Adler J; Wu Z; Green T; Zielinski M; Žídek A; Bridgland A; Cowie A; Meyer C; Laydon A; Velankar S; Kleywegt GJ; Bateman A; Evans R; Pritzel A; Figurnov M; Ronneberger O; Bates R; Kohl SAA; Potapenko A; Ballard AJ; Romera-Paredes B; Nikolov S; Jain R; Clancy E; Reiman D; Petersen S; Senior AW; Kavukcuoglu K; Birney E; Kohli P; Jumper J; Hassabis D Highly Accurate Protein Structure Prediction for the Human Proteome. Nature 2021, 596 (7873), 590–596. [PubMed: 34293799]

(55). Jumper J; Evans R; Pritzel A; Green T; Figurnov M; Ronneberger O; Tunyasuvunakool K; Bates R; Žídek A; Potapenko A; Bridgland A; Meyer C; Kohl SAA; Ballard AJ; Cowie A; Romera-Paredes B; Nikolov S; Jain R; Adler J; Back T; Petersen S; Reiman D; Clancy E; Zielinski M; Steinegger M; Pacholska M; Berghammer T; Silver D; Vinyals O; Senior AW; Kavukcuoglu K; Kohli P; Hassabis D Applying and Improving AlphaFold at CASP14. Proteins 2021, 89 (12), 1711–1721. [PubMed: 34599769]

(56). Varadi M; Anyango S, Deshpande M; Nair S; Natassia C; Yordanova G; Yuan D; Stroe O; Wood G; Laydon A; Žídek A; Green T; Tunyasuvunakool K; Petersen S; Jumper J; Clancy E; Green R; Vora A; Lutfi M; Figurnov M; Cowie A; Hobbs N; Kohli P; Kleywegt G; Birney E; Hassabis D; Velankar S AlphaFold Protein Structure Database: Massively Expanding the Structural Coverage of Protein-Sequence Space with High-Accuracy Models. Nucleic Acids Res. 2022, 50 (Dl), D439–D444. [PubMed: 34791371]

(57). Fleishman SJ; Horovitz A Extending the New Generation of Structure Predictors to Account for Dynamics and Allostery. J. Mol. Biol 2021, 433 (20), 167007. [PubMed: 33901536]

(58). Skolnick J; Gao M; Zhou H; Singh S AlphaFold 2: Why It Works and Its Implications for Understanding the Relationships of Protein Sequence, Structure, and Function. J. Chem. Inf. Model 2021, 61 (10), 4827–4831. [PubMed: 34586808]

(59). Nussinov R; Zhang M; Liu Y; Jang H AlphaFold, Artificial Intelligence (AI), and Allostery. J. Phys. Chem. B 2022, 126 (34), 6372–6383. [PubMed: 35976160]

(60). Schlessinger A; Bonomi M Exploring the conformational diversity of proteins. Elife 2022, 11, No. e78549. [PubMed: 35443909]

(61). Del Alamo D; Sala D; Mchaourab HS; Meiler J Sampling alternative conformational states of transporters and receptors with AlphaFold2. Elife 2022, 11, No. e75751. [PubMed: 35238773]

(62). Stein RA; Mchaourab HS SPEACH_AF: Sampling protein ensembles and conformational heterogeneity with Alphafold2. PLoS Comput. Biol 2022, 18 (8), No. e1010483. [PubMed: 35994486]

(63). Del Alamo D; DeSousa L; Nair RM, Rahman S; Meiler J; Mchaourab HS Integrated AlphaFold2 and DEER investigation of the conformational dynamics of a pH-dependent APC antiporter. Proc. Natl. Acad. Sci. U.S.A 2022, 119 (34), No. e2206129119. [PubMed: 35969794]

(64). Saldaño T; Escobedo N; Marchetti J; Zea DJ; Mac Donagh J; Velez Rueda AJ; Gonik E; García Melani A; Novomisky Nechcoff J; Salas MN; Peters T; Demitroff N; Fernandez Alberti S; Palopoli N; Fornasari MS; Parisi G Impact of Protein Conformational Diversity on AlphaFold Predictions. Bioinformatics 2022, 38 (10), 2742–2748. [PubMed: 35561203]

(65). Chakravarty D; Porter LL AlphaFold2 fails to predict protein fold switching. Protein Sci. 2022, 31 (6), No. e4353. [PubMed: 35634782]

(66). Pak MA; Markhieva KA; Novikova MS; Petrov DS; Vorobyev IS; Maksimova ES; Kondrashov FA; Ivankov DN Using AlphaFold to Predict the Impact of Single Mutations on Protein Stability and Function. bioRxiv 2021, DOI: 10.1101/2021.09.19.460937.

(67). Akdel M; Pires DEV; Pardo EP; Jänes J; Zalevsky AO; Mészáros B; Bryant P; Good LL; Laskowski RA; Pozzati G; Shenoy A; Zhu W; Kundrotas P; Serra VR; Rodrigues CHM Dunham AS; Burke D; Borkakoti N; Velankar S; Frost A; Basquin J; Lindorff-Larsen K; Bateman A; Kajava AV; Valencia A; Ovchinnikov S; Durairaj J; Ascher DB; Thornton JM; Davey NE; Stein A; Elofsson A; Croll TI; Beltrao P A Structural Biology Community Assessment of AlphaFold2 Applications. Nat. Struct. Mol. Biol 2022, 29 (11); 1056–1067. [PubMed: 36344848]

(68). Mirdita M; Schütze K; Moriwaki Y; Heo L; Ovchinnikov S; Steinegger M ColabFold: making protein folding accessible to all. Nat. Methods 2022, 19 (6), 679–682. [PubMed: 35637307]

(69). Bryant P; Pozzati G; Elofsson A Improved prediction of protein-protein interactions using AlphaFold2. Nat. Commun 2022, 13 (1), 1265. [PubMed: 35273146]

(70). Baek M; DiMaio F; Anishchenko I; Dauparas J; Ovchinnikov S; Lee GR; Wang J; Cong Q; Kinch LN; Schaeffer RD; Millán C; Park H; Adams C; Glassman CR; DeGiovanni A; Pereira JH; Rodrigues AV; van Dijk AA; Ebrecht AC; Opperman DJ; Sagmeister T; Buhlheller C; Pavkov-Keller T; Rathinaswamy MK; Dalwadi U; Yip CK; Burke JE; Garcia KC; Grishin NV; Adams PD; Read RJ; Baker D Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. Science 2021, 373 (6557), 871–876. [PubMed: 34282049]

(71). Anishchenko I; Pellock SJ; Chidyausiku TM; Ramelot TA; Ovchinnikov S; Hao J; Bafna K; Norn C; Kang A; Bera AK; DiMaio F; Carter L; Chow CM; Montelione GT; Baker D De Novo Protein Design by Deep Network Hallucination. Nature 2021, 600 (7889), 547–552. [PubMed: 34853475]

(72). Wang J; Lisanza S; Juergens D; Tischer D; Watson JL; Castro KM; Ragotte R; Saragovi A; Milles LF; Baek M; Anishchenko I; Yang W; Hicks DR; Expòsit M; Schlichthaerle T ; Chun J-H;

Dauparas J; Bennett N; Wicky BIM; Muenks A; DiMaio F; Correia B; Ovchinnikov S, Baker D Scaffolding Protein Functional Sites Using Deep Learning. Science 2022, 377 (6604), 387–394. [PubMed: 35862514]

(73). Dauparas J; Anishchenko I; Bennett N; Bai H; Ragotte RJ; Milles LF; Wicky BIM; Courbet A; de Haas RJ; Bethel N; Leung PJY; Huddy TF; Pellock S; Tischer D; Chan F; Koepnick B; Nguyen H; Kang A; Sankaran B; Bera AK; King NP; Baker D Robust Deep Learning–Based Protein Sequence Design Using ProteinMPNN. Science 2022, 378 (6615), 49–56. [PubMed: 36108050]

(74). Sekhar A; Kay LE NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. Proc. Natl. Acad. Sci. U.S.A 2013, 110 (32), 12867–12874. [PubMed: 23868852]

(75). Williamson MP; Kitahara R Characterization of low-lying excited states of proteins by high-pressure NMR. Biochim Biophys Acta Proteins Proteom 2019, 1867 (3), 350–358. [PubMed: 30366154]

(76). Sekhar A; Kay LE An NMR View of Protein Dynamics in Health and Disease. Annu. Rev. Biophys 2019, 48, 297–319. [PubMed: 30901260]

(77). Xie T; Saleh T; Rossi P; Kalodimos CG Conformational states dynamically populated by a kinase determine its function. Science 2020, 370 (6513), No. eabc2754. [PubMed: 33004676]

(78). Laurents DV AlphaFold 2 and NMR Spectroscopy: Partners to Understand Protein Structure, Dynamics and Function. Front. Mol. Biosci 2022, 9, 906437. [PubMed: 35655760]

(79). Wehmeyer C; Noé F Time-lagged autoencoders: deep learning of slow collective variables for molecular kinetics. J. Chem. Phys 2018, 148 (24), 241703. [PubMed: 29960344]

(80). Hernández CX; Wayment-Steele HK; Sultan MM; Husic BE; Pande VS Variational encoding of complex dynamics. Phys. Rev. E 2018, 97 (6–1), 062412. [PubMed: 30011547]

(81). Mardt A; Hempel T; Clementi C; Noé F Deep learning to decompose macromolecules into independent Markovian domains. Nat. Commun 2022, 13 (1), 7101. [PubMed: 36402768]

(82). Ribeiro JML; Bravo P; Wang Y; Tiwary P Reweighted autoencoded variational Bayes for enhanced sampling (RAVE). J. Chem. Phys 2018, 149 (7), 072301. [PubMed: 30134694]

(83). Sultan MM; Pande VS tICA-metadynamics: accelerating metadynamics by using kinetically selected collective variables. J. Chem. Theory Comput 2017, 13 (6), 2440–2447. [PubMed: 28383914]

(84). Shamsi Z; Cheng KJ; Shukla D Reinforcement Learning Based Adaptive Sampling: REAPing Rewards by Exploring Protein Conformational Landscapes. J. Phys. Chem. B 2018, 122 (35), 8386–8395. [PubMed: 30126271]

(85). Kleiman DE; Shukla D Multiagent Reinforcement Learning- Based Adaptive Sampling for Conformational Dynamics of Proteins. J. Chem. Theory Comput 2022, 18 (9), 5422–5434. [PubMed: 36044642]

(86). Zimmerman MI; Porter JR; Sun X; Silva RR; Bowman GR Choice of Adaptive Sampling Strategy Impacts State Discovery, Transition Probabilities, and the Apparent Mechanism of Conformational Changes. J. Chem. Theory Comput 2018, 14 (11), 5459–5475. [PubMed: 30240203]

(87). Noé F; Olsson S; Köhler J; Wu H Boltzmann generators—sampling equilibrium states of manybody systems with deep learning. Science 2019, 365 (6457), No. eaaw1147. [PubMed: 31488660]

(88). Noé F; Tkatchenko A; Müller KR; Clementi C Machine Learning for Molecular Simulation. Annu. Rev. Phys. Chem 2020, 71, 361–390. [PubMed: 32092281]

(89). Bonati L; Piccini G; Parrinello M Deep learning the slow modes for rare events sampling. Proc. Natl. Acad. Sci. U.S.A 2021, 118 (44), No. e2113533118. [PubMed: 34706940]

(90). Bonati L; Zhang YY; Parrinello M Neural networks-based variationally enhanced sampling. Proc. Natl. Acad. Sci. U.S.A 2019, 116 (36), 17641–17647. [PubMed: 31416918]

(91). Wang Y; Ribeiro JML; Tiwary P Past-future information bottleneck for sampling molecular reaction coordinate simultaneously with thermodynamics and kinetics. Nat. Commun 2019, 10 (1), 3573. [PubMed: 31395868]

(92). Wang D; Tiwary P State predictive information bottleneck. J. Chem. Phys 2021, 154 (13), 134111. [PubMed: 33832235]

(93). Mehdi S; Wang D; Pant S; Tiwary P Accelerating All-Atom Simulations and Gaining Mechanistic Understanding of Biophysical Systems through State Predictive Information Bottleneck. J. Chem. Theory Comput 2022, 18 (5), 3231–3238. [PubMed: 35384668]

(94). Tsai ST; Fields E; Xu Y; Kuo EJ; Tiwary P Path sampling of recurrent neural networks by incorporating known physics. Nat. Commun 2022, 13 (1), 7231. [PubMed: 36433982]

(95). Konovalov KA; Unarta IC; Cao S; Goonetilleke EC; Huang X Markov State Models to Study the Functional Dynamics of Proteins in the Wake of Machine Learning. JACS Au 2021, 1 (9), 1330–1341. [PubMed: 34604842]

(96). Litzinger F; Boninsegna L; Wu H; Nuske F; Patel R; Baraniuk R; Noe F; Clementi C Rapid Calculation of Molecular Kinetics Using Compressed Sensing. J. Chem. Theory Comput 2018, 14 (5), 2771–2783. [PubMed: 29660273]

(97). Brandt S; Sittel F; Ernst M; Stock G Machine Learning of Biomolecular Reaction Coordinates. J. Phys. Chem. Lett 2018, 9 (9), 2144–2150. [PubMed: 29630378]

(98). Mardt A; Pasquali L; Wu H; Noé F VAMPnets: deep learning of molecular kinetics. Nat.Commun 2018, 9 (1), 5. [PubMed: 29295994]

(99). Ghorbani M; Prasad S; Klauda JB; Brooks BR GraphVAMPNet, using graph neural networks and variational approach to Markov processes for dynamical modeling of biomolecules. J. Chem. Phys 2022, 156 (18), 184103. [PubMed: 35568532]

(100). Ghorbani M; Prasad S; Klauda JB; Brooks BR Variational embedding of protein folding simulations using Gaussian mixture variational autoencoders. J. Chem. Phys 2021, 155 (19), 194108. [PubMed: 34800961]

(101). Cao SQ; Montoya-Castillo A; Wang W; Markland TE; Huang XH On the advantages of exploiting memory in Markov state models for biomolecular dynamics. J. Chem. Phys 2020, 153 (1), 014105. [PubMed: 32640825]

(102). Fleetwood O; Kasimova MA; Westerlund AM; Delemotte L Molecular Insights from Conformational Ensembles via Machine Learning. Biophys. J 2020, 118 (3), 765–780. [PubMed: 31952811]

(103). Uyar A; Karamyan VT; Dickson A Long-Range Changes in Neurolysin Dynamics Upon Inhibitor Binding. J. Chem. Theory Comput 2018, 14 (1), 444–452. [PubMed: 29179556]

(104). Duro N; Varma S Role of Structural Fluctuations in Allosteric Stimulation of Paramyxovirus Hemagglutinin-Neuraminidase. Structure 2019, 27 (10), 1601–1611. [PubMed: 31402221]

(105). Zhou H; Dong Z; Verkhivker G; Zoltowski BD; Tao P Allosteric mechanism of the circadian protein Vivid resolved through Markov state model and machine learning analysis. PLoS Comput. Biol 2019, 15 (2), No. e1006801. [PubMed: 30779735]

(106). Tian H; Jiang X; Trozzi F; Xiao S; Larson EC; Tao P Explore Protein Conformational Space With Variational Autoencoder. Front. Mol. Biosci 2021, 8, 781635. [PubMed: 34869602]

(107). Tsuchiya Y; Taneishi K; Yonezawa Y Autoencoder-Based Detection of Dynamic Allostery Triggered by Ligand Binding Based on Molecular Dynamics. J. Chem. Inf. Model 2019, 59 (9), 4043–4051. [PubMed: 31386362]

(108). Zhu J; Wang J; Han W; Xu D Neural relational inference to learn long-range allosteric interactions in proteins from molecular dynamics simulations. Nat. Commun 2022, 13 (1), 1661. [PubMed: 35351887]

(109). Ward MD; Zimmerman MI; Meller A; Chung M; Swamidass SJ; Bowman GR Deep learning the structural determinants of protein biochemical properties by comparing structural ensembles with DiffNets. Nat. Commun 2021, 12 (1), 3023. [PubMed: 34021153]

(110). Matsunaga Y; Sugita Y Linking time-series of single-molecule experiments with molecular dynamics simulations by machine learning. Elife 2018, 7, No. e32668. [PubMed: 29723137]

(111). Dimura M; Peulen TO; Sanabria H; Rodnin D; Hemmen K; Hanke CA; Seidel CAM; Gohlke H Automated and optimally FRET-assisted structural modeling. Nat. Commun 2020, 11 (1), 5394. [PubMed: 33106483]

(112). ElGamacy M; Riss M; Zhu H; Truffault V; Coles M Mapping Local Conformational Landscapes of Proteins in Solution. Structure 2019, 27 (5), 853–865. [PubMed: 30930065]

(113). Bostock MJ; Solt AS; Nietlispach D The role of NMR spectroscopy in mapping the conformational landscape of GPCRs. Curr. Opin. Struct. Biol 2019, 57, 145–156. [PubMed: 31075520]

(114). Usher ET; Showalter SA Mapping invisible epitopes by NMR spectroscopy. J. Biol. Chem 2020, 295 (51), 17411–17412. [PubMed: 33453987]

(115). Zorba A; Nguyen V; Koide A; Hoemberger M; Zheng Y; Kutter S; Kim C; Koide S; Kern D Allosteric modulation of a human protein kinase with monobodies. Proc. Natl. Acad. Sci. U.S.A 2019, 116 (28), 13937–13942. [PubMed: 31239342]

(116). Grutsch S; Bruschweiler S; Tollinger M NMR Methods to Study Dynamic Allostery. PLoS Comput. Biol 2016, 12 (3), No. e1004620. [PubMed: 26964042]

(117). Aoto PC; Martin BT; Wright PE NMR Characterization of Information Flow and Allosteric Communities in the MAP Kinase p38gamma. Sri. Rep 2016, 6, 28655.

(118). Selvaratnam R; Chowdhury S; VanSchouwen B; Melacini G Mapping allostery through the covariance analysis of NMR chemical shifts. Proc. Natl. Acad. Sci. U.S.A 2011, 108 (15), 6133–6138. [PubMed: 21444788]

(119). Boulton S; Melacini G Advances in NMR Methods To Map Allosteric Sites: From Models to Translation. Chem. Rev 2016, 116 (11), 6267–6304. [PubMed: 27111288]

(120). Boulton S; Selvaratnam R; Ahmed R; Melacini G Implementation of the NMR CHEmical Shift Covariance Analysis (CHESCA): A Chemical Biologist's Approach to Allostery. Methods Mol. Biol 2018, 1688, 391–405. [PubMed: 29151219]

(121). Akimoto M; Martinez Pomier K; VanSchouwen B; Byun JA; Khamina M; Melacini G Allosteric pluripotency: challenges and opportunities. Biochem. J 2022, 479 (7), 825–838. [PubMed: 35403669]

(122). Khamina M; Martinez Pomier K; Akimoto M; VanSchouwen B; Melacini G Non-Canonical Allostery in Cyclic Nucleotide Dependent Kinases. J. Mol. Biol 2022, 434 (17), 167584. [PubMed: 35427632]

(123). Mohamed H; Baryar U; Bashiri A; Selvaratnam R; VanSchouwen B; Melacini G Identification of core allosteric sites through temperature- and nucleus-invariant chemical shift covariance. Biophys. J 2022, 121 (11), 2035–2045. [PubMed: 35538664]

(124). Bhattacharya S; Margheritis EG; Takahashi K; Kulesha A; D'Souza A; Kim I; Yoon JH; Tame JRH; Volkov AN; Makhlynets OV; Korendovych IV NMR-guided directed evolution. Nature 2022, 610 (7931), 389–393. [PubMed: 36198791]

(125). Köhler C; Carlström G; Gunnarsson A; Weininger U; Tångefjord S; Ullah V; Lepistö M; Karlsson U; Papavoine T; Edman K; Akke M Dynamic allosteric communication pathway directing differential activation of the glucocorticoid receptor. Sci. Adv 2020, 6 (29), No. eabb5277. [PubMed: 32832645]

(126). East KW; Newton JC; Morzan UN; Narkhede YB; Acharya A; Skeens E; Jogl G; Batista VS; Palermo G; Lisi GP Allosteric Motions of the CRISPR-Cas9 HNH Nuclease Probed by NMR and Molecular Dynamics. J. Am. Chem. Soc 2020, 142 (3), 1348–1358. [PubMed: 31885264]

(127). Nierzwicki L; East KW; Morzan UN; Arantes PR; Batista VS; Lisi GP; Palermo G Enhanced specificity mutations perturb allosteric signaling in CRISPR-Cas9. Elife 2021, 10, No. e73601. [PubMed: 34908530]

(128). Biddle JW; Martinez-Corral R; Wong F; Gunawardena J Allosteric conformational ensembles have unlimited capacity for integrating information. Elife 2021, 10, No. e65498. [PubMed: 34106049]

(129). Benton DJ; Wrobel AG; Xu P; Roustan C; Martin SR; Rosenthal PB; Skehel JJ; Gamblin SJ Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. Nature 2020, 588 (7837), 327–330. [PubMed: 32942285]

(130). Lu M; Uchil PD; Li W; Zheng D; Terry DS; Gorman J; Shi W; Zhang B; Zhou T; Ding S; Gasser R; Prevost J; Beaudoin-Bussieres G; Anand SP; Laumaea A; Grover JR; Lihong L; Ho DD; Mascola JR; Finzi A; Kwong PD; Blanchard SC; Mothes W Real-time conformational dynamics of SARS-CoV-2 spikes on virus particles. Cell Host Microbe 2020, 28 (6), 880–891. [PubMed: 33242391]

(131). Díaz-Salinas MA; Li Q; Ejemel M; Yurkovetskiy L; Luban J; Shen K; Wang Y; Munro JB Conformational dynamics and allosteric modulation of the SARS-CoV-2 spike. Elife 2022, 11, No. e75433. [PubMed: 35323111]

(132). Xu C; Wang Y; Liu C; Zhang C; Han W; Hong X; Wang Y; Hong Q; Wang S; Zhao Q; Wang Y; Yang Y; Chen K; Zheng W; Kong L; Wang F; Zuo Q; Huang Z; Cong Y Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. Sci. Adv 2021, 7 (1), No. eabe5575. [PubMed: 33277323]

(133). Dokainish HM; Re S; Mori T; Kobayashi C; Jung J; Sugita Y The inherent flexibility of receptor binding domains in SARS-CoV-2 spike protein. Elife 2022, 11, No. e75720. [PubMed: 35323112]

(134). Fallon L; Belfon KAA; Raguette L; Wang Y; Stepanenko D; Cuomo A; Guerra J; Budhan S; Varghese S; Corbo CP; Rizzo RC; Simmerling C Free Energy Landscapes from SARS-CoV-2 Spike Glycoprotein Simulations Suggest that RBD Opening can be Modulated via Interactions in an Allosteric Pocket. J. Am. Chem. Soc 2021, 143 (30), 11349–11360. [PubMed: 34270232]

(135). Zimmerman MI; Porter JR; Ward MD; Singh S; Vithani N; Meller A; Mallimadugula UL; Kuhn CE; Borowsky JH; Wiewiora RP; Hurley MFD; Harbison AM; Fogarty CA; Coffland JE; Fadda E; Voelz VA; Chodera JD; Bowman GR SARS-CoV-2 Simulations Go Exascale to Predict Dramatic Spike Opening and Cryptic Pockets across the Proteome. Nat. Chem 2021, 13 (7), 651–659. [PubMed: 34031561]

(136). Brotzakis ZF; Löhr T; Vendruscolo M Determination of intermediate state structures in the opening pathway of SARS-CoV-2 spike using cryo-electron microscopy. Chem. Sci 2021, 12 (26), 9168–9175. [PubMed: 34276947]

(137). Verkhivker GM; Di Paola L Dynamic Network Modeling of Allosteric Interactions and Communication Pathways in the SARS-CoV-2 Spike Trimer Mutants: Differential Modulation of Conformational Landscapes and Signal Transmission via Cascades of Regulatory Switches. J. Rhys. Chem. B 2021, 125 (3), 850–873.

(138). Verkhivker GM; Di Paola L Integrated Biophysical Modeling of the SARS-CoV-2 Spike Protein Binding and Allosteric Interactions with Antibodies. J. Phys. Chem. B 2021, 125 (18), 4596–4619. [PubMed: 33929853]

(139). Verkhivker GM; Agajanian S; Oztas DY; Gupta G Dynamic Profiling of Binding and Allosteric Propensities of the SARS-CoV-2 Spike Protein with Different Classes of Antibodies: Mutational and Perturbation-Based Scanning Reveals the Allosteric Duality of Functionally Adaptable Hotspots. J. Chem. Theory Comput 2021, 17 (7), 4578–4598. [PubMed: 34138559]

(140). Verkhivker GM; Agajanian S; Oztas DY; Gupta G Comparative Perturbation-Based Modeling of the SARS-CoV-2 Spike Protein Binding with Host Receptor and Neutralizing Antibodies: Structurally Adaptable Allosteric Communication Hotspots Define Spike Sites Targeted by Global Circulating Mutations. Biochemistry 2021 60 (19), 1459–1484. [PubMed: 33900725]

(141). Verkhivker GM; Agajanian S; Kassab R; Krishnan K Frustration-driven allosteric regulation and signal transmission in the SARS-CoV-2 spike omicron trimer structures: a crosstalk of the omicron mutation sites allosterically regulates tradeoffs of protein stability and conformational adaptability. Phys. Chem. Chem. Phys 2022 24 (29), 17723–17743. [PubMed: 35839100]

(142). Bhattacharjee S; Bhattacharyya R; Sengupta J Dynamics and electrostatics define an allosteric druggable site within the receptor-binding domain of SARS-CoV-2 spike protein. FEBS Lett. 2021, 595 (4), 442–451. [PubMed: 33449359]

(143). Spinello A; Saltalamacchia A; Borišek J; Magistrato A Allosteric Cross-Talk among Spike's Receptor-Binding Domain Mutations of the SARS-CoV-2 South African Variant Triggers an Effective Hijacking of Human Cell Receptor. J. Phys. Chem. Lett 2021, 12 (25), 5987–5993. [PubMed: 34161095]

(144). Rochman ND; Faure G; Wolf YI; Freddolino PL; Zhang F; Koonin EV Epistasis at the SARS-CoV-2 Receptor-Binding Domain Interface and the Propitiously Boring Implications for Vaccine Escape. mBio 2022, 13 (2), No. e0013522. [PubMed: 35289643]

(145). Starr TN; Greaney AJ; Hannon WW; Loes AN; Hauser K; Dillen JR; Ferri E; Farrell AG; Dadonaite B; McCallum M; Matreyek KA; Corti D; Veesler D; Snell G; Bloom JD Shifting mutational constraints in the SARS-CoV-2 receptor-binding domain during viral evolution. Science 2022, 377 (6604), 420–424. [PubMed: 35762884]

(146). Martin DP; Lytras S; Lucaci AG; Maier W; Gruning B; Shank SD; Weaver S; MacLean OA; Orton RJ; Lemey P; Boni MF; Tegally H; Harkins GW; Scheepers C; Bhiman JN; Everatt J; Amoako DG; San JE; Giandhari J; Sigal A; Williamson C; Hsiao N.-y.; von Gottberg A; De Klerk A; Shafer RW; Robertson DL; Wilkinson RJ; Sewell BT; Lessells R; Nekrutenko A; Greaney AJ; Starr TN; Bloom JD; Murrell B; Wilkinson E; Gupta RK; de Oliveira T; Kosakovsky Pond SL Selection analysis identifies clusters of unusual mutational changes in Omicron lineage BA.1 that likely impact Spike function. Mol. Biol Evol 2022, 39 (4), msac061. [PubMed: 35325204]

(147). Ou J; Lan W; Wu X; Zhao T; Duan B; Yang P; Ren Y; Quan L; Zhao W; Seto D; Chodosh J; Luo Z; Wu J; Zhang Q Tracking SARS-CoV-2 Omicron diverse spike gene mutations identifies multiple inter-variant recombination events. Signal Transduct. Target Ther 2022, 7 (l), 138. [PubMed: 35474215]

(148). Fan J; Liu Y; Kong R; Ni D; Yu Z; Lu S; Zhang J Harnessing Reversed Allosteric Communication: A Novel Strategy for Allosteric Drug Discovery. J. Med. Chem 2021, 64 (24), 17728–17743. [PubMed: 34878270]

(149). Tee W-V; Guarnera E; Berezovsky IN Reversing allosteric communication: From detecting allosteric sites to inducing and tuning targeted allosteric response. PLoS Comput. Biol 2018, 14, No. e1006228. [PubMed: 29912863]

(150). Ni D; Wei J; He X; Rehman AU; Li X; Qiu Y; Pu J; Lu S; Zhang J Discovery of cryptic allosteric sites using reversed allosteric communication by a combined computational and experimental strategy. Chem. Sci 2021, 12 (l), 464–476.

(151). Verkhivker G; Agajanian S; Kassab R; Krishnan K Probing Mechanisms of Binding and Allostery in the SARS-CoV-2 Spike Omicron Variant Complexes with the Host Receptor: Revealing Functional Roles of the Binding Hotspots in Mediating Epistatic Effects and Communication with Allosteric Pockets. Int. J. Mol. Sci 2022, 23 (19), 11542. [PubMed: 36232845]

(152). Wang Q; Wang L; Zhang Y; Zhang X; Zhang L; Shang W; Bai F Probing the Allosteric Inhibition Mechanism of a Spike Protein Using Molecular Dynamics Simulations and Active Compound Identifications. J. Med. Chem 2022, 65 (4), 2827–2835. [PubMed: 34415156]

(153). Wang L; Wu Y; Yao S; Ge H; Zhu Y; Chen K; Chen W; Zhang Y; Zhu W; Wang H; Guo Y; Ma P; Ren P; Zhang X; Li H; Ali MA; Xu W; Jiang H; Zhang L; Zhu L; Ye Y; Shang W; Bai F Discovery of Potential Small Molecular SARS-CoV-2 Entry Blockers Targeting the Spike Protein. Acta Pharmacol Sin 2022, 43 (4), 788–796. [PubMed: 34349236]

(154). Li B; Wang L; Ge H; Zhang X; Ren P; Guo Y; Chen W; Li J; Zhu W; Chen W; Zhu L; Bai F Identification of Potential Binding Sites of Sialic Acids on the RBD Domain of SARS-CoV-2 Spike Protein. Front. Chem 2021, 9, 659764. [PubMed: 34368076]

(155). Day CJ; Bailly B; Guillon P; Dirr L; Jen FE-C; Spillings BL; Mak J; von Itzstein M; Haselhorst T; Jennings MP Multidisciplinary Approaches Identify Compounds That Bind to Human ACE2 or SARS-CoV-2 Spike Protein as Candidates to Block SARS-CoV-2–ACE2 Receptor Interactions. mBio 2021, 12 (2), e03681–20.

(156). Taft JM; Weber CR; Gao B; Ehling RA; Han J; Frei L; Metcalfe SW; Overath MD; Yermanos A; Kelton W; Reddy ST Deep mutational learning predicts ACE2 binding and antibody escape to combinatorial mutations in the SARS-CoV-2 receptor-binding domain. Cell 2022, 185 (21), 4008–4022. [PubMed: 36150393]

(157). Zhao LP; Lybrand TP; Gilbert PB; Payne TH; Pyo AW; Geraghty DE; Jerome KR Rapidly identifying new coronavirus mutations of potential concern in the Omicron variant using an unsupervised learning strategy. Sci. Rep 2022, 12 (1), 19089. [PubMed: 36352021]

(158). Gross B; Sharan R Multi-task learning for predicting SARS-CoV-2 antibody escape. Front Genet 2022, 13, 886649. [PubMed: 36035121]

(159). Makowski EK; Schardt JS; Smith MD; Tessier PM Mutational analysis of SARS-CoV-2 variants of concern reveals key tradeoffs between receptor affinity and antibody escape. PFoS Comput. Biol 2022, 18 (5), No. e1010160.

(160). Köchi K; Schopper T; Durmaz V; Parigger L; Singh A; Krassnigg A; Cespugli M; Wu W; Yang X; Zhang Y; Wang WW; Selluski C; Zhao T; Zhang X; Bai C; Lin L; Hu Y; Xie Z; Zhang Z; Yan J; Zatloukal K; Gruber K; Steinkellner G; Gruber CC Optimizing variant-specific
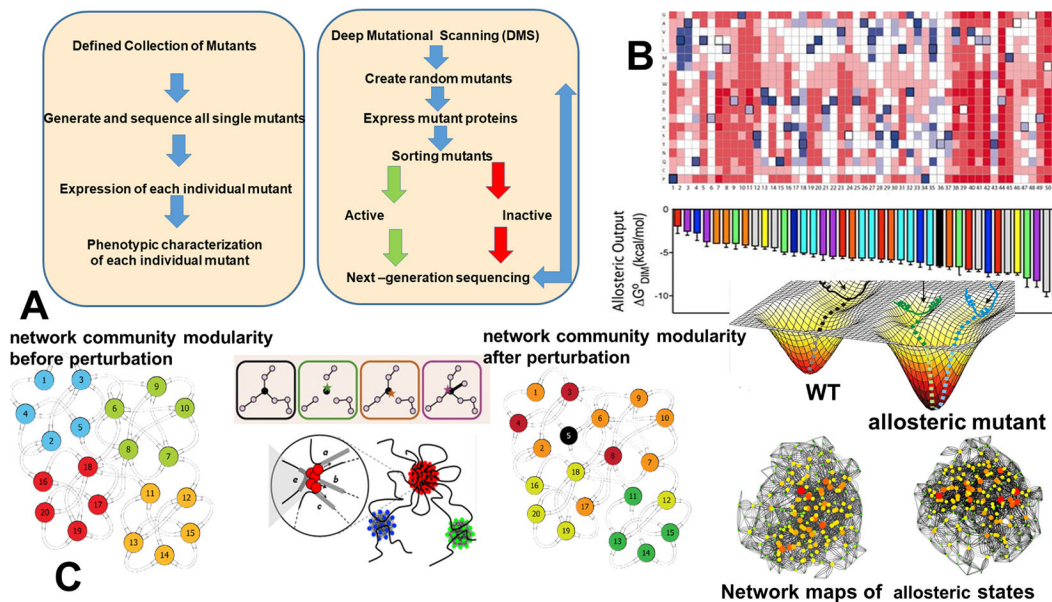
therapeutic SARS-CoV-2 decoys using deep-learning-guided molecular dynamics simulations. Sci. Rep 2023, 13 (1), 774. [PubMed: 36641503]
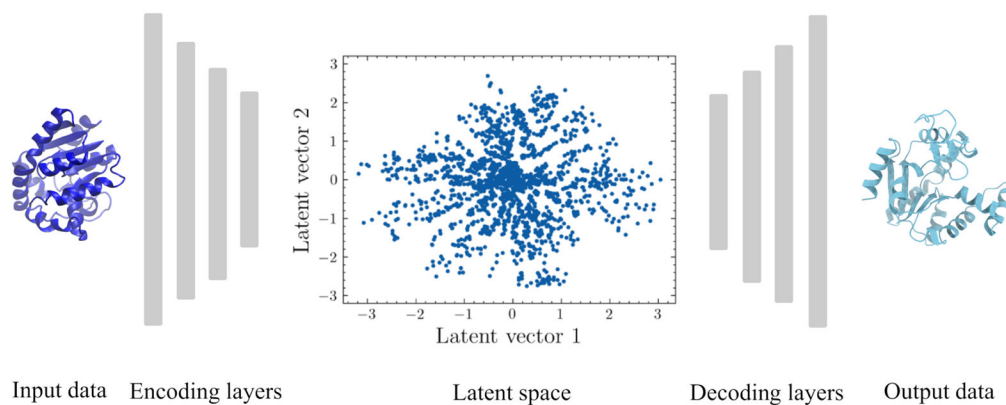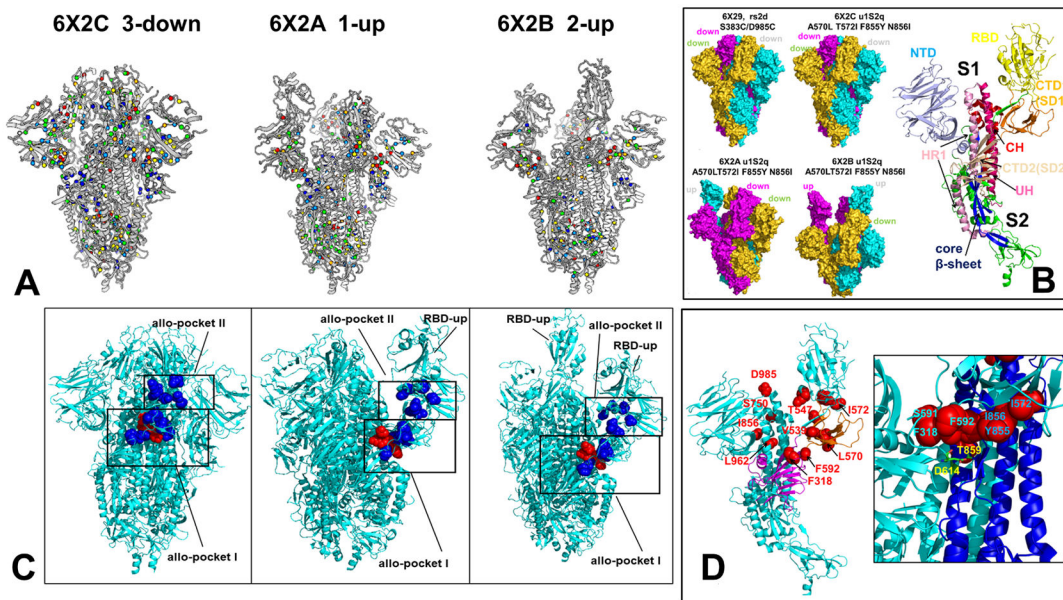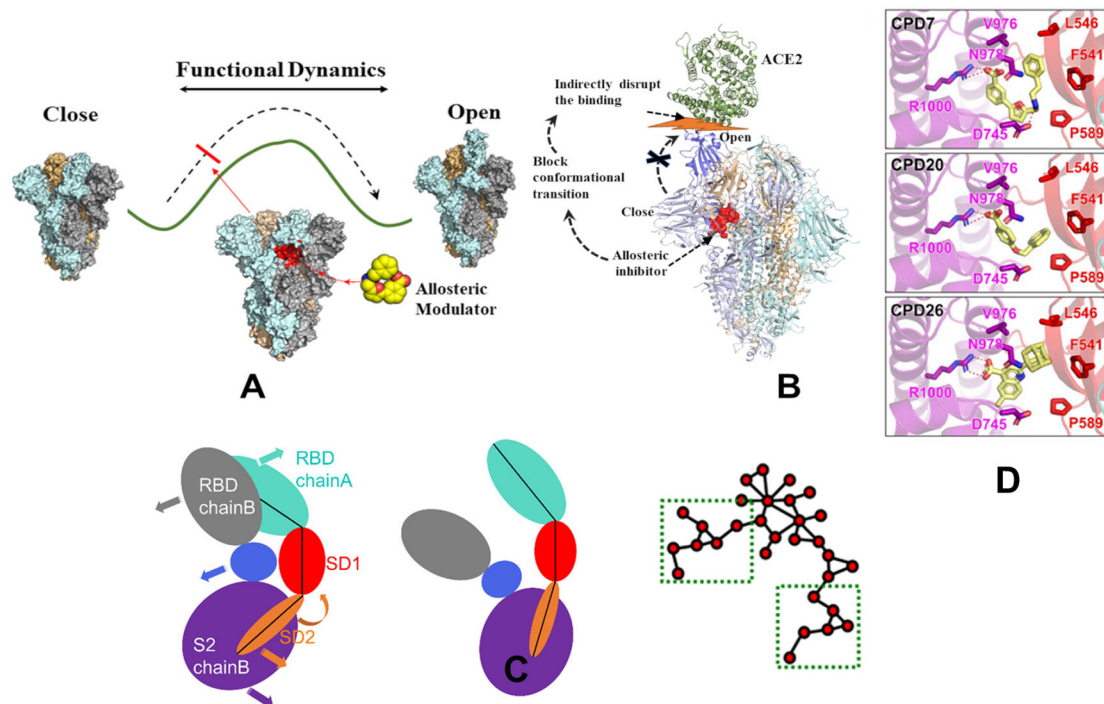
Figure 1.
(A) An overview of mutability landscape mapping approaches. The first approach is characterization of a defined collection of single mutants, and the second is DMS. In DMS the diversity in the protein is created through generation and expression of random mutants followed by high-throughput sorting of active mutants from inactive mutants using flow cytometry, phage display, or growth selection. This allows for the enrichment of active mutants. (B) The heatmap obtained from DMS analysis allows for comprehensive mapping of the phenotypic effects, for example, the decreased, neutral, or increased enzymatic activity, the effect of mutations on protein folding, and the mutational effect on binding, allowing for the identification of allosteric output of mutations and detection of allosteric hotspots. (C) A schematic description of dynamic network analysis of mutations mimicking DMS by monitoring the perturbation-induced changes in rewiring and community modularity of the graph-based residue interaction networks.

Figure 2.
Autoencoder architecture for exploring protein (adenosine kinase ADK as an example) conformational space.[106] Conformational space of a target protein kinase obtained from simulations could be used as training data to construct an encoder module to project protein structures in Cartesian coordinates onto the reduced space. The encoder module is designed with a decreasing number of neurons in hidden layers to encode high-dimensional inputs to a low-dimensional latent space. The decoder module, with an increasing number of neurons in hidden layers, aims to project latent space hack to the protein kinase conformations.

Figure 3.
(A) Molecular topography of allosteric communication pathways in the SARS-CoV-2 S 3RBD-down, 1RBD-up, and 2RBD-up trimers with regulatory hotspots (S383, D985, A570L, I572, I856, F592, and K987) serving as switches of allosteric signaling. (B) The cryo-EM structures of in the SARS-CoV-2 S 3RBD-down, 1RBD-up, and 2RBD-up trimers with protomers shown in green, blue, and cyan surface. A detailed annotation of the S protein elements in the S1 and S2 domains is shown for a single protomer. (C) Protein pocket scanning of network-derived allosteric positions identified major allosteric sites: hinge allo-pocket I at the junction of the subdomain 1 and subdomain 2 SD1/SD2 (F541, I587, K557, I834, Y855, I856, L570, I572 M740, D745, and L981) and allo-pocket II (A520, C361, A522, K528, K529, D389, F329, and T333). (D) Structural maps of the hinge regions in the SARS-CoV-2 spike trimer mutants and network-derived high centrality positions corresponding to allosteric mediating sites shown in red spheres.

Figure 4.
Elucidation of the allosteric binding pockets identified for designing molecular modulators of the S protein. (A) An allosteric path was discovered that correlates the motion of the RBD with the motion of the junction between the SD1 and the SD2 subdomains of the S protein. (B) Building on this finding, we designed non-RBD binding allosteric modulators to inhibit SARS-CoV-2 by prohibiting the conformational change of the S protein. (C) A schematic illustration of a cooperative conformational change of the S protein. (D) Three discovered potential molecular modulators and their binding modes. The top three compounds, i.e., CPD7, CPD20, and CPD26, were proven to be able to inhibit SARS-CoV-2 at a concentration of 100 μM. These compounds, especially the CPD7, show a stronger inhibitory effect, suggesting that these compounds may inhibit viral entry.