Trends in
Biochemical Sciences

CellPress

Review

Machine learning and protein allostery

Sian Xiao,[1,*] Gennady M. Verkhivker,[2,3] and Peng Tao [1,*]

The fundamental biological importance and complexity of allosterically regulated proteins stem from their central role in signal transduction and cellular processes. Recently, machine-learning approaches have been developed and actively deployed to facilitate theoretical and experimental studies of protein dynamics and allosteric mechanisms. In this review, we survey recent developments in applications of machine-learning methods for studies of allosteric mechanisms, prediction of allosteric effects and allostery-related physicochemical properties, and allosteric protein engineering. We also review the applications of machine-learning strategies for characterization of allosteric mechanisms and drug design targeting SARS-CoV-2. Continuous development and task-specific adaptation of machine-learning methods for protein allosteric mechanisms will have an increasingly important role in bridging a wide spectrum of data-intensive experimental and theoretical technologies.

Highlights

Machine-learning methods provide unprecedented opportunities for studies in understanding and exploiting protein allostery.

A large amount of data, including simulations related to protein allostery, were subjected to various types of machine-learning method to provide deeper insight into underlying allosteric mechanisms at levels of allosteric residues, pathways and networks, communities, and protein ensembles.

Machine-learning methods have done exceptionally well to develop prediction models for protein allosteric properties, including allosteric sites and effectors.

Allosteric protein engineering and design are emerging fields accumulating data for further applications of machine-learning methods.

Protein allostery has a key role in many computational studies using machine-learning methods targeting SARS-CoV-2, aiming to mitigate the COVID-19 pandemic.

Protein allostery at the intersection of modern molecular biology and data science

Allosteric regulation serves as an efficient strategy for molecular communication and is a common mechanism used by proteins for regulation of activity and adaptability [1–3]. Allosteric effects ensue when a certain perturbation occurs at a distal site of a protein that is topographically distinct from the orthosteric function site of the protein and consequently modulates the activity of that protein [1–3]. Since the term 'allostery' was introduced in 1961 [4], protein allostery has been one of the focuses of structural biology and is often referred as the 'second secret of life', second only to the genetic code [5,6]. A quantitative elucidation of such a fundamental and elusive phenomenon is critical to understanding life process and disease therapy [7–9]. It has been further proposed that all proteins are allosteric: even if it is not known to be allosteric, the protein could be observed to be allosteric under given conditions, such as the presence of appropriate allosteric effectors or mutations [10,11].

The remarkable progress and recent breakthroughs in **X-ray crystallography** (see Glossary), **nuclear magnetic resonance (NMR) spectroscopy**, **fluorescence resonance energy transfer (FRET)**, and **hydrogen–deuterium exchange mass spectrometry (HDXMS)** have enabled structural and dynamic studies of large biomolecules at atomic resolution and these technologies are often used as diagnostic tools in the study of allosteric interactions and communications in signaling proteins [12]. Recent advances in single-particle **cryogenic electron microscopy (cryo-EM)** have enabled the determination of near-atomic resolution structures for well-ordered proteins and large macromolecular assemblies, breaking resolution barriers for studies of allosteric events and allosteric drug discovery [13–15].

Computational approaches have complemented experimental methods and provided detailed molecular insights into allosteric transformations and regulatory mechanisms. **Molecular dynamics (MD) simulation**-based and **elastic network model (ENM)**-based approaches represent two main types of computational method to interrogate allosteric mechanisms based

[1]Department of Chemistry, Center for Research Computing, Center for Drug Discovery, Design, and Delivery (CD4), Southern Methodist University, Dallas, TX 75205, USA
[2]Graduate Program in Computational and Data Sciences, Schmid College of Science and Technology, Chapman University, Orange, CA 92866, USA
[3]Department of Biomedical and Pharmaceutical Sciences, Chapman University School of Pharmacy, Irvine, CA 92618, USA

*Correspondence:
sxiao@smu.edu (S. Xiao) and
ptao@smu.edu (P. Tao).

on protein dynamics [16–19]. Many other computational approaches correlate protein structural information at various levels with their allosteric functions. Allosteric molecular events involve a complex interplay of thermodynamic and dynamic changes that are difficult to observe, simulate, and interpret. The quantitative elucidation of these highly dynamic processes continues to present formidable technical and conceptual challenges [20].

Due to its universal importance, protein allostery has been studied with a wide range of approaches (Figure 1). The past decade has witnessed the rapid development of **machine-learning** and deep learning (DL) techniques and their applications to model increasingly complex chemical and biological phenomena [21–23]. In this review, we survey recent developments and applications of machine learning to protein allostery along three main themes: prediction and analysis of allosteric mechanisms; property prediction; and allosteric protein design. We also provide a perspective for the future development of computational and machine-learning approaches for studies of protein allostery.

Machine-learning studies of protein allostery

Dynamics-driven allosteric models have described protein allosteric mechanisms as signal propagation through dynamically modulated functional motions that can occur in the absence of visible structural changes. The current view recognizes that allostery can often involve an equilibrium

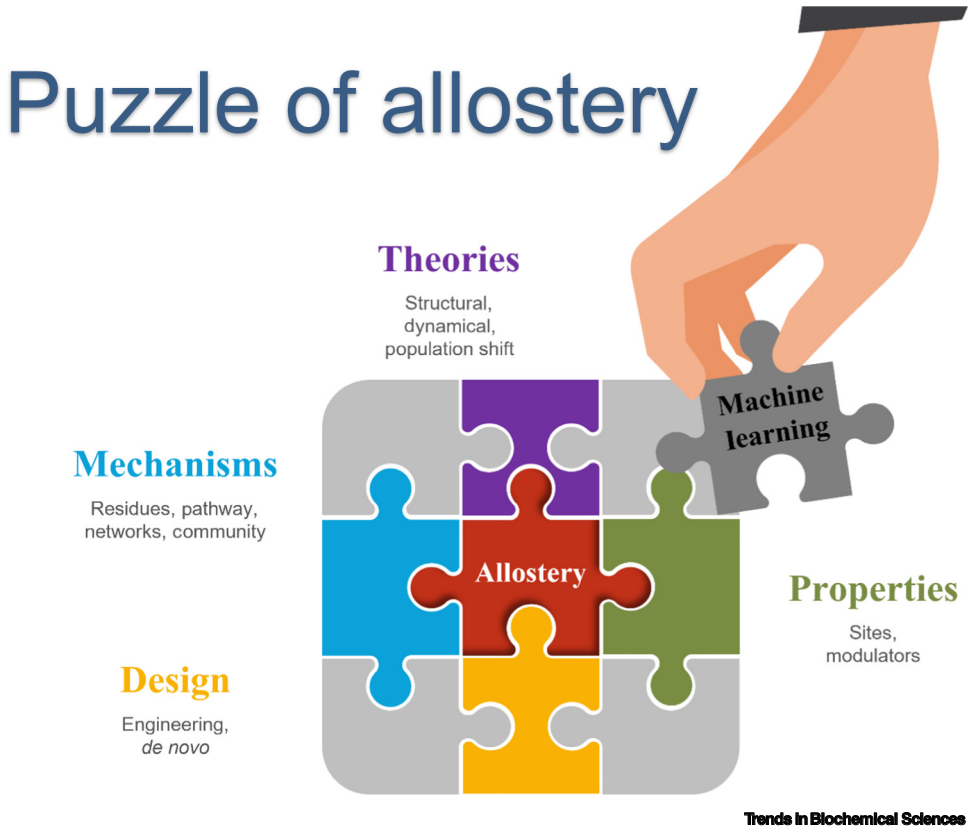

Figure 1. Solving the puzzle of allostery with machine learning. Protein allostery has been investigated from multiple aspects, including fundamental theories, allostery mechanisms, allostery-related properties, and allosteric protein design. With increasing amounts of information and data related to allostery available, machine-learning methods add another piece to the puzzle and have been used more widely to study protein allostery in various areas.

Glossary

Allosterome: systematic identification of protein allosteric interactions. It provides entire allosteric landscapes for related proteins of interest.
Angiotensin-converting enzyme 2 (ACE2): vital element in the renin–angiotensin–aldosterone system (RAAS) pathway that is critical for the regulation of processes such as blood pressure, wound healing, and inflammation. ACE2 helps modulate the many activities of angiotensin II (ANG II). When SARS-CoV-2 binds to ACE2, it prevents ACE2 from performing its normal function to regulate ANG II signaling.
Cryogenic electron microscopy (cryo-EM): microscopy technique applied to samples cooled to cryogenic temperatures. It can be used to provide 3D structural information about biological molecules and assemblies by imaging noncrystalline specimens. The structures of the samples are preserved by embedding them in a low-temperature environment.
Elastic network model (ENM): computational model used to describe proteins as structured elastic objects at a coarse-grained level. Proteins are treated as points in space with mass and connected by springs. ENMs can provide essential vibrational dynamics associated with the given structure and have been widely used to study protein dynamics, function, and conformational changes.
Fluorescence resonance energy transfer (FRET): distance-dependent physical process by which energy is transferred nonradiatively from an excited molecular fluorophore (the donor) to another fluorophore (the acceptor) by means of intramolecular long-range dipole–dipole coupling. FRET can be an accurate measurement of molecular proximity at distances between 10 and 100 Å and highly efficient if the donor and acceptor are positioned within the Förster radius (the distance at which half the excitation energy of the donor is transferred to the acceptor, typically 3–6 nm).
Hydrogen–deuterium exchange mass spectrometry (HDXMS): protein is exposed to D_2O and induces rapid amide H→D exchange in disordered regions that lack stable hydrogen bonding. Tightly folded elements are more protected from HDX, resulting in slow isotope exchange that is

shift of the pre-existing conformational ensembles due to effector binding [11,24]. In some perturbation-based simulation methods, mechanical forces are exerted on the allosteric proteins during simulations to probe protein dynamical and allosteric responses [25–29].

Combined with network models, these approaches can provide insight into mechanistic details of signaling pathways, predict the response to various perturbations, and guide the identification of regulatory sites. Despite the established view that many proteins function as dynamic and versatile allosteric regulatory machines, our atomistic understanding of allosteric mechanisms is still at a rudimentary level, and our knowledge of allosteric functional states and allosteric communication networks that govern diverse protein functions is surprisingly limited. Due to the lack of a universal theory, current studies aim to interpret protein allostery at various protein structural levels (Figure 2). A substantial challenge in investigating the allosteric mechanisms for large multidomain protein systems is the inherent difficulty of adapting experimental and computational methodologies to capture the intrinsic flexibility of these structures essential for functionality. The fundamental biological importance and complexity of these processes require a multifaceted platform of integrated approaches for characterization of allosteric functional states and atomistic reconstruction of allosteric regulatory mechanisms. In this review, we detail how machine-learning methods can be productively used to capitalize on the rapidly growing and rich multidimensional data on protein dynamics and allosteric protein landscapes. We suggest that machine-learning approaches have the potential to become a unifying data-centric research tool for synthesizing advances in theory and experimental technologies, ultimately leading to the development of robust and efficient computational models and expert systems for the prediction of diverse allosteric effects in protein systems.

Machine-learning approaches for molecular simulations and characterization of allosteric functional states

Without a universally accepted fundamental theory, experimental observations remain the foundation of protein allostery. New advances in experimental techniques often provide new insight into this ubiquitous phenomenon. For example, recent breakthroughs in single-molecule (sm) FRET technologies have enabled dynamic studies of large biomolecules. These advances provided semidirect experimental observation of allostery-related protein dynamics. Combined with MD simulations at the microsecond scale, smFRET experiments could directly probe transitions among allosteric states with significant conformational changes [30–32]. Recently, the DeepFRET method was developed using a DL model to bridge experimental data and protein dynamics [33]. These emerging experimental advances provide a solid foundation for computational and theoretical studies of protein allostery seen in recent years.

Machine-learning approaches have been widely used to facilitate conformational sampling with MD simulations via optimal selection of reaction coordinates [34–37], enhanced conformational sampling by active learning [38–40], and even autonomous generation of equilibrium ensembles without performing MD simulations [41]. With help from machine-learning methods, time-dependent structural changes can be quantitatively analyzed to provide insight into underlying allosteric mechanisms. Takami *et al.* [42] applied three time-series clustering methods, the unsupervised machine-learning technique for time-series data, to analyze multiple tight–relaxed state transition trajectories of human adult hemoglobin (HbA). These trajectories were classified by time-series clustering methods and analyzed to investigate the effect of oxygen molecules on the structural change of HbA.

In other cases, the allostery-related structural changes could not be easily recognized or characterized. Therefore, several machine-learning models have been developed to identify structural

mediated by the structural dynamics ('breathing motions') of the protein.
Machine learning: part of artificial intelligence; leverages data to improve performance on sets of tasks. It builds a model based on sample data, known as training data, to make predictions or decisions without being explicitly programmed to do so.
Markov state model (MSM): theoretical model used to study allosteric regulatory events. The first step is using robust dimensionality reduction techniques to identify suitable collective variables. Simulation data can be projected and represented by these collective variables. Clustering methods are applied to divide the simulation projection into metastable states. Transition probabilities among these metastable states can be estimated based on the simulation data.
Molecular dynamics (MD) simulation: computational method for analyzing the movements of atoms and molecules in space. The MD trajectories of atoms and molecules are determined by numerically solving Newton's equations of motion for a system of interacting particles. The forces between the particles and their potential energies are calculated using molecular mechanics force fields. These simulations can capture a variety of important biomolecular processes, including conformational change, ligand binding, and protein folding.
Normal mode analysis (NMA): provides vibrational modes accessible to a system in an equilibrium state, approximating the system in harmonic potentials. This computational model has been applied to identify and characterize the slow and global motions in a macromolecular system.
Nuclear magnetic resonance (NMR) spectroscopy: used to obtain information about the structure and dynamics of proteins, nucleic acids, and their complexes. The sample is placed inside a powerful magnet to measure the absorption of radiofrequency signals. Types of nucleus and distances between adjacent nuclei can be determined from absorption information and can be used to determine the overall structure of the protein. NMR spectroscopy can monitor both conformations and dynamics and can be applied to partially unfolded proteins.
X-ray crystallography: experimental technique to determine the 3D structure of a compound in crystal. The

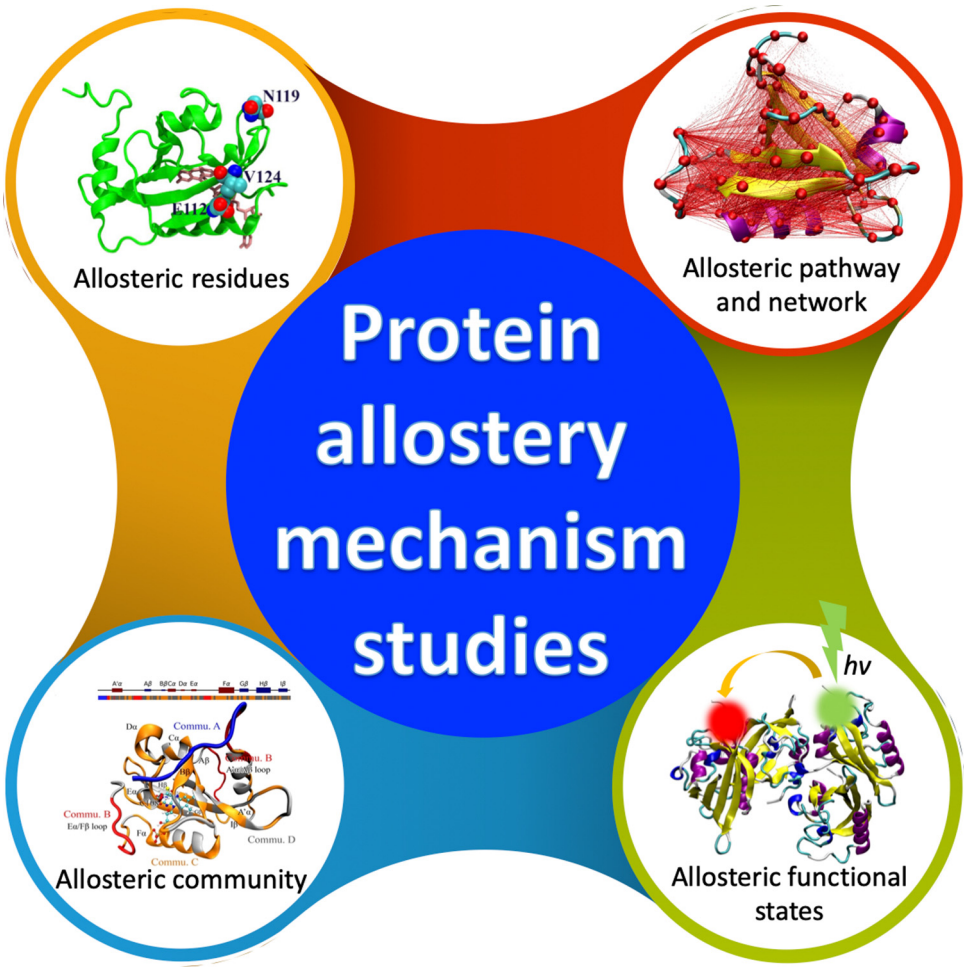

Figure 2. Allostery study facilitated by machine learning. Due to the lack of a universal theoretical framework for protein allostery, the mechanisms of allostery have been elucidated at multiple levels. At the residue level, key individual residues are identified as important for functions of the target allosteric proteins. At the pathway level, allosteric pathways comprising multiple residues are identified as main communication channels between the allosteric site and the main functional site. In some cases, multiple pathways could form networks to enable allosteric signal transduction within the protein structure. The allosteric community comprises a group of closely related residues associated with allostery. Allosteric protein structure could be divided into several communities, which interact with each other synergistically to carry out allosteric functions. From a dynamical point of view, proteins need to transition between different functional states when fulfilling their allosteric functions. These allosteric functional states could be identified through both computational and experimental studies.

features that can properly describe the slowest dynamics underlying conformational changes. These features could be used to model protein kinetics that underlie allosteric processes [43–45].

Identifying key protein allosteric residues

As fundamental building blocks of proteins, individual residues are the focus of many protein allosteric mechanism studies. Machine-learning methods provide quantitative means to correlate global protein allostery with individual residues. Many studies aim to identify key residues for protein allostery through informative and insightful analysis of protein dynamics data using various machine-learning methods. Zhou *et al*. [46] applied supervised learning methods [decision trees and neural networks (NNs)] to build classification models for allosteric states based on

crystallized sample is exposed to an X ray beam to obtain diffraction patterns. These patterns can be processed to yield information about the crystal packing symmetry, the size of the repeating unit, and a map of the electron density. The molecular structure can be built and refined based on electron density information from diffraction patterns.

the simulation data of the second PDZ domain from human PTP1E protein (PDZ2). These accurate classification models provide numerical measurement of the importance of each residue for overall allostery. The key allosteric residues identified based on this importance were in agreement with results from experimental and computational studies. Similarly, Hayatshahi *et al*. [47] applied deep NNs (DNNs) to build a classification model of the PDZ3 domain from the adaptor protein PSD95 to distinguish otherwise-similar allosteric states using MD simulation data. Their classification model, a residue response map as a 2D property-residue map, could be constructed to represent allosteric effects as residue-specific properties. More recently, Do *et al.* [48] introduced a Gaussian-accelerated MD (GaMD), DL, and free energy profiling workflow (GLOW) to characterize both activation and allosteric modulation of a G protein-coupled receptor (GPCR). A convolutional NN (CNN) model was used in GLOW to classify the residue contact maps, from which important residues could be identified.

Mapping of allosteric networks and communication pathways using machine learning

The nature and atomistic details of the allosteric communication between the allosteric site and the functional site are often difficult to dissect. Experimental approaches could reveal allosteric hotspots and potential communication pathways in protein structures. Using a combination of mutagenesis, mass spectrometry, amide HDXMS, and FRET studies, the atomistic details and allosteric pathways of the Hsp70 chaperone regulation mechanisms have been mapped, revealing the previously unrecognized dichotomy of allosteric control in the chaperone [49–51].

There are many machine learning-based methods to identify allosteric pathway or networks. Graph theory-based methods are among the most widely used approaches. By mapping dynamic fluctuations onto a graph, network-based approaches can describe signal transmission via cascades of coupled residue fluctuations and characterize allosteric communication pathways in proteins. Zhu *et al.* [52] applied a graph NN (GNN)-based neural relational inference (NRI) model, which adopts an encoder–decoder architecture, to simultaneously infer latent interactions for probing protein allosteric processes as dynamic networks of interacting residues. From the MD trajectories, this model successfully learned the long-range interactions and pathways that can mediate allosteric communications between distant sites. Machine-learning methods could also be applied to develop various dynamic network models of allosteric interactions to decrypt the underlying mechanisms driving allosteric effects in proteins [53].

Other machine learning-based methods use various correlation relations among residues to identify potential allosteric pathways. Zhou *et al.* [54] used the relative entropy concept from information theory to develop a relative entropy-based dynamical allosteric network (REDAN) model. The relative entropy is used to measure the response of each residue pair to external perturbation. The potential allosteric pathways are identified as a series of short-range residue pairs with the most significant response. Botlani *et al.* [55] extended the underlying mechanism of allostery by exploring correlation between ensembles of protein in different allosteric states. They applied a support vector machine (SVM) approach to quantitively evaluate these correlations using simulations representing different allosteric states of the same protein. Undirected weighted graph theory was also used to identify the shortest pathway possible for allosteric signaling mechanisms. Yan *et al.* [56] proposed the node-weighted amino acid contact energy network (NACEN) to characterize and predict three types of functional residue: hot spots, catalytic residues, and allosteric residues. These studies demonstrate the viability and diversity, as well as uncertainty, of using machine-learning methods to evaluate allosteric contribution from individual residues.

Allosteric community models divide different residues into different groups, referred to as communities. These allosteric communities are not necessarily correlated with protein secondary or tertiary structural components and could provide a higher level of information compared with pathway and network models. Zhou *et al.* [57] and Ibrahim *et al.* [58] developed a community analysis algorithm based on their machine learning-based classification model for protein allostery. The allosteric communities are built in such a way that the impacts of external perturbations on the distribution differences are maximum across different communities and minimum within the same community. This algorithm was applied to reveal the allosteric mechanism of the fungal circadian clock photoreceptor Vivid (VVD), as one member of the light–oxygen–voltage (LOV) domain, upon photoactivation. Interestingly, two distal loop regions were identified in the same community. This means that, despite the distance between these two secondary structures, residue pairs across these two loop regions in VVD carry minimal allosteric significance. By contrast, these two loops together make a significant contribution to the overall allosteric effects.

Stetz and Verkhivker [59] applied a graph-based model to Hsp70 chaperones to construct residue interaction networks. The allosteric communities in Hsp70 were constructed as stable clusters of residues along the simulations. Astl *et al.* [60] and Stetz *et al.* [61] developed allosteric community models for Hsp90 through residue interaction network analysis and noted that different allosteric communities were correlated through intermodular pathways for long-range communications. They also applied community models to characterize functional mechanisms of Hsp90 allosteric modulation through binding with various allosteric modulators as well as other protein domains for its regulation [62,63]. Chen *et al.* [64] applied dynamic network analysis to build a community model to reveal the regulatory effect on GPCR of binding with G protein-mimicking Nanobody80 (Nb80). Both supervised (NN) and unsupervised (principal component analysis) learning methods were used for feature extraction and key residue identification of the dynamical response to the binding event. Compared with pathway and network models, allosteric community models do not target certain sites in a protein and could provide a more comprehensive view of underlying protein allosteric mechanisms. Based on protein dynamics, these community models offer alternative protein structures related to allostery other than conventional primary, secondary, and tertiary structures. The communities within protein structures identified in these allosteric community models provide functional information regarding protein allostery in addition to convention secondary and tertiary structural information.

Machine-learning approaches for prediction of allosteric binding sites, hotspots, and phenotypes, and applications in allosteric drug design

Allosteric drug development is among the most promising fields based on allostery for many reasons: allosteric drugs could be more selective and less toxic with fewer side effects; they can either activate or inhibit proteins; and they can be used in conjunction with orthosteric drugs. Discovery of allosteric drugs presents challenges beyond those encountered in orthosteric drug discovery. To address this challenge, Zhang and coworkers constructed the AlloSteric Database (ASD) [65] and ASBench [66]. ASD is a platform providing comprehensive information about allosteric proteins and their modulators. The database now contains a total of 1949 allosteric protein entries. ASBench, an optimized selection of ASD data, includes a core set with 235 unique allosteric sites and a core-diversity set with 147 structurally diverse allosteric sites. However, in many cases, the location of allosteric sites is unknown. It is also difficult to accurately predict whether the drug will activate or inhibit the protein strength of the allosteric regulation [67,68]. Leveraging existing sample data using machine learning and DL to make predictions or decisions can help predict the allosteric components.

Predicting allosteric sites

Several methods have been developed to detect and predict allosteric sites in proteins. These studies can be classified as sequence-based, structure-based, dynamics-based, **normal mode analysis (NMA)**-based, or combined prediction approaches [69]. Machine learning can help with the detection task since it can deal with numerous input features, including local or static features of pockets and delocalized or dynamic features of proteins (Table 1).

The static features, such as pocket volume, pocket flexibility, and pocket hydrophobicity, characterize the conformation of protein pockets, and also provide information for classifiers to identify allosteric sites. Akbar and Helms [70] characterized allosteric pockets using a set of physicochemical descriptors and trained a predictive model based on Naive Bayes and artificial NNs. The predictive models were capable of prioritizing allosteric pockets in a set of pockets found on a given protein and were encapsulated in the publicly accessible program ALLO. Tian *et al.* [71] and Xiao *et al.* [72] adapted an ensemble learning method combining eXtreme gradient boosting (XGBoost) and graph CNNs (GCNNs), and an automated machine-learning method (AutoGluon and AutoKeras) to predict plausible allosteric sites. They deployed both models to the Protein Allosteric Sites Server[i] [71,72]. Chen *et al.* [73] used the structures of the sites and the co-crystallized ligands to calculate 43 structural descriptors. These structural descriptors were used to build a three-way predictive model based on random forest to characterize protein–ligand binding sites as allosteric, regular, or orthosteric. Huang *et al.* also applied SVM for the prediction of allosteric sites using static pocket features, resulting in the web server Allosite [74].

Dynamic features were also used for allosteric site prediction because allostery is a dynamic behavior of the whole protein. Greener *et al.* used perturbed NMA and pocket descriptors in SVM to sort pockets in proteins and developed the AlloPred web server to predict allosteric pockets [75]. Song *et al.* [76] combined pocket features with NMA-based perturbation analysis to build a logistic regression model, AllositePro, to predict allosteric sites in proteins.

Other features were also explored for allosteric site prediction. Mishra *et al.* [77] used various features at the residue level, including amino acid physicochemical properties, rate of residue evolution, and features for protein geometry and dynamics, to build the Active and Regulatory site Prediction (AR-Pred) model. Fogha *et al.* [78] found that crystal additives (CAs), which stabilize

Table 1. Representative allosteric site prediction methods

Features	Methods	Data sets[a]	Refs
Static pocket features	Naive Bayes and neural networks	ASD and ASBench	[70]
	GCNN with XGBoost	ASD	[71]
	Automated machine learning	ASD and ASBench	[72]
	Random forest	ASD[b]	[73]
	Support vector machine	ASD	[74]
Pocket features with NMA perturbation	Support vector machine	ASBench	[75]
	Logistic regression	ASBench	[76]
Features at residue level	Random forest	ASBench	[77]
Crystal additive location	DBSCAN	ASD and ASBench[c]	[78]

[a]The original data sets used to obtain allosteric site data. The data were filtered for high-resolution and non-redundant structures individually.
[b]The PDBbind database was used to obtain information on orthosteric sites.
[c]The RCSB PDB was used to obtain protein–crystallographic additive complexes.

proteins during the crystallization process, tend to aggregate in protein hotspots, especially near the binding cavities; thus, CAs can be used as a criterion on which to base site-type decisions. The authors proposed an efficient and easy way to use the structural information of CAs to identify allosteric sites.

With comparable accuracy but using different methods, these prediction models for allosteric sites provide ample choice for users. One could also apply methods using different strategies in the same study and use the consensus results for improved outcomes. The workflow of an allosteric site analysis web server AlloFinder is illustrated in Figure 3.

The reversed allosteric communication theory [79] has been used successfully in several studies. It is based on the premise that allosteric signaling in proteins is bidirectional and can propagate from

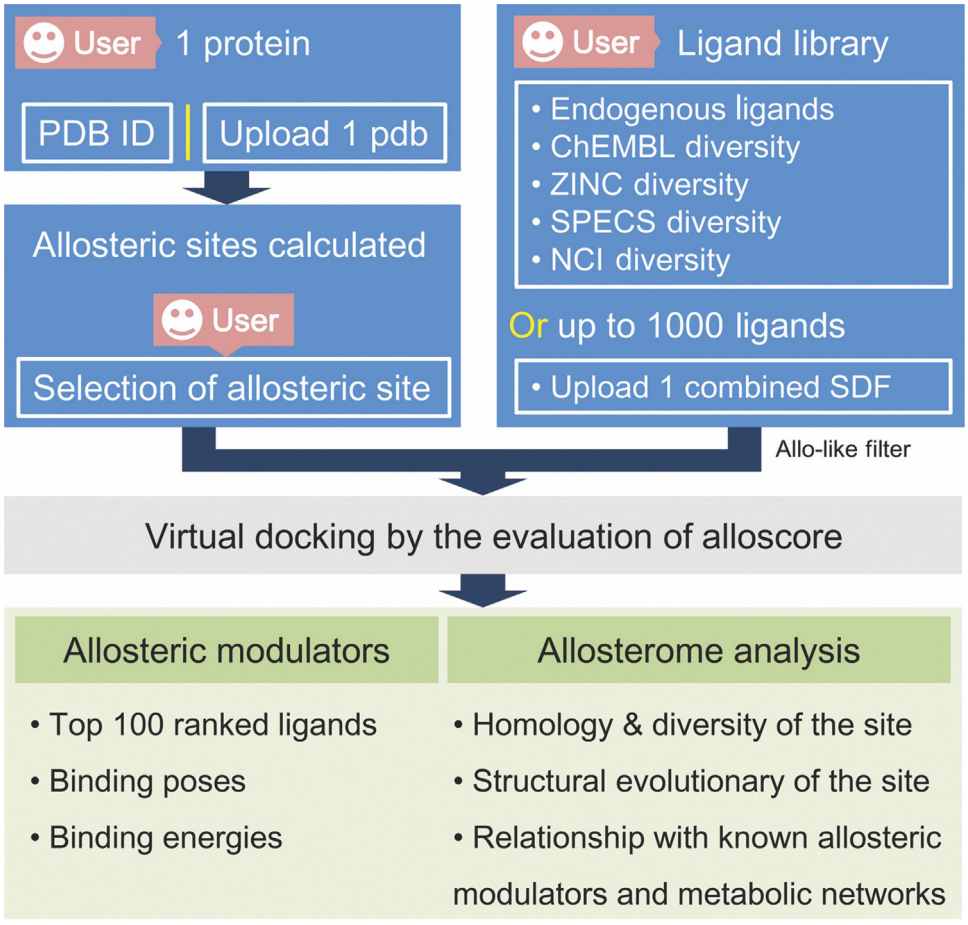

Figure 3. The workflow of the AlloFinder web server. After the user uploads a query protein to AlloFinder, all putative allosteric sites on the protein are predicted. The user can choose one allosteric site to screen a predefined ligand library virtually. The pocket-generated pharmacophore model for the selected allosteric site is generated for quickly ruling out unbound compounds in the library. Conformational sampling of an ensemble of docked conformations is performed for each compound. The most favorable binding energy of each compound is evaluated and ranked. The top 100 compounds are provided by the AlloFinder web server. Finally, the predicted allosteric sites and modulators are harnessed to perform allosterome-mapping analyses of the human proteome [94]. Abbreviation: SDF, Structure Data File.

an allosteric to an orthosteric site and vice versa [80,81]. Some reversed allosteric communication approaches are rooted in dynamic network-based models of inter-residue interactions [82,83]. An integrated computational and experimental strategy exploited reversed allosteric communication concepts to combine MD simulations with **Markov state models (MSMs)** for characterization of binding shifts in protein ensembles and identification of cryptic allosteric sites [84]. In MSMs, dimensionality reduction techniques are used to generate suitable collective variables to characterize protein conformational space. The simulation of allosteric proteins could be projected into the space using these collective variables as the distribution in the conformational space. Clustering methods are generally applied to cluster these distributions into metastable states. Accordingly, transition probabilities among these metastable states could be estimated based on the simulation data.

Using the reversed allosteric communication concept, machine-learning methods enable reconstruction and analysis of the comparative perturbed ensembles of the allosteric states and characterize redistribution of dynamic states in inhibitor-bound versus inhibitor-free systems following allosteric binding [85]. These machine learning-based models as either classification or regression models cannot account for the signal transduction between the distal sites and function-related active sites because no such information is included in the training data to develop these models.

Given that predicted allosteric sites could be used directly for allosteric drug development and due to recent breakthroughs in protein structure prediction, including AlphaFold2 and others, allosteric site prediction methods have huge potential for furthering our understanding of protein allostery.

Machine-learning models based on deep mutational scanning

Currently, experimental data remain the primary foundation for the development of allostery-related computational models for understanding, predicting, and engineering biophysical properties of allosteric proteins. Emerging deep mutational scanning (DMS) experiments combine saturation mutagenesis of a protein with a high-throughput functional test and deep sequencing and provide unbiased and systematic single mutational information of target proteins. Such large and quantitative data sets enable machine-learning approaches to predict allosteric properties from sequences. Leander *et al.* [86] carried out DMS of four homologous bacterial allosteric transcription factors (aTFs). They further developed prediction models using NN models and genetic algorithms to identify hotspots of homolog proteins and to predict the structural and molecular properties of allosteric hotspots. Faure *et al.* [87] generated mutagenesis libraries of the C-terminal SH3 domain of the human growth factor receptor-bound protein 2 (GRB2-SH3) and third PDZ domain from the adaptor protein PSD95 (PSD95-PDZ3) domains, which contain both single and double amino acid substitutions. A NN model was developed using DMS data to predict the binding free energy change upon single amino acid substitutions in both systems. These prediction models were used to map the energetic and allosteric landscapes of the target domains.

These recent studies demonstrate the potential of DMS data to facilitate the development of machine learning-based methods for protein allostery-related properties at the residue level and even theoretical landscaping models for protein allostery.

Evaluating allosteric effectors

Binding with allosteric modulators is the main allosteric perturbation in many cases. Some studies aimed to distinguish allosteric modulators from nonallosteric modulators. Several physically relevant compound descriptors of molecules were computed, and the feature differences were then correlated with chemical property differences. Wang *et al.* [88] and Smith *et al.* [89] concluded

that allosteric modulators are generally more aromatic, structurally rigid, and more hydrophobic. This general idea can help with preliminary screening of allosteric modulators.

Similar to using machine-learning models to identify allosteric sites for proteins, machine-learning models could be developed to classify modulators as allosteric or nonallosteric. For example, Hou *et al.* [90] trained six types of machine-learning model using different combinations of features for an 11-class classification task with ten GPCR subtype classes and a random compounds class. This was the first study of the multiclass classification of GPCR allosteric modulators.

Other studies focus on developing generative models to build and evaluate allosteric inhibitors targeting various receptors. Different methods and training data have been used to develop various machine learning-based models with comparable performance. Bian and Xie [91] first established a general molecule generation model (g-DeepMGM) with a half million compounds collected from the ZINC database, and then constructed a target-specific molecule generation model (t-DeepMGM) based on the transfer learning process of reported cannabinoid receptor 2 (CB2) ligands. Yang *et al.* [92] first trained a Transformer-encoder-based generator on the 1.6 million data sets in ChEMBL to learn the grammatical rules of known drug molecules. Transfer learning is used to introduce the prior knowledge of drugs with known activities against particular targets into the generative model to construct new molecules similar to the known ligands. Reinforcement learning is used to combine the generative model and the predictive model to generate molecules with drug-like properties that are expected to bind well with the target.

Vennila and Elango [93] used the voxelized representation of five different conformational states of the PDK1 allosteric site (the PIF pocket) to predict 1D SMILES imparted in the LiGANN pipeline in the playmolecule platform, in which, for a given protein shape, a generative adversarial NN (GANN) produces complementary ligand shapes in a multimodal fashion. Huang *et al.* [94] built AlloFinder, which identifies potential endogenous or exogenous allosteric modulators and their involvement in the human **allosterome**. AlloFinder automatically amalgamates allosteric site identification, allosteric screening, and allosteric scoring evaluation of modulator–protein complexes to identify allosteric modulators, followed by allosterome mapping analyses of predicted allosteric sites and modulators in the human proteome. More recently, Miljković *et al*. [95] applied random forest, SVM, and DNN models to predict different classes of kinase inhibitor targeting different allosteric sites. Compounds were represented using molecular fingerprints without other structural information being considered. Given that the authors were struck by the consistently good performance across different methods used in this study, this demonstrated that machine-learning methods in general could extract key chemical features for certain properties using appropriate features.

Identifying receptors for allosteric inhibitors

In some scenarios, potential receptors need to be identified for known substrates with significant pharmaceutical effects. These substrates may include allosteric effectors interacting with pharmacology networks. Rodrigues *et al*. [96] developed a novel strategy to identify potential targets of known allosteric effectors using self-organizing map-based prediction of drug equivalence relationships (SPiDER) model. This model uses a consensus of unsupervised self-organizing maps, consensus scoring, and statistical analysis to identify potential targets for known active substrates. Using this approach, the authors identified 5-lipoxygenase as an allosteric inhibiting target for β-lapachone as a clinical-stage, natural product with thorough validation. As an emerging field of computer-aided molecule design, there are many potential directions in which machine-learning methods could be applied specifically for allosteric modulator development.

Machine-learning studies for allosteric protein design

One of the goals of studying protein allostery is developing novel proteins that carry improved or novel allosteric functions. Given that this is a new area, large amounts of data related to the allostery of different proteins have yet to be used in the developing process. In an early study, Zayner *et al*. [97] studied over 100 mutations of *Avena sativa* LOV domain 2 (AsLOV2) as a light-activated protein. In this experimental study, the authors used various experimental methods to characterize the target mutations of AsLOV2. The biggest lesson learned through this study was that most mutations, which were expected to be highly disruptive substitutions, turned out to be modest or had no effect on function, even with many mutations displaying enhanced photoactivity. These counterintuitive results signify the importance of a deeper and more comprehensive understanding of protein allostery in the effort to design an enhanced or novel allosteric molecular apparatus.

Weinkam *et al*. [98] used simulation data of a set of ten proteins and their mutations to build prediction models for allostery. They built a decision-based machine-learning model with a wide range of features, including geometric- and energy-based features, to predict mutational effects on protein allosteric activity. This prediction function will help with protein-engineering efforts to develop modified protein allosteric activities and functions. Xiao *et al*. [99] used systematic machine-learning approaches to analyze the allostery of thrombin as a multifunctional serine protease at the conformational ensemble level. Their study provided mechanistic insight into allostery of one key thrombin mutant with ample intramolecular interaction details.

Successful cases of allosteric protein design are still mainly based on the expertise and experience of researchers. For example, García-Fernández *et al.* [100] developed a novel biosensor by fusing two ion channels, a tetrameric viral Kcv channel and the dimeric mouse TREK-1 channel, to a physiologically unrelated membrane GPCR protein. The GPCR displayed regulatory efforts toward both fused ion channels. The authors fine-tuned the length of linkers connecting GPCR with the two ion channels. The successful fusion between two physiologically unrelated allosteric proteins to design a novel biosensor indicates a direction for computational studies based on structural and simulation data and machine-learning modeling to identify potential candidates and appropriate designs for linkers. In a more traditional study, D'Amico *et al.* [101] developed enhanced tryptophan synthases through mutations at a distant, surface-exposed network residue. It is expected that data-driven strategies using machine-learning methods could catalyze the breakthrough in allosteric proteins designs in the near future.

There is at least one study using a machine-learning method to model evolutionary relations among allosteric proteins. Astl and Verkhivker [102] used a systematic approach and carried out ENM analysis of 235 unique allosteric protein entries from ASBench. Using residue interaction network models of the target proteins, they evaluated the coevolution of key residues for different allosteric proteins and identified unifying molecular signatures shared by allosteric systems. Application of their models to protein kinases revealed molecular signatures of known regulatory allosteric residues. Allostery-related protein evolution is a relatively uncharted area, mainly due to the lack of unified theoretical models of protein allostery. The applications of suitable machine-learning models to correlate protein allosteric mechanisms with evolution point to a new direction for deciphering protein allostery.

Concluding remarks and future perspectives

Allostery is an intrinsic but elusive ubiquitous phenomenon in proteins. We have reviewed research progress in protein allostery using machine-learning methods in various frontiers. Although many theories and models have been developed to interpret this phenomenon, there is no simple equation to quantify allostery. Machine learning helps explain the mechanism in different dimensions, residues, pathways, networks, and communities. Given that there is no

Outstanding questions

How can the underlying mechanisms of protein allostery be formulated at different structural levels, including individual residues, allosteric pathways, and networks?

How can advanced experimental techniques, such as smFRET, be used to characterize protein allostery at the microscopic level?

Could protein ensembles generated from simulations be used directly to shed light on underlying allosteric mechanisms?

With their ability to analyze large amounts of data to build highly performing prediction models, how can machine-learning methods be used to develop prediction models for allosteric sites?

How can potential modulators targeting allosteric proteins with desired properties be effectively developed using machine learning-based approaches?

Is it feasible to engineer or even develop novel allosteric proteins with desired properties? If so, how could machine-learning methods be used to facilitate these developments?

How can machine learning-based computational analysis and prediction methods related to protein allostery be used to address the pharmaceutical challenges caused by the COVID-19 pandemic?

universal theory for all allosteric regulations, it might be that protein allostery theory or mechanisms cannot be unified because of the diversity of protein structures and dynamical behaviors.

Important implementations of protein allostery include the prediction of various protein allostery-related properties. Suitable for processing large amounts of data and developing reliable prediction models in general, data-driven machine-learning methods have been applied to develop computational models to predict protein allosteric binding sites and modulators. Those prediction models have been made available with easy access to the research community and have been widely used in many studies related to protein allostery. The biggest impact made on protein allostery studies using machine-learning methods is mainly in applications, as demonstrated by the emphasis on machine-learning method-based approaches that focused on allosteric mechanisms of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Box 1) and modulators as ligands to target various receptors in this virus (Box 2).

Despite the promising developments presented in this review, readers should also be aware of the limitations of machine learning-based methods for protein allostery study. In general, the use of a machine-learning model is restricted by the training data source and model construction. Machine learning-based models may not lead to a universal theory to explain general allosteric events.

Nevertheless, given the success in numerous studies of protein allostery using machine-learning methods, we expect to see the current trend to continue with more applications using machine-learning methods suitable for protein systems, especially dynamical processes. Due to the uniqueness of protein systems, there is a need to develop machine-learning methods for different purposes, including dimensionality reduction methods with accurate decoding functionality, time-dependent data series analysis, and features suitable for chemical structures, protein structures, and protein assembly structures. With more data available and deeper insight into protein allosteric mechanisms, we expect to see systematic development in allosteric protein engineering and even *de novo* allosteric protein design. With the continuous accumulation of more data and information from chemical and biological sciences related to protein allostery, there are increasing opportunities for advanced and specific machine-learning methods to be integrated into this interdisciplinary field (see Outstanding questions).

Box 1. Allosteric mechanisms against the SARS-CoV-2 viral spike protein

Coronavirus disease 2019 (COVID-19), caused by SARS-CoV-2, emerged in late 2019 and then quickly spread around the globe. The infection involves the attachment of the receptor-binding domain (RBD) of the SARS-CoV-2 viral spike (S) protein to **angiotensin-converting enzyme 2 (ACE2)** receptors on the peripheral membrane of host cells [103]. The open and closed conformations of ACE2 differ from each other by the degree of opening of the catalytic site cleft of the peptidase domain (PD). These structural insights identified ACE2 as a viable target to block S1 recognition through allosteric control of open–closed transitions necessary for S1 recognition [104,105].

Extensive studies have revealed that SARS-CoV-2 shares many biological features with, but has higher infectivity than, SARS-CoV [106]. Delgado *et al.* [107] aimed to understand the host receptor recognition mechanism of SARS-CoV-2 to explain this. Affinity propagation algorithm, an unsupervised machine-learning algorithm, was used for clustering analysis of CoV and CoV-2 spike-ACE2 systems. Trozzi *et al.* [108] developed a collective variable-guided (CV)-CNN model as a novel scheme to capture the functional and structural differences of the ACE2 extracellular N-terminal PD. The REDAN model was used to obtain the pathway information of residue–residue interactions that characterize ACE2 PD functional dynamics. Uyar and Dickson [109] distinguished several all-atom MD simulations by linear discriminant analysis (LDA) to show persistent differences in the ACE2 structure upon binding. This allows the prediction of which compounds lead to free versus bound states and to pinpoint long-range ligand-induced allosteric changes in the ACE2 structure. Ray *et al.* [103] focused on the correlations between the RBD and residues in distant, allosteric sites. These computational studies provided insight at the atomistic level into the infection process of SARS-CoV-2 and paved the way for allosteric drug design to treat COVID-19.

Box 2. Allosteric drug development against SARS-CoV-2

During the COVID-19 pandemic, developing drugs based on an allosteric mechanism of recognition between the SARS-CoV-2 spike protein and ACE2 proteins was an important strategy. Iyengar [110] used the machine-learning method partial order optimum likelihood (POOL) to predict allosteric binding sites in protein structures from SARS-CoV-2. Other studies focused on identifying allosteric modulators for either SARS-CoV-2 spike proteins or ACE2 as potential drugs. Karki *et al.* [104] introduced an application of a DNN-based drug screening method, validating it using a docking algorithm against approved drugs for drug-repurposing efforts, and extending the screen to a library of 750 000 compounds. Jain *et al.* [111] built predictive models, using both machine-learning and pharmacophore-based modeling, with screening data from a SARS-CoV-2 cytopathic effect reduction assay. Experimental testing with live virus provided 100 active compounds out of the predicted hits from the screening result of optimized models. The SARS-CoV-2 main protease (Mpro) is required for maturation of the virus and infection of host cells; thus, the key question is how to block its activity. Kaptan *et al.* [112] combined atomistic simulations with the machine-learning methods, Gaussian mixture model (GMM) and partial least squares-based functional mode analysis (PLS-FMA) model, and found that the enzyme regulates its own activity by a collective allosteric mechanism that involves dimerization and binding of a single substrate. Their results suggest that dimerization of main proteases is a general mechanism to foster coronavirus proliferation and proposes a strategy that does not depend on the frequently mutating spike proteins at the viral envelope. Verkhivker and coworkers [113–119] performed a series of computational studies to explore allosteric mechanisms and potential regulatory effects of SARS-CoV-2 spike proteins for different strains (Figure I). They applied different allosteric models using various machine-learning methods to formulate allosteric interaction pathway and network model for the binding of the SARS-CoV-2 spike proteins. These studies demonstrate the positive impact that advanced and mature computational modeling of protein allostery using machine-learning methods could exert on real global public health emergencies.

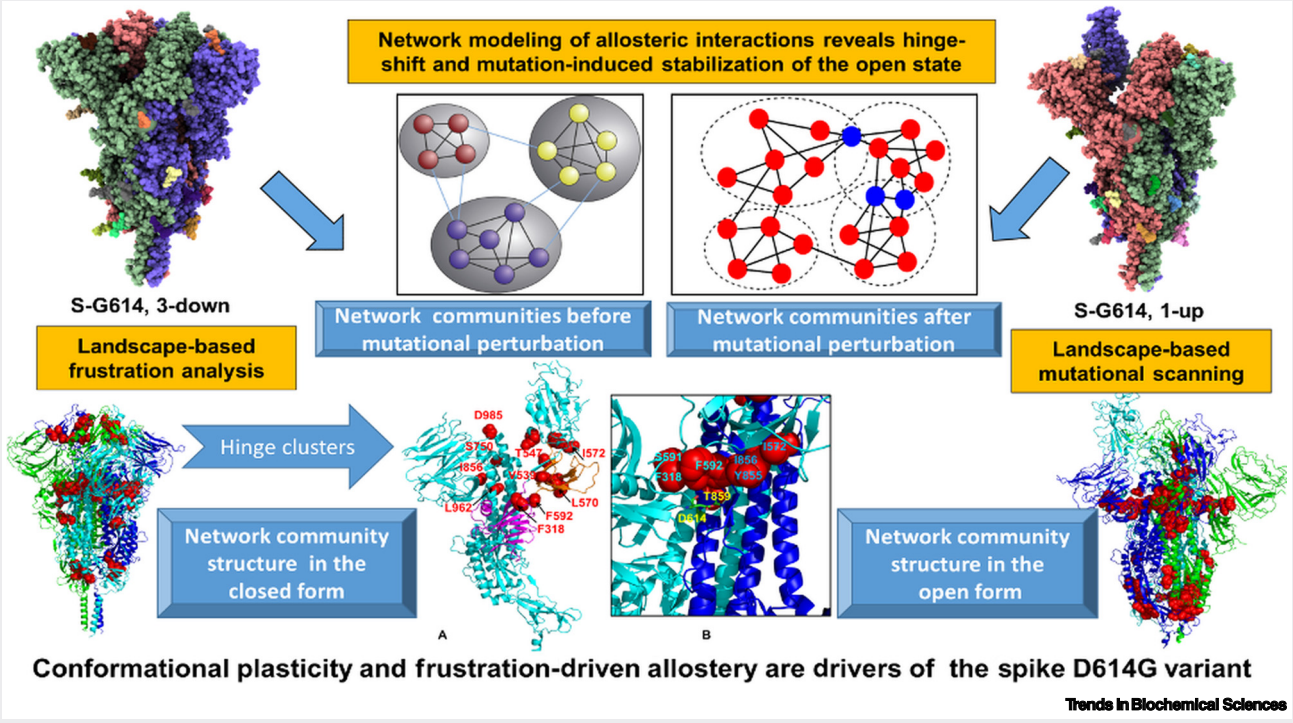

Figure I. Landscape-based protein stability analysis and network modeling of multiple conformational states of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike D614G mutant. Multiple computational methods and models were used in this study of SARS-CoV-2 spike protein allostery focusing on its D614G mutant. Coarse-grained simulations were carried out for trimers of this protein. Residue interaction networks were identified based on both dynamic correlations and coevolutionary residue couplings. A community model was built based on a graph theory representation of protein structure. The impact on protein allostery through mutational perturbation was revealed through both network and community models. The ensemble-based analysis characterized the dynamic signatures of the conformational landscapes for the target protein. The combination of multiple allosteric models revealed a hinge-shift mechanism leading to the increased stability of the open form in the mutant [119].

Acknowledgment

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award No. R15GM122013.

Declaration of interests

There are no conflicts to declare.

Resources

[i]passer.smu.edu

References

1. Zha, J. *et al.* (2022) Explaining and predicting allostery with allosteric database and modern analytical techniques. *J. Mol. Biol.* 434, 167481
2. Lu, S. *et al.* (2019) Allosteric methods and their applications: facilitating the discovery of allosteric drugs and the investigation of allosteric mechanisms. *Acc. Chem. Res.* 52, 492–500
3. Liu, J. and Nussinov, R. (2016) Allostery: an overview of its history, concepts, methods, and applications. *PLoS Comput. Biol.* 12, e1004966
4. Monod, J. and Jacob, F. (1961) General conclusions: teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harb. Symp. Quant. Biol.* 26, 389–401
5. Fenton, A.W. (2008) Allostery: an illustrated definition for the 'second secret of life'. *Trends Biochem. Sci.* 33, 420–425
6. Monod, J. (1971) *Chance and Necessity: An Essay on the Natural Philosophy of Modern Biology*, Vintage Books
7. Motlagh, H.N. *et al.* (2014) The ensemble nature of allostery. *Nature* 508, 331–339
8. Freiburger, L.A. *et al.* (2011) Competing allosteric mechanisms modulate substrate binding in a dimeric enzyme. *Nat. Struct. Mol. Biol.* 18, 288–294
9. Nussinov, R. *et al.* (2013) The (still) underappreciated role of allostery in the cellular network. *Annu. Rev. Biophys.* 42, 169–189
10. Gunasekaran, K. *et al.* (2004) Is allostery an intrinsic property of all dynamic proteins? *Proteins Struct. Funct. Bioinforma.* 57, 433–443
11. Tsai, C.-J. *et al.* (2008) Allostery: absence of a change in shape does not imply that allostery is not at play. *J. Mol. Biol.* 378, 1–11
12. Grutsch, S. *et al.* (2016) NMR methods to study dynamic allostery. *PLoS Comput. Biol.* 12, e1004620
13. Merk, A. *et al.* (2016) Breaking cryo-EM resolution barriers to facilitate drug discovery. *Cell* 165, 1698–1707
14. Coffino, P. and Cheng, Y. (2022) Allostery modulates interactions between proteasome core particles and regulatory particles. *Biomolecules* 12, 764
15. Gulati, S. *et al.* (2019) Cryo-EM structure of phosphodiesterase 6 reveals insights into the allosteric regulation of type I phosphodiesterases. *Sci. Adv.* 5, eaav4322
16. Raman, S. (2018) Systems approaches to understanding and designing allosteric proteins. *Biochemistry* 57, 376–382
17. Yamato, T. and Laprévote, O. (2019) Normal mode analysis and beyond. *Biophys. Physicobiol.* 16, 322–327
18. Hollingsworth, S.A. and Dror, R.O. (2018) Molecular dynamics simulation for all. *Neuron* 99, 1129–1143
19. Na, H. and Song, G. (2014) Bridging between normal mode analysis and elastic network models. *Proteins Struct. Funct. Bioinforma.* 82, 2157–2168
20. Wodak, S.J. *et al.* (2019) Allostery in its many disguises: from theory to applications. *Structure* 27, 566–578
21. Yang, K.K. *et al.* (2019) Machine-learning-guided directed evolution for protein engineering. *Nat. Methods* 16, 687–694
22. Butler, K.T. *et al.* (2018) Machine learning for molecular and materials science. *Nature* 559, 547–555
23. Mater, A.C. and Coote, M.L. (2019) Deep learning in chemistry. *J. Chem. Inf. Model.* 59, 2545–2559
24. Guzel, P. and Kurkcuoglu, O. (2017) Identification of potential allosteric communication pathways between functional sites of the bacterial ribosome by graph and elastic network models. *Biochim. Biophys. Acta BBA Gen. Subj.* 1861, 3131–3141
25. Lu, H.-M. and Liang, J. (2009) Perturbation-based Markovian transmission model for probing allosteric dynamics of large macromolecular assembling: a study of GroEL-GroES. *PLoS Comput. Biol.* 5, e1000526
26. Verkhivker, G.M. (2020) Molecular simulations and network modeling reveal an allosteric signaling in the SARS-CoV-2 spike proteins. *J. Proteome Res.* 19, 4587–4608
27. Villani, G. (2020) A time-dependent quantum approach to allostery and a comparison with light-harvesting in photosynthetic phenomenon. *Front. Mol. Biosci.* 7, 156
28. Hakhverdyan, Z. *et al.* (2021) Dissecting the structural dynamics of the nuclear pore complex. *Mol. Cell* 81, 153–165
29. Brotzakis, Z.F. *et al.* (2021) A method of incorporating rate constants as kinetic constraints in molecular dynamics simulations. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2012423118
30. Matsunaga, Y. and Sugita, Y. (2018) Linking time-series of single-molecule experiments with molecular dynamics simulations by machine learning. *eLife* 7, e32668
31. Matsunaga, Y. and Sugita, Y. (2020) Use of single-molecule time-series data for refining conformational dynamics in molecular simulations. *Curr. Opin. Struct. Biol.* 61, 153–159
32. Zheng, Y. and Cui, Q. (2018) Multiple pathways and time scales for conformational transitions in apo-adenylate kinase. *J. Chem. Theory Comput.* 14, 1716–1726
33. Thomsen, J. *et al.* (2020) DeepFRET, a software for rapid and automated single-molecule FRET data classification using deep learning. *eLife* 9, e60404
34. Wehmeyer, C. and Noé, F. (2018) Time-lagged autoencoders: deep learning of slow collective variables for molecular kinetics. *J. Chem. Phys.* 148, 241703
35. Hernández, C.X. *et al.* (2018) Variational encoding of complex dynamics. *Phys. Rev. E* 97, 062412
36. Mardt, A. *et al.* (2018) VAMPnets for deep learning of molecular kinetics. *Nat. Commun.* 9, 5
37. Ribeiro, J.M.L. *et al.* (2018) Reweighted autoencoded variational Bayes for enhanced sampling (RAVE). *J. Chem. Phys.* 149, 072301
38. Sultan, M. and Pande, V.S. (2017) tICA-Metadynamics: accelerating metadynamics by using kinetically selected collective variables. *J. Chem. Theory Comput.* 13, 2440–2447
39. Doerr, S. and De Fabritiis, G. (2014) On-the-fly learning and sampling of ligand binding by high-throughput molecular simulations. *J. Chem. Theory Comput.* 10, 2064–2069
40. Zimmerman, M.I. and Bowman, G.R. (2015) FAST conformational searches by balancing exploration/exploitation tradeoffs. *J. Chem. Theory Comput.* 11, 5747–5757
41. Noé, F. *et al.* (2019) Boltzmann generators: sampling equilibrium states of many-body systems with deep learning. *Science* 365, eaaw1147
42. Takami, K. *et al.* (2020) Performance research of clustering methods for detecting state transition trajectories in hemoglobin. *J. Comput. Chem. Jpn.* 19, 154–157
43. Konovalov, K.A. *et al.* (2021) Markov state models to study the functional dynamics of proteins in the wake of machine learning. *JACS Au* 1, 1330–1341
44. Bonati, L. *et al.* (2021) Deep learning the slow modes for rare events sampling. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2113533118
45. Brandt, S. *et al.* (2018) Machine learning of biomolecular reaction coordinates. *J. Phys. Chem. Lett.* 9, 2144–2150
46. Zhou, H. *et al.* (2018) Recognition of protein allosteric states and residues: machine learning approaches. *J. Comput. Chem.* 39, 1481–1490
47. Hayatshahi, H.S. *et al.* (2019) Probing protein allostery as a residue-specific concept via residue response maps. *J. Chem. Inf. Model.* 59, 4691–4705
48. Do, H.N. *et al.* (2022) GLOW: a workflow integrating gaussian-accelerated molecular dynamics and deep learning for free energy profiling. *J. Chem. Theory Comput.* 18, 1423–1436
49. Kityk, R. *et al.* (2015) Pathways of allosteric regulation in Hsp70 chaperones. *Nat. Commun.* 6, 8308
50. Mashaghi, A. *et al.* (2016) Alternative modes of client binding enable functional plasticity of Hsp70. *Nature* 539, 448–451
51. Kityk, R. *et al.* (2018) Molecular mechanism of J-domain-triggered ATP hydrolysis by Hsp70 chaperones. *Mol. Cell* 69, 227–237

52. Zhu, J. *et al.* (2022) Neural relational inference to learn long-range allosteric interactions in proteins from molecular dynamics simulations. *Nat. Commun.* 13, 1661
53. Arantes, P.R. *et al.* (2022) Emerging methods and applications to decrypt allostery in proteins and nucleic acids. *J. Mol. Biol.* 434, 167518
54. Zhou, H. and Tao, P. (2019) REDAN: relative entropy-based dynamical allosteric network model. *Mol. Phys.* 117, 1334–1343
55. Botlani, M. *et al.* (2018) Machine learning approaches to evaluate correlation patterns in allosteric signaling: a case study of the PDZ2 domain. *J. Chem. Phys.* 148, 241726
56. Yan, W. *et al.* (2018) Node-weighted amino acid network strategy for characterization and identification of protein functional residues. *J. Chem. Inf. Model.* 58, 2024–2032
57. Zhou, H. *et al.* (2019) Allosteric mechanism of the circadian protein Vivid resolved through Markov state model and machine learning analysis. *PLoS Comput. Biol.* 15, e1006801
58. Ibrahim, M.T. *et al.* (2022) Dynamics of hydrogen bonds in the secondary structures of allosteric protein *Avena Sativa* phototropin 1. *Comput. Struct. Biotechnol. J.* 20, 50–64
59. Stetz, G. and Verkhivker, G.M. (2017) Computational analysis of residue interaction networks and coevolutionary relationships in the Hsp70 chaperones: a community-hopping model of allosteric regulation and communication. *PLoS Comput. Biol.* 13, e1005299
60. Astl, L. *et al.* (2020) Allosteric mechanism of the Hsp90 chaperone interactions with cochaperones and client proteins by modulating communication spines of coupled regulatory switches: integrative atomistic modeling of Hsp90 signaling in dynamic interaction networks. *J. Chem. Inf. Model.* 60, 3616–3631
61. Stetz, G. *et al.* (2020) Exploring mechanisms of communication switching in the Hsp90-Cdc37 regulatory complexes with client kinases through allosteric coupling of phosphorylation sites: perturbation-based modeling and hierarchical community analysis of residue interaction networks. *J. Chem. Theory Comput.* 16, 4706–4725
62. Astl, L. *et al.* (2020) Dissecting molecular principles of the hsp90 chaperone regulation by allosteric modulators using a hierarchical simulation approach and network modeling of allosteric interactions: conformational selection dictates the diversity of protein responses and ligand-specific functional mechanisms. *J. Chem. Theory Comput.* 16, 6656–6677
63. Verkhivker, G.M. (2022) 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.* 434, 167506
64. Chen, Y. *et al.* (2021) Allosteric effect of nanobody binding on ligand-specific active states of the β2 adrenergic receptor. *J. Chem. Inf. Model.* 61, 6024–6037
65. Liu, X. *et al.* (2020) Unraveling allosteric landscapes of allosterome with ASD. *Nucleic Acids Res.* 48, D394–D401
66. Huang, W. *et al.* (2015) ASBench: benchmarking sets for allosteric discovery. *Bioinformatics* 31, 2598–2600
67. Greener, J.G. *et al.* (2017) Predicting protein dynamics and allostery using multi-protein atomic distance constraints. *Structure* 25, 546–558
68. Xie, J. *et al.* (2022) Uncovering the dominant motion modes of allosteric regulation improves allosteric site prediction. *J. Chem. Inf. Model.* 62, 187–195
69. Lu, S. *et al.* (2014) Recent computational advances in the identification of allosteric sites in proteins. *Drug Discov. Today* 19, 1595–1600
70. Akbar, R. and Helms, V. (2018) ALLO: a tool to discriminate and prioritize allosteric pockets. *Chem. Biol. Drug Des.* 91, 845–853
71. Tian, H. *et al.* (2021) PASSer: prediction of allosteric sites server. *Mach. Learn. Sci. Technol.* 2, 035015
72. Xiao, S. *et al.* (2022) PASSer2.0: accurate prediction of protein allosteric sites through automated machine learning. *Front. Mol. Biosci.* 9, 879251
73. Chen, A.S.-Y. *et al.* (2016) A random forest model for predicting allosteric and functional sites on proteins. *Mol. Inform.* 35, 125–135
74. Huang, W. *et al.* (2013) Allosite: a method for predicting allosteric sites. *Bioinformatics* 29, 2357–2359
75. Greener, J.G. and Sternberg, M.J. (2015) AlloPred: prediction of allosteric pockets on proteins using normal mode perturbation analysis. *BMC Bioinforma.* 16, 335
76. Song, K. *et al.* (2017) Improved method for the identification and validation of allosteric sites. *J. Chem. Inf. Model.* 57, 2358–2363
77. Mishra, S.K. *et al.* (2019) Coupling dynamics and evolutionary information with structure to identify protein regulatory and functional binding sites. *Proteins* 87, 850–868
78. Fogha, J. *et al.* (2020) Computational analysis of crystallization additives for the identification of new allosteric sites. *ACS Omega* 5, 2114–2122
79. Tsai, C.-J. and Nussinov, R. (2014) A unified view of "how allostery works". *PLoS Comput. Biol.* 10, e1003394
80. Zhang, W. *et al.* (2019) Correlation between allosteric and orthosteric sites. In *Protein Allostery in Drug Discovery* (Zhang, J. and Nussinov, R., eds), pp. 89–105, Springer
81. Leroux, A.E. and Biondi, R.M. (2020) Renaissance of allostery to disrupt protein kinase interactions. *Trends Biochem. Sci.* 45, 27–41
82. Tee, W.-V. *et al.* (2018) Reversing allosteric communication: from detecting allosteric sites to inducing and tuning targeted allosteric response. *PLoS Comput. Biol.* 14, e1006228
83. Fan, J. *et al.* (2021) Harnessing reversed allosteric communication: a novel strategy for allosteric drug discovery. *J. Med. Chem.* 64, 17728–17743
84. Ni, D. *et al.* (2021) Discovery of cryptic allosteric sites using reversed allosteric communication by a combined computational and experimental strategy. *Chem. Sci.* 12, 464–476
85. Ferraro, M. *et al.* (2021) Machine learning of allosteric effects: the analysis of ligand-induced dynamics to predict functional effects in TRAP1. *J. Phys. Chem. B* 125, 101–114
86. Leander, M. *et al.* (2022) Deep mutational scanning and machine learning reveal structural and molecular rules governing allosteric hotspots in homologous proteins. *eLife* 11, e79932
87. Faure, A.J. *et al.* (2022) Mapping the energetic and allosteric landscapes of protein binding domains. *Nature* 604, 175–183
88. Wang, Q. *et al.* (2012) Toward understanding the molecular basis for chemical allosteric modulator design. *J. Mol. Graph. Model.* 38, 324–333
89. Smith, R.D. *et al.* (2017) Are there physicochemical differences between allosteric and competitive ligands? *PLoS Comput. Biol.* 13, e1005813
90. Hou, T. *et al.* (2021) Integrated multi-class classification and prediction of GPCR allosteric modulators by machine learning intelligence. *Biomolecules* 11, 870
91. Bian, Y. and Xie, X.-Q. (2022) Artificial intelligent deep learning molecular generative modeling of scaffold-focused and cannabinoid CB2 target-specific small-molecule sublibraries. *Cells* 11, 915
92. Yang, L. *et al.* (2021) Transformer-based generative model accelerating the development of novel BRAF inhibitors. *ACS Omega* 6, 33864–33873
93. Vennila, K.N. and Elango, K.P. (2022) Multimodal generative neural networks and molecular dynamics based identification of PDK1 PIF-pocket modulators. *Mol. Syst. Des. Eng.* 7, 1085–1092
94. Huang, M. *et al.* (2018) AlloFinder: a strategy for allosteric modulator discovery and allosterome analyses. *Nucleic Acids Res.* 46, W451–W458
95. Miljković, F. *et al.* (2020) Machine learning models for accurate prediction of kinase inhibitors with different binding modes. *J. Med. Chem.* 63, 8738–8748
96. Rodrigues, T. *et al.* (2018) Machine intelligence decrypts β-lapachone as an allosteric 5-lipoxygenase inhibitor. *Chem. Sci.* 9, 6899–6903
97. Zayner, J.P. *et al.* (2013) Investigating models of protein function and allostery with a widespread mutational analysis of a light-activated protein. *Biophys. J.* 105, 1027–1036
98. Weinkam, P. *et al.* (2013) Impact of mutations on the allosteric conformational equilibrium. *J. Mol. Biol.* 425, 647–661
99. Xiao, J. *et al.* (2019) Probing light chain mutation effects on thrombin via molecular dynamics simulations and machine learning. *J. Biomol. Struct. Dyn.* 37, 982–999

100. García-Fernández, M.D. *et al.* (2021) Distinct classes of potassium channels fused to GPCRs as electrical signaling biosensors. *Cell Rep. Methods* 1, 100119
101. D'Amico, R.N. *et al.* (2021) Substitution of a surface-exposed residue involved in an allosteric network enhances tryptophan synthase function in cells. *Front. Mol. Biosci.* 8, 679915
102. Astl, L. and Verkhivker, G.M. (2019) Data-driven computational analysis of allosteric proteins by exploring protein dynamics, residue coevolution and residue interaction networks. *Biochim. Biophys. Acta Gen. Subj.* Published online July 19, 2019. https://doi.org/10.1016/j.bbagen.2019.07.008
103. Ray, D. *et al.* (2021) Distant residues modulate conformational opening in SARS-CoV-2 spike protein. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2100943118
104. Karki, N. *et al.* (2021) Predicting potential SARS-COV-2 drugs—in depth drug database screening using deep neural network framework SSnet, classical virtual screening and docking. *Int. J. Mol. Sci.* 22, 1573
105. Bhattarai, A. *et al.* (2021) Mechanism and pathways of inhibitor binding to the human ACE2 receptor for SARS-CoV1/2. *Biophys. J.* 120, 204a
106. Nishiga, M. *et al.* (2020) COVID-19 and cardiovascular disease: from basic mechanisms to clinical perspectives. *Nat. Rev. Cardiol.* 17, 543–558
107. Delgado, J.M. *et al.* (2021) Molecular basis for higher affinity of SARS-CoV-2 spike RBD for human ACE2 receptor. *Proteins Struct. Funct. Bioinforma.* 89, 1134–1144
108. Trozzi, F. *et al.* (2022) Allosteric control of ACE2 peptidase domain dynamics. *Org. Biomol. Chem.* 20, 3605–3618
109. Uyar, A. and Dickson, A. (2021) Perturbation of ACE2 structural ensembles by SARS-CoV-2 spike protein binding. *J. Chem. Theory Comput.* 17, 5896–5906
110. Iyengar, S.M. *et al.* (2021) Prediction and analysis of multiple sites and inhibitors of SARS-CoV-2 proteins. *Biophys. J.* 120, 204a
111. Jain, S. *et al.* (2021) Hybrid In silico approach reveals novel inhibitors of multiple SARS-CoV-2 variants. *ACS Pharmacol. Transl. Sci.* 4, 1675–1688
112. Kaptan, S. *et al.* (2022) Maturation of the SARS-CoV-2 virus is regulated by dimerization of its main protease. *Comput. Struct. Biotechnol. J.* 20, 3336–3346
113. Verkhivker, G.M. *et al.* (2021) Atomistic simulations and in silico mutational profiling of protein stability and binding in the SARS-CoV-2 spike protein complexes with nanobodies: molecular determinants of mutational escape mechanisms. *ACS Omega* 6, 26354–26371
114. Verkhivker, G.M. *et al.* (2021) Computational analysis of protein stability and allosteric interaction networks in distinct conformational forms of the SARS-CoV-2 spike D614G mutant: reconciling functional mechanisms through allosteric model of spike regulation. *J. Biomol. Struct. Dyn.* 40, 9724–9741
115. Verkhivker, G.M. *et al.* (2021) Allosteric control of structural mimicry and mutational escape in the SARS-CoV-2 spike protein complexes with the ACE2 decoys and miniprotein inhibitors: a network-based approach for mutational profiling of binding and signaling. *J. Chem. Inf. Model.* 61, 5172–5191
116. Verkhivker, G.M. and Di Paola, L. (2021) 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. Phys. Chem. B* 125, 850–873
117. Verkhivker, G. (2022) Allosteric determinants of the SARS-CoV-2 spike protein binding with nanobodies: examining mechanisms of mutational escape and sensitivity of the omicron variant. *Int. J. Mol. Sci.* 23, 2172
118. Verkhivker, G. *et al.* (2022) Computer simulations and network-based profiling of binding and allosteric interactions of SARS-CoV-2 spike variant complexes and the host receptor: dissecting the mechanistic effects of the delta and omicron mutations. *Int. J. Mol. Sci.* 23, 4376
119. Verkhivker, G.M. *et al.* (2022) Landscape-based protein stability analysis and network modeling of multiple conformational states of the SARS-CoV-2 spike D614G mutant: conformational plasticity and frustration-induced allostery as energetic drivers of highly transmissible spike variants. *J. Chem. Inf. Model.* 62, 1956–1978