



Review

From Deep Mutational Mapping of Allosteric Protein Landscapes to Deep Learning of Allostery and Hidden Allosteric Sites: Zooming in on “Allosteric Intersection” of Biochemical and Big Data Approaches

Gennady Verkhivker ^{1,2,*}, Mohammed Alshahrani ¹, Grace Gupta ¹, Sian Xiao ³ and Peng Tao ³

¹ Keck Center for Science and Engineering, Graduate Program in Computational and Data Sciences, Schmid College of Science and Technology, Chapman University, Orange, CA 92866, USA; alshahrani@chapman.edu (M.A.); grgupta@chapman.edu (G.G.)

² Department of Biomedical and Pharmaceutical Sciences, Chapman University School of Pharmacy, Irvine, CA 92618, USA

³ Department of Chemistry, Center for Research Computing, Center for Drug Discovery, Design, and Delivery (CD4), Southern Methodist University, Dallas, TX 75275, USA; sxiao@smu.edu (S.X.); ptao@smu.edu (P.T.)

* Correspondence: verkhivk@chapman.edu; Tel.: +1-714-516-4586

Abstract: The recent advances in artificial intelligence (AI) and machine learning have driven the design of new expert systems and automated workflows that are able to model complex chemical and biological phenomena. In recent years, machine learning approaches have been developed and actively deployed to facilitate computational and experimental studies of protein dynamics and allosteric mechanisms. In this review, we discuss in detail new developments along two major directions of allosteric research through the lens of data-intensive biochemical approaches and AI-based computational methods. Despite considerable progress in applications of AI methods for protein structure and dynamics studies, the intersection between allosteric regulation, the emerging structural biology technologies and AI approaches remains largely unexplored, calling for the development of AI-augmented integrative structural biology. In this review, we focus on the latest remarkable progress in deep high-throughput mining and comprehensive mapping of allosteric protein landscapes and allosteric regulatory mechanisms as well as on the new developments in AI methods for prediction and characterization of allosteric binding sites on the proteome level. We also discuss new AI-augmented structural biology approaches that expand our knowledge of the universe of protein dynamics and allostery. We conclude with an outlook and highlight the importance of developing an open science infrastructure for machine learning studies of allosteric regulation and validation of computational approaches using integrative studies of allosteric mechanisms. The development of community-accessible tools that uniquely leverage the existing experimental and simulation knowledgebase to enable interrogation of the allosteric functions can provide a much-needed boost to further innovation and integration of experimental and computational technologies empowered by booming AI field.

Keywords: allosteric regulation; deep mutational scanning; molecular dynamics; protein stability; network modeling; deep machine learning; binding energetics; allosteric communications; allosteric binding pockets



Citation: Verkhivker, G.; Alshahrani, M.; Gupta, G.; Xiao, S.; Tao, P. From Deep Mutational Mapping of Allosteric Protein Landscapes to Deep Learning of Allostery and Hidden Allosteric Sites: Zooming in on “Allosteric Intersection” of Biochemical and Big Data Approaches. *Int. J. Mol. Sci.* **2023**, *24*, 7747. <https://doi.org/10.3390/ijms24097747>

Academic Editor: Samuel De Visser

Received: 6 April 2023

Revised: 22 April 2023

Accepted: 23 April 2023

Published: 24 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Over the past years, artificial intelligence (AI) and machine learning (ML) approaches have increasingly found their way into biomedical research, healthcare applications, and drug discovery. AI augments the capabilities of human researchers, often outperforming alternative methods, and leverages large amounts of high-quality data obtained in

biochemical and biophysical experiments for model training which is a key bottleneck that can unlock the full potential of AI technologies. The advances in the ML field have driven the design of new tools and systems to model increasingly complex chemical and biological phenomena [1–5]. Recent years have witnessed the remarkable growth of data-intensive experimental high-throughput technologies and machine learning tools aiming for atomistic-level interrogation and characterization and engineering of protein functions and mechanisms. Allosteric interactions and communications in proteins are central to diverse biological mechanisms, and allosteric regulation phenomenon has been long recognized as the ‘second secret of life’ [6–14]. Nonetheless, dynamic molecular mechanisms that give rise to allosteric molecular events continue to be fairly elusive and are often characterized at a phenomenological level, lacking rigorous and universal theoretical and computational foundation that can explain the enormous diversity of these processes. Diverse spatial and temporal scales of thermodynamic and dynamic changes underlie allosteric regulatory mechanisms and communications. Among important reasons that limited the fundamental progress in characterization of allosteric mechanisms across human proteome is the lack of precise and robust high-throughput biochemical methods that can comprehensively quantify allosteric effects at the residue level resolution and the sparse data on quantitative mappings of the genotype-phenotype landscapes for allosterically regulated proteins.

Structural and biophysical studies of allosteric molecular events that involve the formation of short-lived allosteric states and intermediates have benefited from advances in nuclear magnetic resonance (NMR) technologies that can be employed as powerful tools for probing of allosteric interactions and communications in proteins [15–18]. NMR studies highlighted progress in the determination of hidden allosteric states [19–26]. AI and ML approaches can capitalize on the growing experimental and simulation data on protein dynamics to enable robust exploration of conformational landscapes for complex biomolecular systems [27–35]. In our recent review, we discussed potentials and limitations of integrative computational and experimental studies empowered by AI/ML tools for quantitative characterization of allostery from first principles [36] focusing on applications of ML approaches for efficient conformational sampling and characterization of protein dynamics. In the present review, we examine and discuss in detail new developments along two major directions of allosteric research through the lens of data-intensive biochemical approaches and AI-based computational methods (Figure 1). We focus on the latest remarkable progress in deep high-throughput mining and comprehensive mapping of allosteric protein landscapes and allosteric regulatory mechanisms. This review also analyzes the latest developments in AI and ML methods for atomistic modeling of allosteric proteins, experiment-guided mapping of allosteric energy landscapes, and new predictive models of allosteric binding sites on the proteome level (Figure 1). We also discuss the recent progress in the AI-augmented structural biology approaches that empower and expand high-throughput capabilities of X-ray crystallography, cryo-electron microscopy (cryo-EM), and single-molecule Förster resonance energy transfer (smFRET) technologies to expand our knowledge of the universe of protein dynamics and allostery. We conclude with an outlook and highlight the importance of developing an open science infrastructure for machine learning studies of allosteric regulation and validation of computational approaches using integrative studies of allosteric mechanisms.

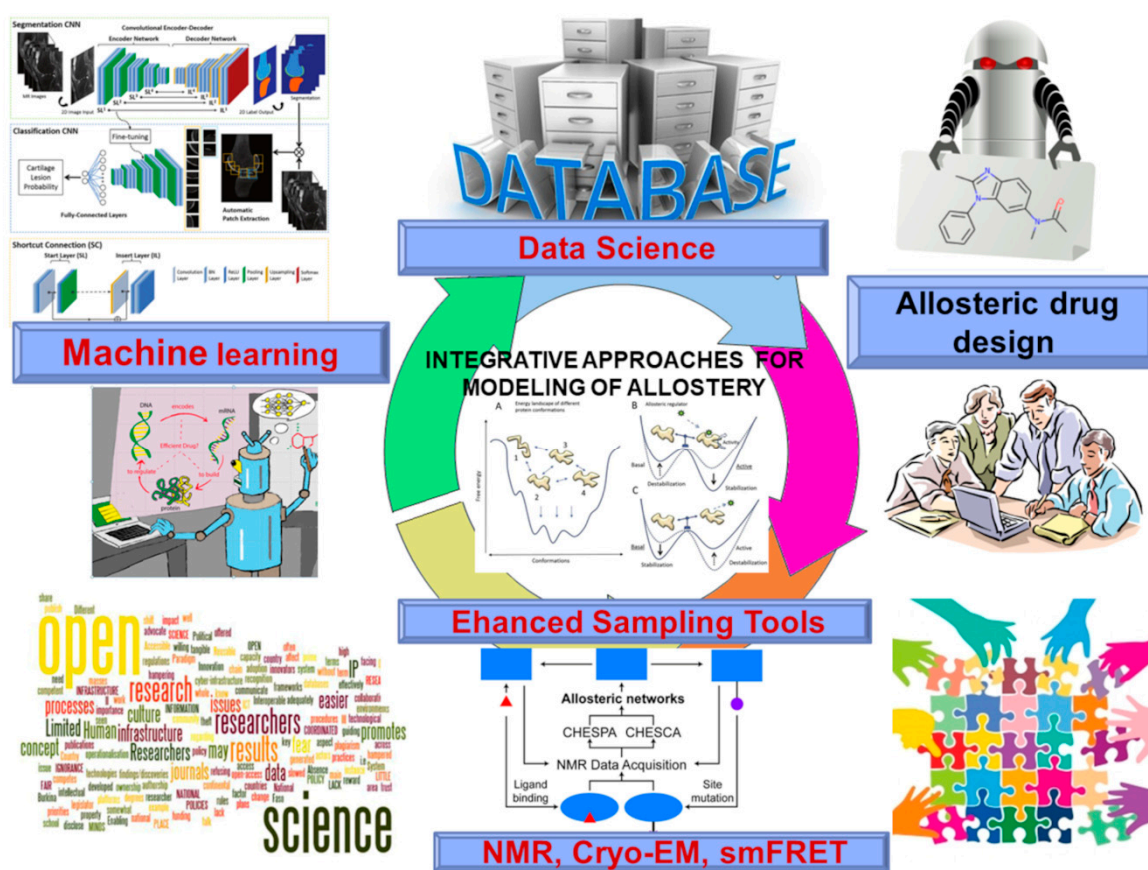


Figure 1. The integrative computational and experimental approaches for studies of conformational dynamics and allosteric regulation mechanisms. (A) The energy landscape of different allosteric conformations. (B,C) Different effects of allosteric regulators on modulating the thermodynamic equilibrium between allosteric conformations.

2. Mining of Allosteric Protein Landscapes and Mechanisms Using Multidimensional Deep Mutational Scanning Approaches

Recent biochemical studies exploited and further advanced the Deep Mutational Scanning (DMS) approach to enable direct mapping of residue-based biophysical effects of mutations on protein stability, binding, and expression (Figure 2) [37–49]. Biochemical and functional studies using protein engineering and deep mutagenesis have quantified binding mechanisms of SARS-CoV-2 interactions with the host receptor ACE2 and a wide range of antibodies [38–43]. The landscapes of antibody binding for SARS-CoV-2 spike protein variants were systematically mapped by measuring the binding affinity of all possible combinations of these 15 receptor binding domains (RBD) mutations to a panel of monoclonal antibodies with distinct epitopes [42]. A powerful DMS platform based on non-replicative pseudotyped lentiviruses quantified how SARS-CoV-2 spike mutations impact pseudo-virus infection by providing a high-throughput approach to measure the effect of $\sim 10^5$ combinations of mutations on antibody neutralization [43].

A pioneering approach for global mapping of the energetic and allosteric landscapes of protein interaction domains leverages the massively parallel nature of DMS and employs multidimensional mutagenesis to causally quantify the phenotypic effects of thousands of perturbations for multiple molecular phenotypes in diverse genetic backgrounds [44]. The new technique known as Double Deep Protein-Fragment Complementation Assay (ddPCA) employs two separate selection assays based on the protein-fragment complementation (PCA) assay so that one assay can detect mutational effects on the stability of the protein (AbundancePCA) while the other assay detects both stability and protein–protein interactions (BindingPCA). Compared to other experimental approaches, ddPCA can disentangle

the stability and binding affinity effects of mutations and systematically quantify allosteric sites that are distant from the binding interface in which mutations induce significant changes on the binding via long-range [44]. These illuminating studies have provided invaluable insights into the underlying global features of allosteric landscapes that are driven by evolutionary requirements for multiple phenotypic fitness factors. Among common “rules” of allostery emerged from this study was the realization that allosteric mutations and allosteric sites are common in protein domains and the frequency of allosterically influential positions can increase near the protein binding interfaces, suggesting that local allosteric perturbations may present an important mechanism of allosteric modulation of binding [44]. Remarkably, this study also discovered the abundance of mutations in both the protein core and protein surfaces that can alter stability and binding affinity, which signified presence of a large genotypic space available for evolution and modulation of diverse allosteric regulatory mechanisms.

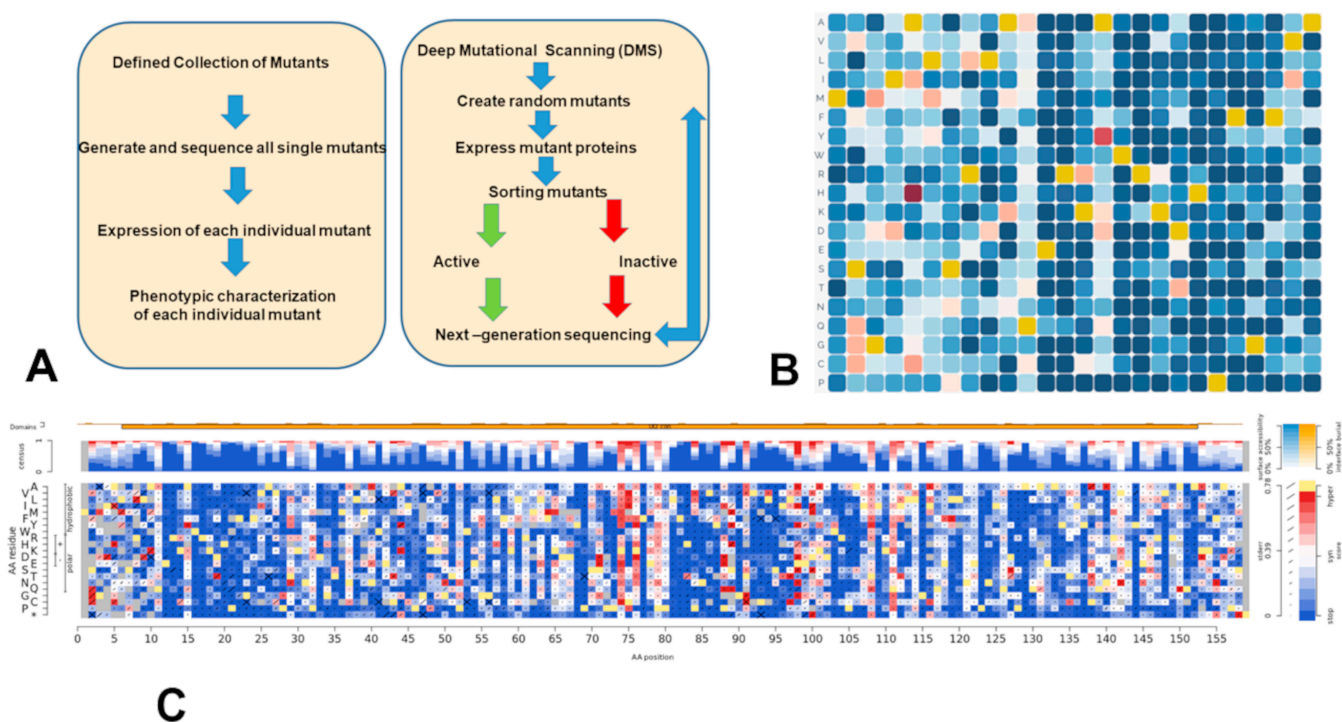


Figure 2. (A) Mutational profiling approach is typically characterized by a defined collection of single mutants. In the DMS approach, the generation and expression of random mutants are followed by high-throughput sorting of active mutants from inactive mutants using flow cytometry, phage display, or growth selection. (B) An example of visualization of deep mutational scanning data from MaveDB (<https://www.mavedb.org/>, accessed on 10 April 2023). MaveDB is an open public repository for datasets from Multiplexed Assays of Variant Effect (MAVEs), such as those generated by DMS [46]. Each cell in the heatmap is a different genetic variant measured in the experiment, where the color corresponds to the functional impact of the variant. (C) An example of DMS visualization for the human SUMO E2 conjugase UBE2I using functional complementation in yeast via DMS-TileSeq obtained from MaveDB [46]. DMS-TileSeq read counts were used to establish relative allele frequencies in each condition. Non-mutagenized control counts were subtracted from counts (as estimates of sequencing error). log ratios of selection over non-selection counts were calculated. The resulting TileSeq fitness values were normalized to a 0–1 scale, where 0 corresponds to the median nonsense variant score and 1 corresponds to the median synonymous score [47].

In a very recent follow-up study, Lehner and coworkers developed a comprehensive map of allostery and reported multiple global atlases of the inhibitory allosteric communications in KRAS protein by measuring the impact of mutations on the protein folding

and binding to six interaction partners [45]. By using the developed enhanced DMS technologies, this study systematically quantified and disentangled mutational effects of protein residues in different genetic backgrounds which allowed to infer biophysical effects of >26,500 variants of KRAS. In addition, this approach reconstructed the free energy landscapes of KRAS binding with six interaction partners by producing >22,000 precise thermodynamic measurements of the mutation-induced free energy changes—a truly monumental achievement and a revolutionary step towards a comprehensive atomistic mapping of genotype-phenotype landscapes on a large scale [45]. In particular, this study quantified the inhibitory allosteric landscape for RAF1 binding by discovering a total of 361 allosteric mutations in 74 distally located from the binding interface allosteric positions, showing that allosteric mutations are highly enriched in the nucleotide binding pocket site of KRAS. This remarkable study revealed principles of allosteric communication in KRAS which may be generalizable to other signaling proteins, showing that allosteric mutations can inhibit binding to all three KRAS effectors and allosterically control binding specificity indicative of fundamental role of allosteric sites in modulation and fine-tuning of binding, regulation, and signaling [45].

DMS studies of the FabZ, LpxC, and MurA proteins that are involved in cell envelope synthesis were performed with the aid of high-throughput CRISPR genome editing, revealing the effect of >17,000 protein variants on function [48]. Mutational libraries generated in this study were deployed to characterize resistance development against antimicrobial compounds that target the selected protein, dissecting the role of specific amino acids in supporting viability. Nonetheless, this powerful approach displayed certain limitations by assessing the effects of single mutations while synergistic and compensatory effects of multiple mutations arising during evolution or due to protein responses eliciting drug resistance remained elusive [48]. The importance of nonadditive, epistatic relationships was assessed in the systematic biochemical analysis of all mutants at the binding interfaces of the SARS-CoV-2 spike variants [49]. A systematic DMS-based mapping of the epistatic interactions between the SARS-CoV-2 Spike BA.1 variant mutations relative to the Wu-Hu-1 original strain showed evidence of compensatory epistasis in which immune escape mutations are compensated by the emergence of affinity-enhancing mutations with the host cell receptor ACE2 [50]. While mutations in proteins are typically deleterious and can impair to a different degree protein function, and stability, compensatory neutral mutations often facilitate protein evolution through positive epistasis with deleterious mutations in the same protein. By enabling high-throughput quantitative analysis of functional effects for thousands of mutations, DMS tools can facilitate understanding of complex synergistic mechanisms and dissect functional cross-talks between mutational sites in addition to quantifying the impact of single mutations. This venue presents a new frontier for further development of DMS technologies, a challenge that requires integration of the DMS datasets [47] with emerging ML models of epistasis and allostery [51].

A high-throughput platform was developed for DMS analysis of caspase-3 (CASP3) and caspase-7 (CASP7) allowing for a large-scale assessment of mutations on caspase structure and function and revealing differences between CASP3 and CASP7 proteolytic activity and protein stability [52]. This study highlighted important limitations of “generic” DMS studies in differentiating between specific molecular mechanisms as mutations that decrease caspase activity could impact functional changes across multiple phenotypes, including protein expression, stability, catalytic activity, substrate specificity, and allosteric regulation. Strikingly, the results of this study further underscored the unique nature of Lehner’s studies [45] that brilliantly overcame these inherent biophysical uncertainties in their ddPCA-based DMS assays and systematically quantified specific functional consequences of every mutation in the system resolving biophysical ambiguities that are inherently present in standard DMS experiments.

Allosteric mapping of GTPase switch Gsp1/Ran system in the context of its in vivo protein interactions by comprehensive mutational scanning enabled characterization of allosteric sites for highly regulated molecular switches that catalyze GTP hydrolysis and

nucleotide exchange [53]. Using a DMS experimental platform, this study discovered new regulatory sites that are allosterically coupled with the GTPase active site allowing allosteric mutations to alter Gsp1/Ran cellular function through modulation of GTPase switching. A large-scale mapping of the genotype-phenotype landscape for the lac repressor, LacI quantified the effect of mutations on LacI allostery, showing that a complex balance between contributions from many residues and interactions can ultimately drive the phenotypic outcomes [54]. Consistent with other studies, it was found that allosteric genotype-phenotype landscapes can allow for rapid evolutionary innovation resulting in novel allosteric phenotypes emerging from sampling a broader genotype space with higher-order mutations [54]. The range and experimental details of DMS methods with a specific focus on recently developed techniques in mutation library generation, high-throughput methods, and data analysis were discussed in a recent insightful review [55].

3. Machine Learning Models Infer Genotype-Phenotype Relationships from DMS Experiments

ML models leveraged the data-rich DMS results to generate predictive models and obtain mechanistic insights into genotype-phenotype maps [56–61]. A positive-unlabeled (PU) learning framework can infer sequence–function relationships from large-scale DMS data showing strong predictive performance across ten large-scale sequence–function datasets, representing proteins of different folds, functions, and library types [56]. DMS experiments were combined with molecular dynamics (MD) simulations and network analysis to examine the functional landscape of a bacterial transcription factor showing the role of diverse and broad ensembles of mutational communication pathways in propagating allosteric phenotypic effects [57]. Deep learning (DL) models built upon DMS data for several bacterial allosteric transcription factors predicted the distribution of allosteric hotspots revealing that regulatory sites mediating allostery can be broadly distributed on the protein rather than being confined to narrow pathways linking the allosteric and active sites [58]. An important revelation of this study was the ability of DL models to accurately predict functional sites and link these predictions with molecular mechanisms of allosteric regulation. A supervised DL architecture helped to infer sequence–function relationships using training on large-scale DMS datasets [59]. Using different network architectures ranging from linear regression to convolutional networks, this study displayed a predictive accuracy on five diverse DMS datasets and enabled the design of a protein sequence that binds to immunoglobulin G with substantially enhanced affinity compared to the wild-type GB1 [59]. A statistical learning framework was developed to infer the protein fitness landscape for DHFR protein from directed evolution data in which the inferred landscape model revealed numerous examples of epistasis arising from interactions between residues and allow for analysis of evolutionary trajectories for protein engineering [60]. Recent studies showed that ML models can successfully reconstruct sequence–function landscapes from large datasets and facilitate efficient protein engineering with less experimental screening than traditional directed evolution approaches [61]. Understanding the relationship between promoter sequence, expression phenotype, and fitness are central in understanding evolution and gene regulation. The genomic analysis of sequence-to-expression regulatory fitness landscapes leveraged data on millions of DNA sequences and their experimentally measured expression to develop ML models and transformer encoders that represented each sequence using a sorted vector of expression changes, allowing for large scale exploration of novel genotypes and predictions of expression levels [62].

Multiplex assays of variant effect (MAVEs) are a type of DMS experiment that utilize a combination of reporter assays, protein expression assays, and next-generation sequencing (NGS) to screen thousands of variants in a single experiment and can be used to study the effect of genetic variants on gene expression, protein stability, and protein–protein interactions. Quantitative modeling of MAVE data is based on the use of latent phenotype models, which describe the relationship between the observed data and a latent phenotype. The framework is designed to capture the complex structure of MAVE data, including both

the observed phenotype data and the underlying latent phenotype. The latent phenotype model is then used to estimate the latent phenotype from the observed data, and to quantify the effect of different environmental factors on the underlying latent phenotype [63]. Using this framework, a Python package MAVE-NN implements a range of ML models for learning genotype-phenotype maps using latent phenotype neural networks and parameters inferred from MAVE data using a TensorFlow 2 backend [63]. With the improving accuracy of DMS datasets and the enhanced predictive power and robustness of generative ML models, the integrative tools combining these approaches can discover functional protein sequences with the designed fitness profile and guide exploration towards effective protein engineering solutions.

4. Learning Protein Conformational Dynamics Using AI Approaches

Exploring and learning the protein allostery universe becomes increasingly plausible using the proliferation of AI/ML tools and methods to solve long-standing problems in molecular biology and biophysics. The remarkable success of AlphaFold2 (AF2) in protein structure modeling and looming resolution of the protein folding problem in the last few years have marked a revolutionary change in structural biology that was inconceivable even several years ago [64,65]. However, the key shortcomings of the AF2 technology in resolving conformational dynamics and allosteric phenomenon challenges are real and are mainly associated with the limitations to accurately capture the ensembles and energetics of functional conformational changes as well as mutational effects on protein dynamics and functions [66]. In recent illuminating reviews, Nussinov and colleagues noted that expanding AF2 tools for prediction of allosteric regulatory mechanisms would require a considerable step forward beyond an accurate single structure prediction—from mapping of the conformational ensembles to detection of low-populated allosteric conformations and hidden functional states [67,68]. Several recent studies have provided strong evidence that structure prediction capabilities of the AF2 methods are not trivially expandable for the prediction of complex allosteric landscapes and mechanisms of allosteric regulation [69,70]. These formidable challenges will test the limits and further accelerate the developments of AI technologies that are currently incorporating new concepts and architectures—from natural language processing to generative transfer learning to enable unprecedented in its scale prediction and design of protein structure, functions, and mechanisms. The latest breakthrough in AI-based applications to large-scale protein structure predictions involved language models of protein sequences at the scale of evolution trained on up to 15 billion parameters that resulted in high accuracy end-to-end family of transformer protein language models, ESMFold for atomic level structure prediction directly from the individual sequence of a protein [71]. Compared to AF2, the trained language model directly learns at scale evolutionary patterns of sequences linked to structure, strikingly eliminating the need for the evolutionary databases, multiple sequence alignments, and templates. This results in a speed improvement for structure prediction of more than an order of magnitude while maintaining high resolution accuracy comparable to AF2 models [71]. A new method for accurate protein design by integrating structure prediction networks and diffusion generative model called RoseTTAFold Diffusion (RFdiffusion) enables the design of complex functional proteins from simple molecular specifications [72]. This method can generate novel protein structures of more than 600 residues that are accurately predicted by AF2 and ESMFold, with the experimental data demonstrating that designs are extremely thermostable [72]. These truly revolutionary developments in the development and application of innovative AI technologies to fundamental challenges in protein science and molecular biology are driven by unified efforts from a myriad of academic laboratories and are accelerated by the growing attention to these problems and competition between industry giants DeepMind/Google and Meta.

The biological importance and complexity of protein conformational dynamics and allosteric mechanisms require next-generation computational approaches and innovative AI-based integrative tools synergizing advances in computation and experiments that

can enable atomistic characterization of allosteric landscapes, states, and interactions. Development of AI-augmented simulation approaches that leverage ML tools to represent physics-based thermodynamic and kinetic determinants of conformational sampling as deep neural networks has gained a significant momentum in recent years. ML approaches were employed to facilitate exploration of conformational landscapes using MD simulations via optimal selection of reaction coordinates. By combining deep learning (DL) approaches and biased MD simulations, physically meaningful collective variables can be determined to resolve the bottlenecks that often hinder the reliable characterization of conformational transitions and rare events [73]. Utilizing variational inference and a predictive information bottleneck concept implemented through deep neural networks, a new framework was recently proposed that utilizes brief MD simulations to estimate the reaction coordinates and to conduct iterative biased simulations, which enables improved exploration of conformational landscapes and reliable prediction of the corresponding thermodynamic and kinetic features of the system [74–77]. The ML model can be trained with simulation data to generate realistic conformational ensembles of proteins without sampling. A recently introduced generative adversarial network, idpGAN, based on a transformer architecture with self-attention on coarse-grained simulations can be used to train an ML model to generate physically realistic conformational ensembles of proteins without needing to sample data and with minimal computational cost [78]. This proof-of-principle model can predict sequence-dependent conformational ensembles for sequences that are not present in the training set and can be further generalized for atomistic ensembles.

VAMPNets, an end-to-end learning system, can be used to predict molecular kinetics with neural networks [79]. The combination of VAMPNet and graph-level dynamics with neural networks creates GraphVAMPNet, which can accurately identify high-resolution metastable states from long-term MD trajectories [80]. GMVAE can be used to construct a reduced representation of the free energy landscape of protein folding, which has clusters that correspond to the metastable states during folding [81]. A method that applies ML models and MD simulations to explore protein conformational space uses an autoencoder to map MD snapshots onto a user-defined conformational landscape defined by principal components analysis or specific structural features [82]. Variational autoencoders (VAE) models were adapted for encoding protein structures into a two-dimensional latent space to describe conformational changes between different allosteric states [83]. The recently introduced latent space-assisted adaptive sampling method for protein trajectories (LAST) [84] can accelerate the exploration of protein conformational space by using VAE and exploitation of the iteratively optimized training data for the latent-based sampling (Figure 3). Using enhanced simulation schemes and machine learning models, researchers have studied the molecular factors that lead to allosteric changes and ligand-induced ensemble changes in proteins. Linear Discriminant Analysis was used to identify differences between the apo and allosteric inhibitor-bound ensembles [85], and another ML was developed to understand the temporal relationships of allosteric stimulation in hemagglutinin-neuraminidase [86]. Zhou and colleagues employed an allosteric community model based on a machine learning method to examine the allosteric mechanism of the Vivid (VVD) protein [87]. An autoencoder-based detection method for characterizing ligand-induced dynamic allostery was used to compare the time fluctuations of protein structures in the form of distance matrices acquired from MD simulations [88].

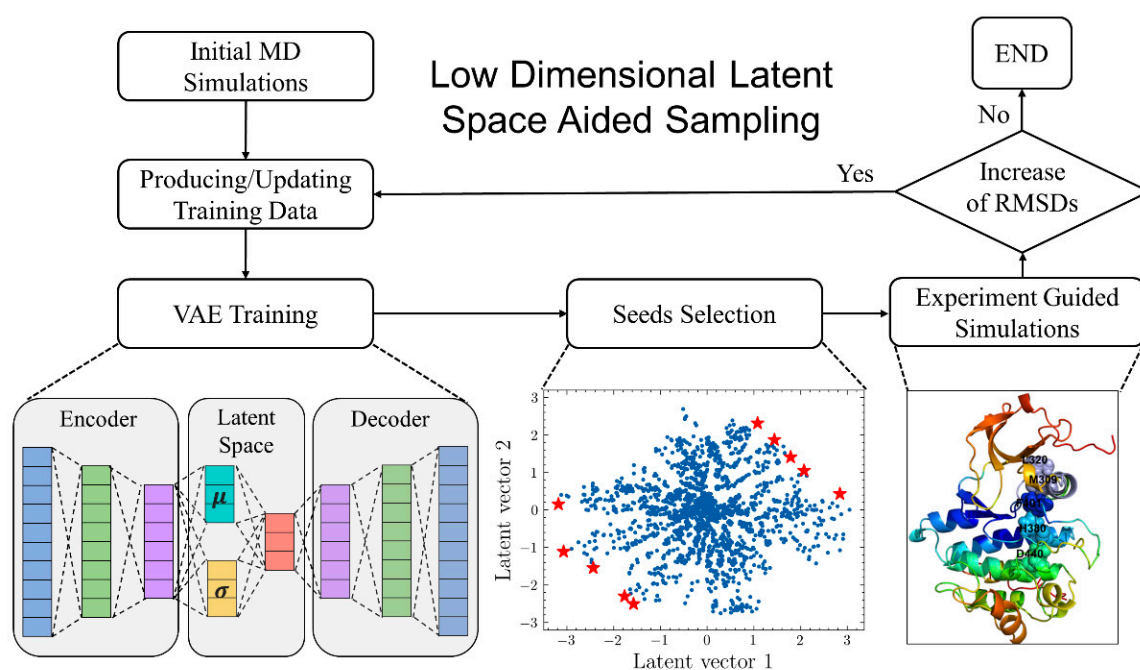


Figure 3. The architecture of latent space-assisted adaptive sampling method for protein trajectories LAST [84] includes cycles of VAE training, seeding structure selection on the latent space, and conformational sampling through iterative cycles of short MD simulations. The proposed approach is validated using simulation studies of metastable states of adenosine kinase (ADK) and native states of Vivid protein (VVD). The red stars refer to different protein conformations in the latent space of the model.

5. From AI-Enabled Mapping of Allosteric Conformational Landscapes to ML Discovery of Cryptic Allosteric Sites

The nature and atomistic details of the allosteric communication between the allosteric site and the functional site are often difficult to dissect. While experimental approaches could reveal allosteric hotspots and potential communication pathways in protein structures, robust and large scale-based tools for mapping of allosteric communications and allosteric binding sites are still missing. The problem of identifying putative allosteric sites can be addressed by considering the communication with orthosteric sites and uncovering the allosteric communication pathways between orthosteric and allosteric sites. Computational methods for predicting binding sites and allosteric pockets can be classified as geometric, energetic, evolution-based, knowledge-based, and machine learning approaches. AlloPred uses normal mode perturbations along with pocket descriptors in a machine learning approach to rank allosteric sites, with a good performance of 28 out of 40 cases [89]. The Allosite approach, which uses support vector machine classifiers of topological and physicochemical characteristics of allosteric and non-allosteric sites, has shown promise in predicting allosteric pockets [90]. AllositePro, an extension of Allosite, combines pocket features with perturbation analysis to identify allosteric sites and was able to validate a novel allosteric site in cyclin-dependent kinase 2 (CDK2) [91]. The Allosite and AllositePro tools, along with the Alloscore scoring function and Allosterome database service, have been essential components of allosteric services and applications developed in the past decade for the evaluation of allosteric protein–modulator interactions and evolutionary analysis of query allosteric sites/modulators within the human proteome [92]. The Active and Regulatory site Prediction (AR-Pred) model was developed by utilizing various residue-based features such as amino acid physicochemical properties, rate of residue evolution, and features describing protein geometry and dynamics [93].

Recently, autonomous predictors of binding sites based solely on ML methods have emerged, replacing traditional template-free methods such as Fpocket [94], SiteHound [95],

and MetaPocket 2.0 [96]. ISMBLab-LIG is an ML algorithm that uses 3D probability density maps of interacting atoms as input attributes to predict ligand binding sites (LBSs), which are relatively insensitive to local conformational variations [97]. DeepSite utilizes 3D voxelization of the protein and deep convolutional neural networks (DCNN) for binding site prediction, outperforming other structure-based algorithms on a set of 7622 proteins from the scPDB database [98]. Other ML-based approaches such as FRSite [99] and Kalasanty [100] also use 3D voxelization of the entire protein, with FRSite based on Faster-RCNN and Kalasanty employing a common segmentation architecture U-Net.

In recent years, the concept of end-to-end learning has been gaining attention. This approach involves three main steps: (a) reducing the dimensions of inputs, (b) creating various neural networks based on the available data, and (c) applying backpropagation across the entire architecture to reduce loss and update weights. For example, a neural relational inference model that utilized a graph neural network and an autoencoder architecture was used to explore the latent embedding of allosteric systems, learn about the long-range interactions and communication between distant sites in the ligand-induced allosteric regulation of Pin1, conformational transition of SOD1 protein, and the activation of MEK1 caused by oncogenic mutations [101]. Another machine learning (ML) approach, referred to as DiffNets, was proposed which does not assume whether large structural changes are more important than local changes, rather it utilizes a self-supervised autoencoder to learn features of the conformational ensembles which are pertinent to distinguish between protein systems [102]. DeepSurf introduces a new representation of the 3D protein surface based on local voxel grids located around the surface, as well as a novel residual network, LDS-ResNet, which has been extended in three dimensions to be applicable to volumetric data [103]. This proposed method has been assessed in terms of binding site prediction using various benchmark datasets, evidencing its superiority when compared to other state-of-the-art approaches.

A template-free P2Rank approach is among the most efficient and fast available ML tools for prediction of ligand binding sites [104,105]. It combines information from both sequence and structural data to rank potential binding sites based on their likelihood of binding a specific ligand. P2Rank uses support vector machine (SVM), random forests (RF), and artificial neural networks (ANNs) to learn the features that are most important for binding site prediction. This method is based on the prediction of the ligandability of a local chemical environment that is centered on points placed on the protein's solvent-accessible surface [104,105]. P2Rank works by scoring and clustering these points on the protein's surface. The ligandability score of individual points is determined by a machine learning model that is trained on a dataset of known protein–ligand complexes. The P2Rank approach assigns structural, physical-chemical, and evolutionary features to points on a mesh that covers the protein surface and builds a ML model from this representation. Rather than learning to memorize ligand binding sites, P2Rank learns what makes local neighborhoods around the protein surface intrinsically ligandable. This then allows the detection of ligandable points, which are then clustered to produce a list of surface patches corresponding to the predicted binding sites [104,105].

Recent advances in convolutional neural networks (CNNs) have made them very effective in image recognition tasks. They have been widely applied to detect protein binding sites which are represented as 3D-voxels with atomic attributes. A novel, rapid, and accurate DL approach, BiteNet (Binding site neural Network) was developed for large-scale and spatiotemporal identification of protein binding sites. Drawing inspiration from computer vision problems such as object detection in images and video, BiteNet considers protein conformations as three-dimensional images with channels corresponding to the atomic densities, binding sites as the objects to detect on these images, and conformational ensembles of proteins as three-dimensional videos to analyze [106]. In comparison to the more established techniques, BiteNet has the ability to enable an extensive examination of conformational ensembles and has the potential to recognize allosteric binding sites in both soluble and transmembrane protein domains. The method has demonstrated

its efficacy in identifying the most challenging binding site detection problems on three-dimensional structures of pharmacological targets, such as ATP-gated cation channels, epidermal growth factor receptors (EGFRs), and G protein-coupled receptors (GPCRs). Notably, it successfully identified an oligomer-specific allosteric binding site made up of the subunits of the trimeric P2X3 receptor complex, as well as a conformation-specific allosteric binding site of the epidermal growth factor receptor kinase domain [106]. This method was validated on a large Holo4K benchmark dataset of holo protein structures used for evaluation of binding site prediction methods [107] showing high accuracy and speed, outperforming both DeepSite and P2Rank tools. BiteNet offers an alternative to traditional methods of predicting binding sites in holo protein structures, by utilizing large-scale analysis of conformational ensembles to assess protein dynamics and flexibility. The sites of interest which are detected by this method can then be utilized for structure-based drug design.

An ensemble learning method combining eXtreme gradient boosting (XGBoost) and graph convolutional neural networks (GCNNs), and an automated machine learning method (AutoGluon and AutoKeras) was developed to predict plausible allosteric sites and both models were incorporated to Prediction of Allosteric Sites Server (PASSer) (Figure 4) [108,109]. Most ML models are designed to make predictions either as labels or probabilities for all pockets detected in studied proteins, which is still a challenging and time-consuming task. Learning to Rank (LTR), an emerging area, and LTR models provide 'relative' predictions by ranking objects from the most to the least relevant, making detection and ranking of the allosteric sites more tractable. LambdaMART model is implemented for allosteric sites ranking with gradient boosting decision tree (GBDT) and LambdaRank loss function [110]. Compared with other ML models such as XGBoost, SVM, and RF, LambdaMART achieved the highest efficiency and showed a better ability to rank the experimentally known allosteric sites at top positions [110].

Cryptic binding sites are small and hidden pockets in proteins that are difficult to detect using traditional methods such as X-ray crystallography and NMR spectroscopy as they may emerge in meta-stable intermediate states that are typically short-lived and dynamic in nature, often appearing during allosteric conformational changes. Additionally, cryptic pockets may contain non-polar or hydrophobic residues that do not form strong hydrogen bonds with other molecules in the protein, making them difficult to detect by computational methods. Accurate and rapid discovery of potential cryptic pockets in a protein structure is an important and challenging problem and robust approaches capable of uncovering hidden druggable sites that can greatly accelerate the scope of therapeutic interventions and modernize drug discovery of the future.

While the repertoire of available methods for detection of protein binding pockets and ligand binding sites is rich and diverse, the reliable identification of allosteric binding sites and hidden allosteric pockets available in short-lived intermediate states is far more challenging. The emergence of recent AI and ML models to incorporate the elements of protein dynamics and sophisticated CNN and DL architectures may provide a fruitful avenue for future developments. These directions are organically linked with the increasing interest to adaptation of AF2 tools for prediction of conformational ensembles and functional states with previously unknown hidden pockets making them an appealing target for a new generation of drug discovery technologies. An important step forward to address challenges of detecting cryptic binding pockets emerging from conformational ensembles is made in a novel method PocketMiner that employs DL-based architecture to identify pockets in large protein structures that have not been identified through crystallography or other experimental techniques. In contrast to previous methods, PocketMiner utilizes molecular simulations to evaluate if each residue in a protein structure has the ability to rearrange its orientation to participate in a cryptic pocket as part of its thermal fluctuations. The proposed DL model is trained on extensive datasets of structural ensembles derived from simulations that contain examples of pocket opening events. PocketMiner can detect hidden binding sites more than 1000 times faster than existing methods when applied

to single structures from a newly created dataset of 39 experimentally confirmed cryptic pockets [111,112]. PocketMiner was also applied to the human proteome discovering that more than half of proteins, which were assumed to lack pockets based on the available structures, likely contain cryptic pockets [111,112].

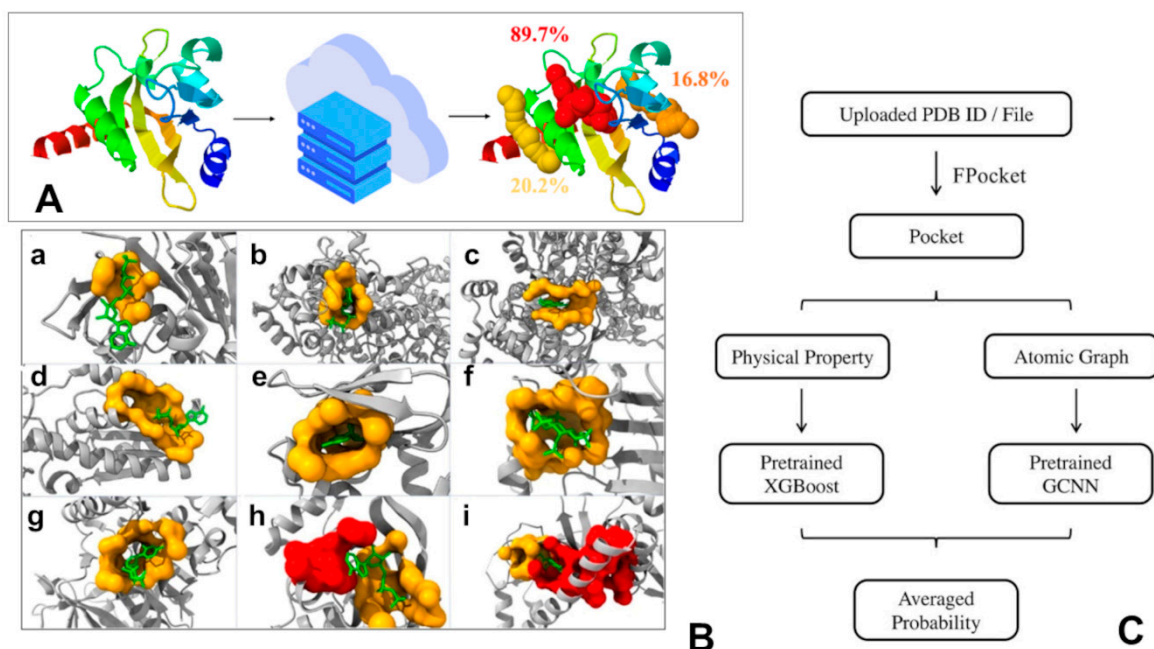


Figure 4. (A) A web server workflow for the Prediction of Allosteric Sites Server (PASSer) approach [108,109] for prediction of allosteric binding sites is presented. Users may submit a PDB ID or a PDB file, and the server will use a pretrained XGBoost model to calculate and predict physical properties for each pocket, and a pretrained GCNN model to construct and process an atomic graph. (B) Structures of nine proteins with modulators and predicted pockets [109]. The following protein structures were analyzed: (a) 2FPL, (b) 2R1R, (c) 3BCR, (d) 4PFK, (e) 1Q5O, (f) 3PEE, (g) 4HO6, (h) 1XMV, and (i) 2OZ6 [109]. The yellow pockets are labeled as allosteric, and the lime molecules are modulators. For (a–g), the allosteric pockets are successfully predicted as top one by our model. For (h,i), the red pockets are predicted as the first place, and the allosteric pockets are predicted as the second place. (C) A schematic flowchart of PASSer algorithm [108,109].

6. AI-Augmented Cryo-EM and smFRET Approaches Expand the Universe of Protein Conformational Dynamics

Recent advances in single-particle cryo-electron microscopy (cryo-EM) have enabled the determination of numerous near-atomic resolution structures for well-ordered proteins and large macromolecular assemblies breaking resolution barriers for studies of allosteric events and allosteric drug discovery [113–115]. The processing of increasingly large, complex datasets using traditional data processing strategies is exceedingly expensive in both user time and computational resources.

The recent development of data processing capitalizes on AI and ML models to enhance the efficacy of data analysis and validation and automate the steps of particle picking, 3D map reconstruction, and local resolution determination [116]. Furthermore, DL algorithms have been used to improve the accuracy of cryo-EM image reconstruction, by providing more accurate estimates of the resolution and signal-to-noise ratio. DL-based image processing approaches have been used in processes of cryo-electron microscopy 3D reconstruction and atomic structure determination [117–120]. CryoDRGN (Deep Reconstructing Generative Networks) is a network that enables heterogeneous reconstruction by learning a deep generative model of 3D structure from diverse cryo-EM density maps. This neural network representation of structure can model single density maps with a higher resolution, and it can be used for unsupervised heterogeneous reconstruction [121].

Cryo-EM reconstructs multiple configurations of the same sample from images, which usually needs several rounds of 2D and 3D classifications to understand the conformational heterogeneity. CryoDRGN can produce a large number of 3D density maps, and automated processing methods for analyzing the results of CryoDRGN have also been created [122]. ResNet is a type of deep learning model that uses a deep neural network architecture composed of residual layers. It has been shown to be effective in a variety of computer vision and natural language processing tasks. The ResNet model uses a set of convolutional layers to learn the representation of the input data, and then a set of residual layers to refine the representation. The residual layers are used to learn the difference between the input and the output of the convolutional layers, making the model more efficient and accurate. The DeepHEMNMA method uses this architecture to learn the distribution of the conformational heterogeneity captured in the single particle images and to use this distribution to generate a single, ensemble-averaged structure [123].

Single-molecule Förster resonance energy transfer (smFRET) paved the way for studying dynamics in macromolecular structures under biologically relevant conditions [124]. The recent breakthroughs in smFRET technologies have enabled dynamic studies of large biomolecules at atomic resolution. MD simulations have been widely applied with smFRET experiments to provide atomistic insights into the dynamic behavior of biomolecules [125,126].

Microsecond MD simulations have been used in combination with ML methods and restraints derived from smFRET experiments to successfully recapitulate and interpret smFRET states and allosteric conformational transitions [127–129]. This integration of single-molecule experiments, MD simulations, and ML models involves supervised learning, in which an initial Markov State Model (MSM) is constructed from raw simulation data, and unsupervised learning, which entails performing hidden Markov modeling to optimize the initial MSM using smFRET time-series measurements [127]. MD simulations can be used in conjunction with information from smFRET experiments to create an automated FRET-assisted structural model that can increase the accuracy of an integrated structure, as well as allow for FRET-assisted coarse-grained structural modeling and MD simulation-based refinement [130]. Recently, ML developments in single-molecule imaging automation and single-molecule feature recognition, including smFRET intensity trace analysis, have been reviewed [131].

The DeepFRET method utilizes DL architectures for rapid and automated classification of smFRET data, which is a significant challenge in smFRET experiments and the analysis of protein dynamics [132]. This approach, trained on simulated smFRET time-series, accurately and effectively distinguishes and categorizes smFRET experimental series both in simulated ground truth data and in a real-world dataset. The AutoSiM approach, which is based on DL, is used for automated screening of smFRET data to identify high-quality traces for further investigation [133].

By leveraging smFRET and MD simulations, researchers have explored the molecular basis and dynamic events associated with allosteric regulatory mechanisms. For example, a study of the NMDA receptor uncovered the mechanism of allostery underlying negative cooperativity, revealing that binding of one agonist leads to flexibility and increased conformational spread at the second agonist site [134]. Additionally, smFRET imaging tools have been used to analyze the conformational dynamics and allosteric modulation of the SARS-CoV-2 Spike protein by antibody binding [135].

By combining smFRET, fluorescence correlation spectroscopy (FCS), and fluorescence lifetime measurements with atomistic MD simulations, researchers were able to show how ATP hydrolysis in Hsp90 chaperones induces allosteric changes at a distant protein binding site and modulates chaperone activities [136]. Furthermore, the effects of positive allosteric modulators on structural dynamics of metabotropic glutamate receptors (mGlu2) were explored in optimized detergent micelles using smFRET at sub-millisecond timescales, demonstrating how the modulators can increase mGlu2 receptor efficacy through fast inter-subunit rearrangements [137].

Despite considerable progress in applications of AI and ML methods in cryo-EM and smFRET studies, the intersection between allosteric regulation, the emerging structural biology technologies, and AI approaches are largely unexplored, calling for the development of AI-augmented integrative structural biology. This is an exciting area of research, with the potential to reveal the underlying mechanisms of allosteric regulation and to develop new ways of controlling the behavior of proteins. By better understanding the mechanisms of allosteric regulation and developing new AI approaches, we can gain a better understanding of how proteins work and how they can be manipulated to produce desired biological outcomes. In addition, by combining these technologies, AI-enabled integrative tools could unveil the invisible aspects of protein 'life' revealing hidden protein states and reconstructing allosteric landscapes for therapeutically important regulatory proteins operating under allosteric control. Through potential synergies of these innovative technologies, AI-enhanced biochemical and structural biology tools can create novel ways of controlling the structure and activity of proteins to achieve desired outcomes, leading to a new era of protein engineering and synthetic biology.

7. Concluding Remarks and Future Perspectives

Despite the accepted understanding that intricate protein systems and regulatory complexes often function as dynamic and diverse allosteric machines, the characterization of hidden and uncommon protein functional states, allosteric structural changes, and allosteric pathways is still surprisingly scarce, necessitating the combination of new structural, biophysical, and computational approaches to address these problems. Research in the area of allosteric mechanisms is advancing in several new directions, such as the use of computational methods for the detection and visualization of allosteric networks, as well as the development of novel experimental methods to study allosteric processes. The latest remarkable progress in deep high-throughput mining and comprehensive mapping of allosteric protein landscapes and allosteric mechanisms using novel experimental technologies created a wealth of high-quality data that were used in the development of new experiment-guided AI systems. The recent studies have shown that the continuous developments and adaptations of AI and ML tools in high throughput biochemistry and their integration with DMS tools can reveal atomistic details of allosteric landscapes and result in a comprehensive characterization of allosteric mutations, yielding unprecedented insights into molecular mechanisms of complex biological systems.

The latest advances in structural characterization of allosteric molecular events and hidden functional states using cryo-EM, NMR, smFRET spectroscopy have highlighted the growing need for data-centric integrative biophysics approaches. These emerging methods and tools, including high-throughput sequencing, microscopy imaging, proteomics, metabolomics, and systems biology, involve utilization of AI and ML models trained on large datasets from different sources. AI-augmented integrative biophysics approaches leverage the power of AI to facilitate the integration of multiple data sources and to discover novel physical principles underlying biological systems. The latest developments reviewed in this study suggest that AI can be used to develop more accurate predictive models of diverse and complex biological systems, identify relevant features from large scale datasets, and provide insights into the underlying physical mechanisms of biological processes. AI-augmented biophysics approaches in structural biology can also be used to automate the experimental design process and optimize laboratory techniques, allowing scientists to conduct research and identify meaningful patterns in the data quickly and efficiently. In addition, AI-augmented integrative biophysics approaches will provide new strategies for drug design and development—from allosteric drug discovery to pathway-targeted design, systems medicine, and protein engineering applications. By developing an open science infrastructure for machine learning studies of allosteric regulation and validating computational approaches using integrative studies of allosteric mechanisms, the scientific community can expand the toolkit of approaches 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 knowledgebase to enable interrogation of the allosteric functions can provide a much-needed boost to further innovation and integration of experimental and computational technologies empowered by a booming AI field.

Author Contributions: Conceptualization, G.V. and P.T.; methodology, G.V. and P.T.; software, G.V., S.X., M.A., G.G. and P.T.; validation, G.V.; formal analysis, G.V., M.A. and G.G.; investigation, G.V. and P.T.; resources, G.V., M.A. and G.G.; data curation, G.V.; writing—original draft preparation, G.V.; writing—review and editing, G.V., M.A. and G.G.; visualization, G.V.; supervision, G.V.; project administration, G.V.; funding acquisition, G.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Kay Family Foundation Grant A20-0032 and National Institutes of Health under Award No. R15GM122013.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are fully contained within the article. Crystal structures were obtained and downloaded from the Protein Data Bank (<http://www.rcsb.org>) (accessed on 5 April 2023). 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/>) (accessed on 5 April 2023). All the software tools are freely available in the GitHub sites <https://github.com/smu-tao-group>; <https://github.com/smu-tao-group/protein-VAE>; <https://github.com/smu-tao-group/PASSer2.0>; <https://github.com/stevenagajanian/KInGAN>; and https://github.com/kassabry/Perturbation_Experiment (accessed on 12 April 2023).

Acknowledgments: G.V. acknowledges support from Schmid College of Science and Technology at Chapman University for providing computing resources at the Keck Center for Science and Engineering.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationship that could be construed as a potential conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Goh, G.B.; Hodas, N.O.; Vishnu, A. Deep learning for computational chemistry. *J. Comput. Chem.* **2017**, *38*, 1291–1307. [[CrossRef](#)]
2. Dimitrov, T.; Kreisbeck, C.; Becker, J.S.; Aspuru-Guzik, A.; Saikin, S.K. Autonomous Molecular Design: Then and Now. *ACS Appl. Mater. Interfaces* **2019**, *11*, 24825–24836. [[CrossRef](#)] [[PubMed](#)]
3. Mater, A.C.; Coote, M.L. Deep Learning in Chemistry. *J. Chem. Inf. Model.* **2019**, *59*, 2545–2559. [[CrossRef](#)]
4. Chen, H.; Engkvist, O.; Wang, Y.; Olivecrona, M.; Blaschke, T. The rise of deep learning in drug discovery. *Drug Discov. Today* **2018**, *23*, 1241–1250. [[CrossRef](#)]
5. Popova, M.; Isayev, O.; Tropsha, A. Deep reinforcement learning for de novo drug design. *Sci. Adv.* **2018**, *4*, eaap7885. [[CrossRef](#)]
6. Changeux, J.-P. Allostery and the Monod-Wyman-Changeux Model after 50 Years. *Annu. Rev. Biophys.* **2012**, *41*, 103–133. [[CrossRef](#)] [[PubMed](#)]
7. Changeux, J.-P.; Edelstein, S.J. Allosteric Mechanisms of Signal Transduction. *Science* **2005**, *308*, 1424–1428. [[CrossRef](#)]
8. Popovych, N.; Sun, S.; Ebright, R.; Kalodimos, C.G. Dynamically driven protein allostery. *Nat. Struct. Mol. Biol.* **2006**, *13*, 831–838. [[CrossRef](#)] [[PubMed](#)]
9. Hilser, V.J.; Wrabl, J.O.; Motlagh, H.N. Structural and Energetic Basis of Allostery. *Annu. Rev. Biophys.* **2012**, *41*, 585–609. [[CrossRef](#)] [[PubMed](#)]
10. Motlagh, H.N.; Wrabl, J.O.; Li, J.; Hilser, V.J. The ensemble nature of allostery. *Nature* **2014**, *508*, 331–339. [[CrossRef](#)] [[PubMed](#)]
11. Tsai, C.-J.; Nussinov, R. A Unified View of “How Allostery Works”. *PLoS Comput. Biol.* **2014**, *10*, e1003394. [[CrossRef](#)]
12. Liu, J.; Nussinov, R. Allostery: An Overview of Its History, Concepts, Methods, and Applications. *PLoS Comput. Biol.* **2016**, *12*, e1004966. [[CrossRef](#)]
13. Wodak, S.J.; Paci, E.; Dokholyan, N.V.; Berezovsky, I.N.; Horovitz, A.; Li, J.; Hilser, V.J.; Bahar, I.; Karanicolas, J.; Stock, G.; et al. Allostery in Its Many Disguises: From Theory to Applications. *Structure* **2019**, *27*, 566–578. [[CrossRef](#)]
14. Nussinov, R.; Tsai, C.-J.; Jang, H. Allostery, and how to define and measure signal transduction. *Biophys. Chem.* **2022**, *283*, 106766. [[CrossRef](#)] [[PubMed](#)]
15. Huang, C.; Kalodimos, C.G. Structures of Large Protein Complexes Determined by Nuclear Magnetic Resonance Spectroscopy. *Annu. Rev. Biophys.* **2017**, *46*, 317–336. [[CrossRef](#)] [[PubMed](#)]

16. Jiang, Y.; Kalodimos, C.G. NMR Studies of Large Proteins. *J. Mol. Biol.* **2017**, *429*, 2667–2676. [[CrossRef](#)]
17. Kay, L.E. New Views of Functionally Dynamic Proteins by Solution NMR Spectroscopy. *J. Mol. Biol.* **2016**, *428*, 323–331. [[CrossRef](#)] [[PubMed](#)]
18. Lisi, G.P.; Loria, J.P. Solution NMR Spectroscopy for the Study of Enzyme Allostery. *Chem. Rev.* **2016**, *116*, 6323–6369. [[CrossRef](#)]
19. Vallurupalli, P.; Bouvignies, G.; Kay, L.E. Studying “Invisible” Excited Protein States in Slow Exchange with a Major State Conformation. *J. Am. Chem. Soc.* **2012**, *134*, 8148–8161. [[CrossRef](#)]
20. Sekhar, A.; Kay, L.E. NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 12867–12874. [[CrossRef](#)]
21. Munte, C.E.; Erlach, M.B.; Kremer, W.; Koehler, J.; Kalbitzer, H.R. Distinct Conformational States of the Alzheimer β -Amyloid Peptide Can Be Detected by High-Pressure NMR Spectroscopy. *Angew. Chem. Int. Ed.* **2013**, *52*, 8943–8947. [[CrossRef](#)]
22. Kalbitzer, H.R.; Rosnizeck, I.C.; Munte, C.E.; Narayanan, S.P.; Kropf, V.; Spoerner, P.-D.D.M. Intrinsic Allosteric Inhibition of Signaling Proteins by Targeting Rare Interaction States Detected by High-Pressure NMR Spectroscopy. *Angew. Chem. Int. Ed.* **2013**, *52*, 14242–14246. [[CrossRef](#)] [[PubMed](#)]
23. ElGamacy, M.; Riss, M.; Zhu, H.; Truffault, V.; Coles, M. Mapping Local Conformational Landscapes of Proteins in Solution. *Structure* **2019**, *27*, 853–865.e5. [[CrossRef](#)] [[PubMed](#)]
24. Bostock, M.J.; Solt, A.S.; Nietlispach, D. The role of NMR spectroscopy in mapping the conformational landscape of GPCRs. *Curr. Opin. Struct. Biol.* **2019**, *57*, 145–156. [[CrossRef](#)] [[PubMed](#)]
25. Usher, E.T.; Showalter, S.A. Mapping invisible epitopes by NMR spectroscopy. *J. Biol. Chem.* **2020**, *295*, 17411–17412. [[CrossRef](#)]
26. Xie, T.; Saleh, T.; Rossi, P.; Kalodimos, C.G. Conformational states dynamically populated by a kinase determine its function. *Science* **2020**, *370*, eabc2754. [[CrossRef](#)] [[PubMed](#)]
27. Banerjee, A.; Saha, S.; Tvedt, N.C.; Yang, L.-W.; Bahar, I. Mutually beneficial confluence of structure-based modeling of protein dynamics and machine learning methods. *Curr. Opin. Struct. Biol.* **2023**, *78*, 102517. [[CrossRef](#)] [[PubMed](#)]
28. Hernández, C.X.; Wayment-Steele, H.K.; Sultan, M.M.; Husic, B.E.; Pande, V.S. Variational encoding of complex dynamics. *Phys. Rev. E* **2018**, *97*, 062412. [[CrossRef](#)]
29. Novelli, P.; Bonati, L.; Pontil, M.; Parrinello, M. Characterizing Metastable States with the Help of Machine Learning. *J. Chem. Theory Comput.* **2022**, *18*, 5195–5202. [[CrossRef](#)]
30. Ribeiro, J.M.L.; Bravo, P.; Wang, Y.; Tiwary, P. Reweighted autoencoded variational Bayes for enhanced sampling (RAVE). *J. Chem. Phys.* **2018**, *149*, 072301. [[CrossRef](#)]
31. Audagnotto, M.; Czechtizky, W.; De Maria, L.; Käck, H.; Papoian, G.; Tornberg, L.; Tyrchan, C.; Ulander, J. Machine learning/molecular dynamic protein structure prediction approach to investigate the protein conformational ensemble. *Sci. Rep.* **2022**, *12*, 10018. [[CrossRef](#)] [[PubMed](#)]
32. Shamsi, Z.; Cheng, K.J.; Shukla, D. Reinforcement Learning Based Adaptive Sampling: REAPing Rewards by Exploring Protein Conformational Landscapes. *J. Phys. Chem. B* **2018**, *122*, 8386–8395. [[CrossRef](#)]
33. Kleiman, D.E.; Shukla, D. Multiagent Reinforcement Learning-Based Adaptive Sampling for Conformational Dynamics of Proteins. *J. Chem. Theory Comput.* **2022**, *18*, 5422–5434. [[CrossRef](#)]
34. Zimmerman, M.I.; Porter, J.; Sun, X.; Silva, R.R.; Bowman, G.R. Choice of Adaptive Sampling Strategy Impacts State Discovery, Transition Probabilities, and the Apparent Mechanism of Conformational Changes. *J. Chem. Theory Comput.* **2018**, *14*, 5459–5475. [[CrossRef](#)] [[PubMed](#)]
35. Noé, F.; Olsson, S.; Köhler, J.; Wu, H. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science* **2019**, *365*, eaaw1147. [[CrossRef](#)] [[PubMed](#)]
36. Agajanian, S.; Alshahrani, M.; Bai, F.; Tao, P.; Verkhivker, G.M. Exploring and Learning the Universe of Protein Allostery Using Artificial Intelligence Augmented Biophysical and Computational Approaches. *J. Chem. Inf. Model.* **2023**, *63*, 1413–1428. [[CrossRef](#)] [[PubMed](#)]
37. Fowler, D.M.; Fields, S. Deep mutational scanning: A new style of protein science. *Nat. Methods* **2014**, *11*, 801–807. [[CrossRef](#)]
38. Starr, T.N.; Greaney, A.J.; Hilton, S.K.; Ellis, D.; Crawford, K.H.D.; Dingens, A.S.; Navarro, M.J.; Bowen, J.E.; Tortorici, M.A.; Walls, A.C.; et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* **2020**, *182*, 1295–1310.e20. [[CrossRef](#)] [[PubMed](#)]
39. Greaney, A.J.; Starr, T.N.; Gilchuk, P.; Zost, S.J.; Binshtein, E.; Loes, A.N.; Hilton, S.K.; Huddleston, J.; Eguia, R.; Crawford, K.H.; et al. Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition. *Cell Host Microbe* **2021**, *29*, 44–57.e9. [[CrossRef](#)]
40. Greaney, A.J.; Loes, A.N.; Crawford, K.H.; Starr, T.N.; Malone, K.D.; Chu, H.Y.; Bloom, J.D. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* **2021**, *29*, 463–476.e6. [[CrossRef](#)]
41. Greaney, A.J.; Starr, T.N.; Barnes, C.O.; Weisblum, Y.; Schmidt, F.; Caskey, M.; Gaebler, C.; Cho, A.; Agudelo, M.; Finkin, S.; et al. Mapping mutations to the SARS-CoV-2 RBD that escape binding by different classes of antibodies. *Nat. Commun.* **2021**, *12*, 4196. [[CrossRef](#)]
42. Moulana, A.; Dupic, T.; Phillips, A.M.; Chang, J.; Roffler, A.; Greaney, A.J.; Starr, T.N.; Bloom, J.D.; Desai, M.M. The landscape of antibody binding affinity in SARS-CoV-2 Omicron BA.1 evolution. *Elife* **2023**, *12*, e83442. [[CrossRef](#)] [[PubMed](#)]

43. Dadonaite, B.; Crawford, K.H.; Radford, C.E.; Farrell, A.G.; Yu, T.C.; Hannon, W.W.; Zhou, P.; Andrabi, R.; Burton, D.R.; Liu, L.; et al. A pseudovirus system enables deep mutational scanning of the full SARS-CoV-2 spike. *Cell* **2023**, *186*, 1263–1278. [\[CrossRef\]](#)
44. Faure, A.J.; Domingo, J.; Schmiedel, J.M.; Hidalgo-Carcedo, C.; Diss, G.; Ben Lehner, B. Mapping the energetic and allosteric landscapes of protein binding domains. *Nature* **2022**, *604*, 175–183. [\[CrossRef\]](#)
45. Weng, C.; Faure, A.J.; Lehner, B. The energetic and allosteric landscape for KRAS inhibition. *bioRxiv* **2022**. [\[CrossRef\]](#)
46. Esposito, D.; Weile, J.; Shendure, J.; Starita, L.M.; Papenfuss, A.T.; Roth, F.P.; Fowler, D.M.; Rubin, A.F. MaveDB: An open-source platform to distribute and interpret data from multiplexed assays of variant effect. *Genome Biol.* **2019**, *20*, 223. [\[CrossRef\]](#)
47. Weile, J.; Sun, S.; Cote, A.G.; Knapp, J.; Verby, M.; Mellor, J.C.; Wu, Y.; Pons, C.; Wong, C.; van Lieshout, N.; et al. A framework for exhaustively mapping functional missense variants. *Mol. Syst. Biol.* **2017**, *13*, 57. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Dewachter, L.; Brooks, A.N.; Noon, K.; Cialek, C.; Clark-ElSayed, A.; Schalck, T.; Krishnamurthy, N.; Versées, W.; Vranken, W.; Michiels, J. Deep mutational scanning of essential bacterial proteins can guide antibiotic development. *Nat. Commun.* **2023**, *14*, 241. [\[CrossRef\]](#)
49. Starr, T.N.; Greaney, A.J.; Hannon, W.W.; Loes, A.N.; Hauser, K.; Dillen, J.R.; Ferri, E.; Farrell, A.G.; Dadonaite, B.; McCallum, M.; et al. Shifting mutational constraints in the SARS-CoV-2 receptor-binding domain during viral evolution. *Science* **2022**, *377*, 420–424. [\[CrossRef\]](#)
50. Moulana, A.; Dupic, T.; Phillips, A.M.; Chang, J.; Nieves, S.; Roffler, A.A.; Greaney, A.J.; Starr, T.N.; Bloom, J.D.; Desai, M.M. Compensatory epistasis maintains ACE2 affinity in SARS-CoV-2 Omicron BA.1. *Nat. Commun.* **2022**, *13*, 7011. [\[CrossRef\]](#)
51. Azbukina, N.; Zharikova, A.; Ramensky, V. Intragenic compensation through the lens of deep mutational scanning. *Biophys. Rev.* **2022**, *14*, 1161–1182. [\[CrossRef\]](#)
52. Roychowdhury, H.; Romero, P.A. Microfluidic deep mutational scanning of the human executioner caspases reveals differences in structure and regulation. *Cell Death Discov.* **2022**, *8*, 7. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Mathy, C.J.; Mishra, P.; Flynn, J.M.; Perica, T.; Mavor, D.; Bolon, D.N.; Kortemme, T. A complete allosteric map of a GTPase switch in its native cellular network. *Cell Syst.* **2023**, *14*, 237–246.e7. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Tack, D.S.; Tonner, P.D.; Pressman, A.; Olson, N.D.; Levy, S.F.; Romantseva, E.F.; Alperovich, N.; Vasilyeva, O.; Ross, D. The genotype-phenotype landscape of an allosteric protein. *Mol. Syst. Biol.* **2021**, *17*, e10179. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Wei, H.; Li, X. Deep mutational scanning: A versatile tool in systematically mapping genotypes to phenotypes. *Front. Genet.* **2023**, *14*, 1087267. [\[CrossRef\]](#)
56. Song, H.; Bremer, B.J.; Hinds, E.C.; Raskutti, G.; Romero, P.A. Inferring Protein Sequence-Function Relationships with Large-Scale Positive-Unlabeled Learning. *Cell Syst.* **2021**, *12*, 92–101.e8. [\[CrossRef\]](#)
57. Leander, M.; Yuan, Y.; Meger, A.; Cui, Q.; Raman, S. Functional plasticity and evolutionary adaptation of allosteric regulation. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 25445–25454. [\[CrossRef\]](#)
58. 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*, e79932. [\[CrossRef\]](#)
59. Gelman, S.; Fahlberg, S.A.; Heinzelman, P.; Romero, P.A.; Gitter, A. Neural networks to learn protein sequence–function relationships from deep mutational scanning data. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2104878118. [\[CrossRef\]](#)
60. D’costa, S.; Hinds, E.C.; Freschlin, C.R.; Song, H.; Romero, P.A. Inferring protein fitness landscapes from laboratory evolution experiments. *PLoS Comput. Biol.* **2023**, *19*, e1010956. [\[CrossRef\]](#)
61. Freschlin, C.R.; Fahlberg, S.; Romero, P. Machine learning to navigate fitness landscapes for protein engineering. *Curr. Opin. Biotechnol.* **2022**, *75*, 102713. [\[CrossRef\]](#) [\[PubMed\]](#)
62. Vaishnav, E.D.; de Boer, C.G.; Molinet, J.; Yassour, M.; Fan, L.; Adiconis, X.; Thompson, D.A.; Levin, J.Z.; Cubillos, F.A.; Regev, A. The evolution, evolvability and engineering of gene regulatory DNA. *Nature* **2022**, *603*, 455–463. [\[CrossRef\]](#) [\[PubMed\]](#)
63. Tareen, A.; Kooshkbaghi, M.; Posfai, A.; Ireland, W.T.; McCandlish, D.M.; Kinney, J.B. MAVE-NN: Learning genotype-phenotype maps from multiplex assays of variant effect. *Genome Biol.* **2022**, *23*, 98. [\[CrossRef\]](#) [\[PubMed\]](#)
64. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [\[CrossRef\]](#) [\[PubMed\]](#)
65. Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Židek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon, A.; et al. Highly accurate protein structure prediction for the human proteome. *Nature* **2021**, *596*, 590–596. [\[CrossRef\]](#)
66. Fleishman, S.J.; Horovitz, A. Extending the New Generation of Structure Predictors to Account for Dynamics and Allostery. *J. Mol. Biol.* **2021**, *433*, 167007. [\[CrossRef\]](#)
67. Nussinov, R.; Zhang, M.; Liu, Y.; Jang, H. AlphaFold, Artificial Intelligence (AI), and Allostery. *J. Phys. Chem. B* **2022**, *126*, 6372–6383. [\[CrossRef\]](#)
68. Nussinov, R.; Zhang, M.; Liu, Y.; Jang, H. AlphaFold, allosteric, and orthosteric drug discovery: Ways forward. *Drug Discov. Today* **2023**, *28*, 103551. [\[CrossRef\]](#)
69. del Alamo, D.; Sala, D.; Mchaourab, H.S.; Meiler, J. Sampling alternative conformational states of transporters and receptors with AlphaFold2. *Elife* **2022**, *11*, e75751. [\[CrossRef\]](#)
70. Saldaño, T.; Escobedo, N.; Marchetti, J.; Zea, D.J.; Mac Donagh, J.; Rueda, A.J.V.; Gonik, E.; Melani, A.G.; Nechcoff, J.N.; Salas, M.N.; et al. Impact of protein conformational diversity on AlphaFold predictions. *Bioinformatics* **2022**, *38*, 2742–2748. [\[CrossRef\]](#)

71. Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **2023**, *379*, 1123–1130. [[CrossRef](#)] [[PubMed](#)]
72. Watson, J.L.; Juergens, D.; Bennett, N.R.; Trippe, B.L.; Yim, J.; Eisenach, H.E.; Ahern, W.; Borst, A.J.; Ragotte, R.J.; Milles, L.F.; et al. Broadly Applicable and Accurate Protein Design by Integrating Structure Prediction Networks and Diffusion Generative Models, Cold Spring Harbor Laboratory. *bioRxiv* **2022**. [[CrossRef](#)]
73. Bonati, L.; Piccini, G.; Parrinello, M. Deep learning the slow modes for rare events sampling. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2113533118. [[CrossRef](#)] [[PubMed](#)]
74. Wang, Y.; Ribeiro, J.M.L.; Tiwary, P. Past–future information bottleneck for sampling molecular reaction coordinate simultaneously with thermodynamics and kinetics. *Nat. Commun.* **2019**, *10*, 3573. [[CrossRef](#)] [[PubMed](#)]
75. Wang, D.; Tiwary, P. State predictive information bottleneck. *J. Chem. Phys.* **2021**, *154*, 134111. [[CrossRef](#)]
76. 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*, 3231–3238. [[CrossRef](#)]
77. Tsai, S.-T.; Fields, E.; Xu, Y.; Kuo, E.-J.; Tiwary, P. Path sampling of recurrent neural networks by incorporating known physics. *Nat. Commun.* **2022**, *13*, 7231. [[CrossRef](#)]
78. Janson, G.; Valdes-Garcia, G.; Heo, L.; Feig, M. Direct generation of protein conformational ensembles via machine learning. *Nat. Commun.* **2023**, *4*, 774. [[CrossRef](#)]
79. Mardt, A.; Pasquali, L.; Wu, H.; Noé, F. VAMPnets for deep learning of molecular kinetics. *Nat. Commun.* **2018**, *9*, 5. [[CrossRef](#)]
80. Ghorbani, M.; Prasad, S.; Klauda, J.B.; Brooks, B.R. Variational embedding of protein folding simulations using Gaussian mixture variational autoencoders. *J. Chem. Phys.* **2021**, *155*, 194108. [[CrossRef](#)]
81. Ghorbani, M.; Prasad, S.; Klauda, J.B.; Brooks, B.R. GraphVAMPNet, using graph neural networks and variational approach to Markov processes for dynamical modeling of biomolecules. *J. Chem. Phys.* **2022**, *156*, 184103. [[CrossRef](#)] [[PubMed](#)]
82. Jin, Y.; Johannissen, L.O.; Hay, S. Predicting new protein conformations from molecular dynamics simulation conformational landscapes and machine learning. *Proteins* **2021**, *89*, 915–921. [[CrossRef](#)] [[PubMed](#)]
83. Tian, H.; Jiang, X.; Trozzi, F.; Xiao, S.; Larson, E.C.; Tao, P. Explore Protein Conformational Space with Variational Autoencoder. *Front. Mol. Biosci.* **2021**, *8*, 781635. [[CrossRef](#)]
84. Tian, H.; Jiang, X.; Xiao, S.; La Force, H.; Larson, E.C.; Tao, P. LAST: Latent Space-Assisted Adaptive Sampling for Protein Trajectories. *J. Chem. Inf. Model.* **2022**, *63*, 67–75. [[CrossRef](#)] [[PubMed](#)]
85. Uyar, A.; Karamyan, V.; Dickson, A. Long-Range Changes in Neurolysin Dynamics Upon Inhibitor Binding. *J. Chem. Theory Comput.* **2018**, *14*, 444–452. [[CrossRef](#)]
86. Duro, N.; Varma, S. Role of Structural Fluctuations in Allosteric Stimulation of Paramyxovirus Hemagglutinin-Neuraminidase. *Structure* **2019**, *27*, 1601–1611.e2. [[CrossRef](#)]
87. Zhou, H.; Dong, Z.; Verkhivker, G.; Zoltowski, B.D.; Tao, P. Allosteric mechanism of the circadian protein Vivid resolved through Markov state model and machine learning analysis. *PLoS Comput. Biol.* **2019**, *15*, e1006801. [[CrossRef](#)]
88. 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*, 4043–4051. [[CrossRef](#)]
89. Greener, J.G.; Sternberg, M.J. AlloPred: Prediction of allosteric pockets on proteins using normal mode perturbation analysis. *BMC Bioinform.* **2015**, *16*, 335. [[CrossRef](#)]
90. Huang, W.; Lu, S.; Huang, Z.; Liu, X.; Mou, L.; Luo, Y.; Zhao, Y.; Liu, Y.; Chen, Z.; Hou, T.; et al. AlloSite: A method for predicting allosteric sites. *Bioinformatics* **2013**, *29*, 2357–2359. [[CrossRef](#)]
91. Song, K.; Liu, X.; Huang, W.; Lu, S.; Shen, Q.; Zhang, L.; Zhang, J. Improved Method for the Identification and Validation of Allosteric Sites. *J. Chem. Inf. Model.* **2017**, *57*, 2358–2363. [[CrossRef](#)]
92. Lu, S.; Shen, Q.; Zhang, J. Allosteric Methods and Their Applications: Facilitating the Discovery of Allosteric Drugs and the Investigation of Allosteric Mechanisms. *Accounts Chem. Res.* **2019**, *52*, 492–500. [[CrossRef](#)]
93. Mishra, S.K.; Kandoi, G.; Jernigan, R.L. Coupling dynamics and evolutionary information with structure to identify protein regulatory and functional binding sites. *Proteins Struct. Funct. Bioinform.* **2019**, *87*, 850–868. [[CrossRef](#)] [[PubMed](#)]
94. Le Guilloux, V.; Schmidtke, P.; Tuffery, P. Fpocket: An open source platform for ligand pocket detection. *BMC Bioinform.* **2009**, *10*, 168. [[CrossRef](#)] [[PubMed](#)]
95. Ghersi, D.; Sanchez, R. EasyMIFs and SiteHound: A toolkit for the identification of ligand-binding sites in protein structures. *Bioinformatics* **2009**, *25*, 3185–3186. [[CrossRef](#)] [[PubMed](#)]
96. Zhang, Z.; Li, Y.; Lin, B.; Schroeder, M.; Huang, B. Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. *Bioinformatics* **2011**, *27*, 2083–2088. [[CrossRef](#)]
97. Jian, J.-W.; Elumalai, P.; Pitti, T.; Wu, C.Y.; Tsai, K.-C.; Chang, J.-Y.; Peng, H.-P.; Yang, A.-S. Predicting Ligand Binding Sites on Protein Surfaces by 3-Dimensional Probability Density Distributions of Interacting Atoms. *PLoS ONE* **2016**, *11*, e0160315. [[CrossRef](#)]
98. Jiménez, J.; Doerr, S.; Martínez-Rosell, G.; Rose, A.S.; De Fabritiis, G. DeepSite: Protein-binding site predictor using 3D-convolutional neural networks. *Bioinformatics* **2017**, *33*, 3036–3042. [[CrossRef](#)]
99. Jiang, M.; Wei, Z.; Zhang, S.; Wang, S.; Wang, X.; Li, Z. FRSite: Protein drug binding site prediction based on faster R-CNN. *J. Mol. Graph. Model.* **2019**, *93*, 107454. [[CrossRef](#)]

100. Stepniewska-Dziubinska, M.M.; Zielenkiewicz, P.; Siedlecki, P. Improving detection of protein-ligand binding sites with 3D segmentation. *Sci. Rep.* **2020**, *10*, 5035. [[CrossRef](#)]
101. 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*, 1661. [[CrossRef](#)] [[PubMed](#)]
102. Ward, M.D.; Zimmerman, M.I.; Meller, A.; Chung, M.; Swamidass, S.J.; Bowman, G.R. Deep learning the structural determinants of protein biochemical properties by comparing structural ensembles with DiffNets. *Nat. Commun.* **2021**, *12*, 3023. [[CrossRef](#)] [[PubMed](#)]
103. Mylonas, S.K.; Axenopoulos, A.; Daras, P. DeepSurf: A surface-based deep learning approach for the prediction of ligand binding sites on proteins. *Bioinformatics* **2021**, *37*, 1681–1690. [[CrossRef](#)] [[PubMed](#)]
104. Krivák, R.; Hoksza, D. P2Rank: Machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *J. Cheminform.* **2018**, *10*, 39. [[CrossRef](#)] [[PubMed](#)]
105. Jakubec, D.; Skoda, P.; Krivak, R.; Novotny, M.; Hoksza, D. PrankWeb 3: Accelerated ligand-binding site predictions for experimental and modelled protein structures. *Nucleic Acids Res.* **2022**, *50*, W593–W597. [[CrossRef](#)] [[PubMed](#)]
106. Kozlovskii, I.; Popov, P. Spatiotemporal identification of druggable binding sites using deep learning. *Commun. Biol.* **2020**, *3*, 618. [[CrossRef](#)]
107. Schmidtke, P.; Souaille, C.; Estienne, F.; Baurin, N.; Kroemer, R.T. Large-Scale Comparison of Four Binding Site Detection Algorithms. *J. Chem. Inf. Model.* **2010**, *50*, 2191–2200. [[CrossRef](#)]
108. Tian, H.; Jiang, X.; Tao, P. PASSer: Prediction of allosteric sites server. *Mach. Learn. Sci. Technol.* **2021**, *2*, 035015. [[CrossRef](#)]
109. Xiao, S.; Tian, H.; Tao, P. PASSer2.0: Accurate Prediction of Protein Allosteric Sites through Automated Machine Learning. *Front. Mol. Biosci.* **2022**, *9*, 879251. [[CrossRef](#)]
110. Tian, H.; Xiao, S.; Jiang, X.; Tao, P. PASSerRank: Prediction of Allosteric Sites with Learning to Rank. *arXiv* **2023**, arXiv:2302.01117v1.
111. Meller, A.; Ward, M.; Borowsky, J.; Kshirsagar, M.; Lotthammer, J.M.; Oviedo, F.; Ferres, J.L.; Bowman, G.R. Predicting locations of cryptic pockets from single protein structures using the PocketMiner graph neural network. *Nat. Commun.* **2023**, *14*, 1177. [[CrossRef](#)]
112. Meller, A.; Lotthammer, J.M.; Smith, L.G.; Novak, B.; Lee, L.; Kuhn, C.C.; Greenberg, L.; Leinwand, L.; Greenberg, M.J.; Bowman, G.R. Drug specificity and affinity are encoded in the probability of cryptic pocket opening in myosin motor domains. *Elife* **2023**, *12*, e83602. [[CrossRef](#)] [[PubMed](#)]
113. Merk, A.; Bartesaghi, A.; Banerjee, S.; Falconieri, V.; Rao, P.; Davis, M.I.; Pragani, R.; Boxer, M.B.; Earl, L.A.; Milne, J.L.; et al. Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell* **2016**, *165*, 1698–1707. [[CrossRef](#)]
114. Coffino, P.; Cheng, Y. Allostery Modulates Interactions between Proteasome Core Particles and Regulatory Particles. *Biomolecules* **2022**, *12*, 764. [[CrossRef](#)] [[PubMed](#)]
115. Gulati, S.; Palczewski, K.; Engel, A.; Stahlberg, H.; Kovacic, L. Cryo-EM structure of phosphodiesterase 6 reveals insights into the allosteric regulation of type I phosphodiesterases. *Sci. Adv.* **2019**, *5*, eaav4322. [[CrossRef](#)]
116. Chung, J.M.; Durie, C.L.; Lee, J. Artificial Intelligence in Cryo-Electron Microscopy. *Life* **2022**, *12*, 1267. [[CrossRef](#)] [[PubMed](#)]
117. Wang, F.; Gong, H.; Liu, G.; Li, M.; Yan, C.; Xia, T.; Li, X.; Zeng, J. DeepPicker: A deep learning approach for fully automated particle picking in cryo-EM. *J. Struct. Biol.* **2016**, *195*, 325–336. [[CrossRef](#)]
118. Nguyen, N.P.; Ersoy, I.; Gotberg, J.; Bunyak, F.; White, T.A. DRPnet: Automated particle picking in cryo-electron micrographs using deep regression. *BMC Bioinform.* **2021**, *22*, 55. [[CrossRef](#)]
119. Al-Azzawi, A.; Oquadou, A.; Max, H.; Duan, Y.; Tanner, J.J.; Cheng, J. DeepCryoPicker: Fully automated deep neural network for single protein particle picking in cryo-EM. *BMC Bioinform.* **2020**, *21*, 509. [[CrossRef](#)]
120. Zhu, Y.; Ouyang, Q.; Mao, Y. A deep convolutional neural network approach to single-particle recognition in cryo-electron microscopy. *BMC Bioinform.* **2017**, *18*, 1–10. [[CrossRef](#)]
121. Zhong, E.D.; Bepler, T.; Berger, B.; Davis, J.H. CryoDRGN: Reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat. Methods* **2021**, *18*, 176–185. [[CrossRef](#)] [[PubMed](#)]
122. Kinman, L.F.; Powell, B.M.; Zhong, E.D.; Berger, B.; Davis, J.H. Uncovering structural ensembles from single-particle cryo-EM data using cryoDRGN. *Nat. Protoc.* **2023**, *18*, 319–339. [[CrossRef](#)] [[PubMed](#)]
123. Hamitouche, I.; Jonic, S. DeepHEMNMA: ResNet-based hybrid analysis of continuous conformational heterogeneity in cryo-EM single particle images. *Front. Mol. Biosci.* **2022**, *9*, 965645. [[CrossRef](#)] [[PubMed](#)]
124. Lerner, E.; Cordes, T.; Ingargiola, A.; Alhadid, Y.; Chung, S.; Michalet, X.; Weiss, S. Toward dynamic structural biology: Two decades of single-molecule Förster resonance energy transfer. *Science* **2018**, *359*, eaan1133. [[CrossRef](#)] [[PubMed](#)]
125. Lerner, E.; Barth, A.; Hendrix, J.; Ambrose, B.; Birkedal, V.; Blanchard, S.C.; Börner, R.; Chung, H.S.; Cordes, T.; Craggs, T.D.; et al. FRET-based dynamic structural biology: Challenges, perspectives and an appeal for open-science practices. *Elife* **2021**, *10*, e60416. [[CrossRef](#)]
126. Agam, G.; Gebhardt, C.; Popara, M.; Mächtel, R.; Folz, J.; Ambrose, B.; Chamachi, N.; Chung, S.Y.; Craggs, T.D.; de Boer, M.; et al. Reliability and accuracy of single-molecule FRET studies for characterization of structural dynamics and distances in proteins. *Nat. Methods* **2023**, *20*, 523–535. [[CrossRef](#)]
127. Matsunaga, Y.; Sugita, Y. Linking time-series of single-molecule experiments with molecular dynamics simulations by machine learning. *Elife* **2018**, *7*, e32668. [[CrossRef](#)]

128. Matsunaga, Y.; Sugita, Y. Use of single-molecule time-series data for refining conformational dynamics in molecular simulations. *Curr. Opin. Struct. Biol.* **2020**, *61*, 153–159. [[CrossRef](#)]
129. Zheng, Y.; Cui, Q. Multiple Pathways and Time Scales for Conformational Transitions in apo-Adenylate Kinase. *J. Chem. Theory Comput.* **2018**, *14*, 1716–1726. [[CrossRef](#)]
130. Dimura, M.; Peulen, T.-O.; Sanabria, H.; Rodnin, D.; Hemmen, K.; Hanke, C.A.; Seidel, C.A.M.; Gohlke, H. Automated and optimally FRET-assisted structural modeling. *Nat. Commun.* **2020**, *11*, 5394. [[CrossRef](#)]
131. Liu, X.; Jiang, Y.; Cui, Y.; Yuan, J.; Fang, X. Deep learning in single-molecule imaging and analysis: Recent advances and prospects. *Chem. Sci.* **2022**, *13*, 11964–11980. [[CrossRef](#)] [[PubMed](#)]
132. Thomsen, J.; Sletfjerding, M.B.; Jensen, S.B.; Stella, S.; Paul, B.; Malle, M.G.; Montoya, G.; Petersen, T.C.; Hatzakis, N.S. DeepFRET, a software for rapid and automated single-molecule FRET data classification using deep learning. *Elife* **2020**, *9*, e60404. [[CrossRef](#)] [[PubMed](#)]
133. Li, J.; Zhang, L.; Johnson-Buck, A.; Walter, N.G. Automatic classification and segmentation of single-molecule fluorescence time traces with deep learning. *Nat. Commun.* **2020**, *11*, 5833. [[CrossRef](#)]
134. Durham, R.J.; Paudyal, N.; Carrillo, E.; Bhatia, N.K.; Maclean, D.M.; Berka, V.; Dolino, D.M.; Gorfe, A.A.; Jayaraman, V. Conformational spread and dynamics in allostery of NMDA receptors. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 3839–3847. [[CrossRef](#)]
135. Díaz-Salinas, M.A.; Li, Q.; Ejemel, M.; Yurkovetskiy, L.; Luban, J.; Shen, K.; Wang, Y.; Munro, J.B. Conformational dynamics and allosteric modulation of the SARS-CoV-2 spike. *Elife* **2022**, *11*, e75433. [[CrossRef](#)]
136. 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*, 3350–3359. [[CrossRef](#)] [[PubMed](#)]
137. Cao, A.-M.; Quast, R.B.; Fatemi, F.; Rondard, P.; Pin, J.-P.; Margeat, E. Allosteric modulators enhance agonist efficacy by increasing the residence time of a GPCR in the active state. *Nat. Commun.* **2021**, *12*, 5426. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.