AlphaFold, Artificial Intelligence (AI), and Allostery

Published as part of The Journal of Physical Chemistry virtual special issue "Protein Folding and Dynamics—An Overview on the Occasion of Harold Scheraga's 100th Birthday".

Ruth Nussinov,* Mingzhen Zhang, Yonglan Liu, and Hyunbum Jang

ABSTRACT: AlphaFold has burst into our lives. A powerful algorithm that underscores the strength of biological sequence data and artificial intelligence (AI). AlphaFold has appended projects and research directions. The database it has been creating promises an untold number of applications with vast potential impacts that are still difficult to surmise. AI approaches can revolutionize personalized treatments and usher in better-informed clinical trials. They promise to make giant leaps toward reshaping and revamping drug discovery strategies, selecting and prioritizing combinations of drug targets. Here, we briefly overview AI in structural biology, including in molecular dynamics simulations and prediction of microbiota–human protein–protein interactions. We highlight the

advancements accomplished by the deep-learning-powered AlphaFold in protein structure prediction and their powerful impact on the life sciences. At the same time, AlphaFold does not resolve the decades-long protein folding challenge, nor does it identify the folding pathways. The models that AlphaFold provides do not capture conformational mechanisms like frustration and allostery, which are rooted in ensembles, and controlled by their dynamic distributions. Allostery and signaling are properties of populations. AlphaFold also does not generate ensembles of intrinsically disordered proteins and regions, instead describing them by their low structural probabilities. Since AlphaFold generates single ranked structures, rather than conformational ensembles, it cannot elucidate the mechanisms of allosteric activating driver hotspot mutations nor of allosteric drug resistance. However, by capturing key features, deep learning techniques can use the single predicted conformation as the basis for generating a diverse ensemble.
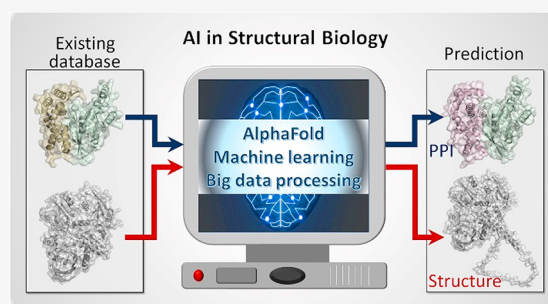
■ INTRODUCTION

AlphaFold has overcome age-long bottlenecks and forcefully bared the power of artificial intelligence (AI) in biological research.[1–3] AlphaFold has combined numerous deep learning innovations to predict the three-dimensional (3D) structures of proteins at or near experimental scale resolution, inspiring the community (including us) to rethink studies of function, evolution, and disease (e.g., refs 4–13). The sheer volume of the rapidly generated accurate structures argues that new, ambitious, frontier-pushing studies will emerge. It also points to research projects that should be reconsidered. The richness of high quality data that are being compiled in databases (e.g., refs 5 and 14–25) is already strengthening studies that require protein structures, such as mapping binding sites and interactions in signaling pathways, and identification of hot spots, including latent and rare cancer driver mutations.[26–34] The most profound impact will likely be in accelerating and improving production of new medications (e.g., ref 35), and in generating data that can be used toward this vital aim (e.g., refs 5, 17, 18, and 36–39). AI developments and applications[40] may further help foretell whether the signal propagating downstream will be strong enough to reach its genomic target

to activate (suppress) gene expression,[41] and predict pathways.[42–49] Altogether, these powerful approaches and the databases that they create revamp and transform traditional and ongoing research involving the use of structures. They also embolden us to step back, rethink, and innovate our projects.

AlphaFold's achievements have been made possible by the protein databank (PDB), currently with a size nearing 200 000 experimentally determined structures. It has been trained on protein chains from the PDB and uses the input sequence to query databases of protein sequences to construct a multiple sequence alignment.[4] However, its striking success has not led us to a deeper mechanistic understanding of exactly how a protein sequence folds, thus not assisting in the folding of a protein from its sequence. Below, we first briefly describe the protein folding problem,[50,51] and strategies to predict protein
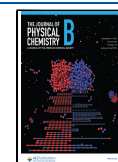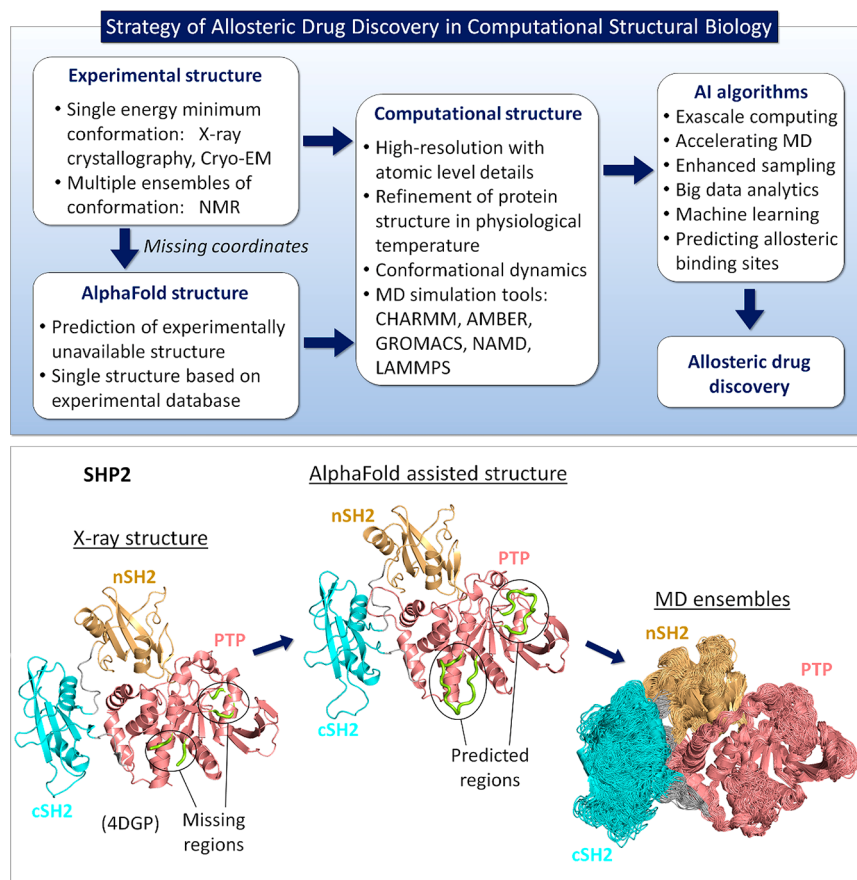
Figure 1. Current strategy of allosteric drug discovery in computational structural biology employing the AlphaFold program with artificial intelligence (AI)-powered methods (*top panel*). Experimental instruments, such as X-ray crystallography, cryo-electron microscopy (cryo-EM), and nuclear magnetic resonance (NMR) can resolve protein structures, but often miss the coordinates of highly fluctuating regions in the protein structure. AlphaFold can predict the missing coordinates of these regions. The resulting structure can be subjected to molecular dynamics (MD) simulations that provide conformational dynamics, conformational changes, and folding characteristics of the protein. An example is shown for Src homology region 2-containing protein tyrosine phosphatase 2 (SHP2) (*bottom panel*). The X-ray structure of SHP2 (PDB ID: 4DGP) misses residues in two flexible regions, which can be predicted by AlphaFold. SHP2 contains two Src homology 2 (SH2) domains (nSH2 and cSH2) and a protein tyrosine phosphatase (PTP) domain.

structures. We describe key conceptual and computational developments and the transformative AlphaFold advances. We outline its strengths and some weaknesses. We emphasize what it has accomplished, and what it has not, and the magnitude of the challenges, underscoring the difference between the *theoretical folding problem*, which was not solved,[51−57] and *practical predictions* by incorporating additional evolution information that generally have been.[58−65] We proceed to AI approaches to the complementary problem of protein−protein interactions (PPIs) by these methods and others,[66−75] with the human−microbiome PPI as a relevant and topical example.[66] AI-powered prediction of human−microbe PPIs can accelerate research into questions such as how microbiota hijack cell signaling and provide drug targets.[76−80] We discuss how AI can reshape drug discovery, for example by amplifying repurposing of FDA-approved drugs,[81−91] an area which is already thriving. AI can also select combinations of drug targets, powerfully guiding and accelerating experiments by providing specific testable hypotheses. Machine learning has already proved its merit in the life and medical sciences.[1,92−98] Coupled with harnessed exascale computing,[99,100] advanced, AI-powered methods are set to revolutionize therapeutic development, providing prioritized drug combinations for the attending physicians.

Finally, we note that AlphaFold, which predicts single ranked structures for a protein sequence, is unable to address directly allosteric mechanisms, which are *based on the populations of conformational states in the ensembles.*[101−117] Allostery, where the signal propagates dynamically with the shifts in the populations, underlies regulation and thus cell life.[118,119] Due to its higher specificity and consequently lower toxicity, which results from targeting nonconserved allosteric sites, allostery also increasingly features in allosteric drugs.[120−125]

Can we then foresee AlphaFold assisting in unraveling the mechanisms of allosteric hotspot mutations and allosteric drug discovery? Indirectly it can and does, even in our hands. The rigid structures that AlphaFold predicts can be submitted to MD simulations that generate such ensembles (Figure 1). At the same time, as we discuss here, other AI-based strategies can assist directly in such efforts, most effectively via accelerating and enhancing MD simulations. Efforts are also likely to persist in exploiting AI toward prediction of allosteric binding sites. Nevertheless, it behooves us to recall that the effectiveness of allosteric sites is determines by both stable interactions at the site, which is something that AI can help with, and initiation of effective allosteric signals, which would be more challenging. Current approaches to predict allosteric binding sites address

only the former. In that sense, they resemble the characterization of orthosteric sites, except that their scoring is based on statistics of allosteric sites.

Orthosteric drugs block the active site; allosteric drugs alter the population of the active state of the protein, including the active site, through binding at a site far away.[126] We suggested that allosteric drugs can constitute of "anchors" and "drivers" atoms, where the anchor atoms bind to the allosteric pocket, without changing the conformation of the binding site. The interactions of the anchor atoms stabilize ligand binding, resembling protein–ligand binding at the orthosteric site. The binding of driver atoms "pulls" or "pushes" atoms in the protein pocket. This initiates the allosteric signal, which shifts the receptor population from the inactive to the active state. Driver atoms can trigger agonism and antagonism. AlphaFold cannot handle population shifts. AI strategies can but will need to go beyond prediction of stabilizing interactions.

Finally, not surprisingly, prediction of the structures of intrinsically disordered proteins (IDPs) and regions (IDRs) is another problem where AlphFold falls short. Disordered proteins (regions) are characterized by broad and heterogeneous ensembles where the differences in the relative conformational stabilities are small, or even minor and the barriers are low.[127−135] The conformations interconvert, leading to low probabilities of AlphaFold's reliably capturing those most favored, or the conformational distribution. Nevertheless, leveraging, learning, and mining of the conformations can exploit AI.[136−138]

AI-powered algorithms, which are fed vast compiled data, and enabled by the emerging massive compute power are propelling a revolution in computational biology (Figure 1). Unlike quantum computing, in the case of AI and data-driven computing, the technological innovations at the requisite scales are already at hand.

PROTEIN FOLDING VERSUS PREDICTION OF PROTEIN STRUCTURE

Protein Folding. The protein folding problem embraces two questions:[51] first, the conceptual question of how a protein's amino acid sequence dictates its 3D atomic structure, and second, how, starting from a single amino acid sequence, to successfully predict the 3D structure, without using information related to other available (homologous, same family) sequences nor structures of any related sequences. Such computational prediction methods are guided by the conceptual notion that this is how the protein folds in nature. Single sequence-based prediction in solution considers forces related to hydrogen bonds, ion pairs, van der Waals attractions,[139] and chiefly water-mediated hydrophobic interactions, with the hydrophobic effect the driving force for protein folding. This formal folding problem emerged six decades ago, alongside the first atomic-resolution protein (globins) structure. The structure led to thermodynamic questions of the balance of interatomic forces that determines the structure of the protein, how the protein can fold so quickly, that is the kinetics of the pathways, and the computational problem of protein structure prediction. The landmark thermodynamic hypothesis of Christian Anfinsen and his colleagues[140,141] stated that the native structure of a protein is its thermodynamically most stable structure, and it is determined only by its amino acid sequence and the conditions it is at, with kinetics playing no role. No other considerations are at play, that is, whether it is synthesized in the lab or on the

ribosome or undergoes chaperone assisted folding. The folding paradigm stipulated that unfolded molecules will always spontaneously fold into the same shape; that is, the linear amino acid sequence specifies a protein's folded native state.[58,142−144] Anfinsen's thermodynamic hypothesis emphasized the shape of the energy landscape where the native state is the one with the lowest free energy.[141,145] Computationally, that description posed the problem of prediction of protein structure, forming the basis for approaches that dominated the field for scores of years. If only the sequence matters, along with the physicochemical forces, it should be possible for "good" algorithms to fold it. Assuming that the crystal structure represents the minimum energy state, the "goodness" of the predicted structure can then be assessed by comparison with it. Anfinsen's description combines sampling of alternative conformations, ranking them by energy and identifying the lowest energy state.[51,146−148] Subsequent efforts focused on prediction of secondary structures, although the dominant role of the hydrophobic interactions suggested that secondary structure is an outcome of the 3D structure and its cause.[149,150] The small (5−10 kcal/mol) difference in the stability of the native structures as compared to the denatured states[151] compounded the challenge that predicting methods faced.

Already early on, Cyrus Levinthal conceptualized the key problem facing the protein and the prediction algorithms:[152] the vast time scales for the protein to search the folding space and reach its most stable native state under biological conditions.[58] For prediction algorithms' sampling backbone states, the search space size grows exponentially with chain length, becoming an impossibility. Levinthal argued that there is no need to search this vast space since the energy landscape is funnel-like, rather than flat, and thus can guide sampling toward the biological conformational basin.[51,153] Packed hydrophobic cores optimize their van der Waals (vdW) interactions, restrict torsion angles, and abolish internal "holes", with hydrogen bonds and salt bridges balancing the loss of interactions with water. Harold Scheraga employed physical chemistry to pioneer studies to decipher how amino acid sequences influence the 3D folding pathways, thermodynamics, and biological activity of proteins. Neither AlphaFold nor broadly other protein structure prediction algorithms consider folding pathways. Physical chemistry is accounted for implicitly; in the case of AlphaFold, via AI.

Protein Structure Prediction. Prediction of protein structure can be template-based or template-free, which does not use global similarity to an experimental (protein data bank, PDB) structure.[58] Template-free modeling exploits physics-based energy functions. Both can exploit machine learning and AI to use data in the PDB. Template-based modeling selects a structural template and uses sequence alignment. Template-free modeling uses conformational sampling and ranking. It may start with multiple-sequence alignment to related sequences to predict local structural features, which will guide the 3D modeling followed by refining and ranking.

Integrative modeling[154−156] that assembles structures from individual components may suffer from high false-positive rate. Computational integrative approach can combine data from experimental methods, bioinformatics, physics, and statistics for rapid and accurate structure determination of protein complexes. The algorithms can integrate experimental data, such as X-ray crystallography, NMR spectroscopy, 2D and 3D electron microscopy (EM), small-angle X-ray scattering (SAXS), mass spectrometry (MS), hydrogen−deuterium

exchange (HDX), mutations, sequence conservation and covariation, and statistical analysis of known structures. Computationally, the algorithms can derive from computer vision, image processing, computational geometry, machine learning, robotics, and graph algorithms. Machine learning has however been used toward protein structure prediction.[157−163]

AlphaFold is not the first in being a machine-learning model. Its remarkable success (with scores of near 90 even in the difficult targets in the 2020 Critical Assessment of Protein Structure Prediction, CASP) was influenced by its training not only on all the PDB structures but also on structures it predicted, and it uses the structure and correlation data to predict the pairs of amino acids that are in contact as well as all amino acid pairwise distances. It also ensured that the distances between the amino acids satisfy the triangle inequality, saving time at intermediate steps.[164] To date, AlphaFold illuminates half of the dark human proteins.[10]

Still, questions remain, such as which structural states exist for a give protein, and what is the population of each state. Addressing these questions is vital to relate protein structure to function. This is where AlphaFold falls short. However, the models it produces can serve as input to generate ensembles, for example by MD simulations, which, if carried out at sufficiently longtime scales, in parallel, it should be able to produce. Simulations can sample the relevant states, can enumerate possible state combinations (multistate models), and can determine the population sizes for the states.

The Structure−Function Paradigm Overlooked Ensembles and Dynamic Energy Landscapes: AlphaFold Is Attuned but Is Unable to Address Them. The sequence−structure−function dogma was the touchstone of a generation. It dominated molecular biology for decades. It was introduced by physical chemists who explained that biological macromolecules function when they are folded. Thus, to understand how molecules function, one needs to consider their 3D structures, a transformative paradigm that became a tenet of modern biology. Today, it is broadly recognized that rigid molecules cannot perform a function, leading the way to the appreciation that to sustain life, molecular flexibility is a necessity. That however has not fully translated to the understanding of the powerful concept of the energy landscape.[165] That is, that biomolecules are dynamical objects that are always interconverting between a variety of structures with varying energies,[166] and that this is the origin of allosteric mechanisms.[167−169] This notion of flexibility as interconversion between conformations is critical for understanding biological processes and their regulation, such as protein activation as a shift of the ensemble from the inactive to the active state, how allosteric drugs work, cell signaling, and binding mechanisms through conformational selection rather than induced fit. The conceptual evolution from the classic structure−function paradigm to *dynamic energy landscapes* of biomolecular function and allosteric mechanisms, poses a challenge to AlphaFold's powerful predictions. To understand biological regulation, structure should be linked to function through protein ensembles in terms of populations and relative energies, which is the foundation of allostery (Figure 2). Despite its transformative power and vast broad impact, the AlphaFold predictions are unable to address it directly. It is only through their sampling that this functional aim can be accomplished.

Around their native states, protein landscapes consist of rapidly interconverting conformations. The ensembles are
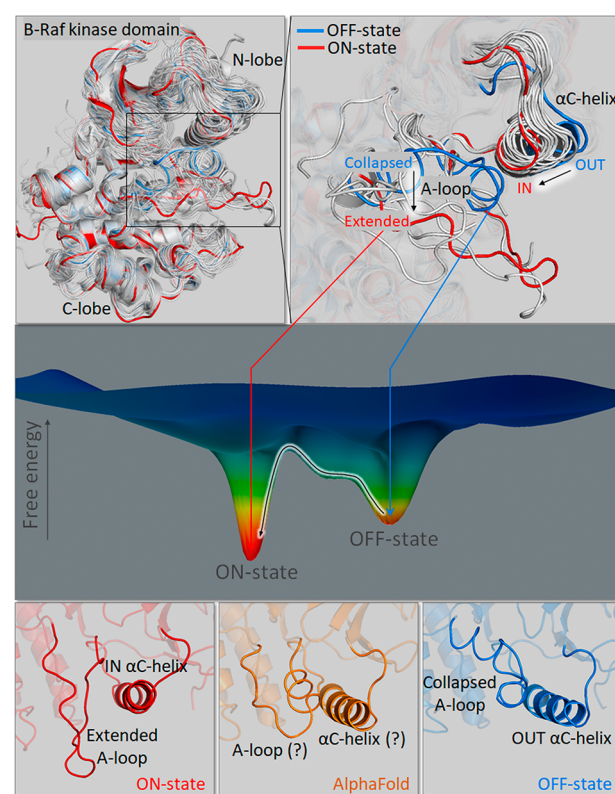

Figure 2. Structural ensembles for B-Raf activation. The snapshots for B-Raf kinase domains (*top panels*) are generated from the protein databank (PDB). The representative inactive OFF-state conformation (PDB ID: 3SKC) and active ON-state conformation (PDB ID: 6UAN) are highlighted in blue and red, respectively. The free energy landscape of B-Raf kinase domain depicting the population shift from OFF-state to ON-state upon activation (*middle panel*). Highlighted activation segments of αC-helix and A-loop representing the side by side comparisons between the single structure predicted by AlphaFold and the representative B-Raf conformations of both inactive OFF-state and active ON-state (*bottom panels*). The AlphaFold structure falls into neither the active ON-state nor the inactive OFF-state.

"fuzzy".[170,171] Events associated with their environments and functions, such as changes in pH, interactions with ions, water, and lipids, and binding of small or macromolecules, promote conformational changes. These are frustrated by their local restricted molecular environment.[172] The cooperative, accommodating structural changes shift the ensemble. The shifted, now populated states are frustrated by their current neighboring residues conformations. Binding and catalysis involve making and breaking covalent and noncovalent interactions at the interaction site. These propagate through frustration, influencing the conformational states of the ensemble. The shifts in the ensemble alter the relative stabilities, i.e., the populations of the states, thus influencing the allosteric transitions. Importantly, frustration does not create new conformations; instead, it alters the number of molecules populating it.[173] *Frustration is thus a powerful tool harnessed by evolution for function.*[174]

Biomolecules must be described statistically, not statically.[166,175] Static descriptions were the norm for decades. Yet, a static description cannot capture function. It cannot describe protein activation from the inactive to the active state upon some activation event, such as binding a hormone, or being covalently modified by a post-translational modification,

or the presence of oncogenic driver mutations. It is also unable to describe how high affinity binding to an activator shifts protein molecules to their active state.[176,177] It will further fail when attempting to describe how allosteric "rescue mutations"[178,179] work (albeit not other rescue mutations, e.g., refs 180−183), how allosteric drugs are able to block the active site, and how mutations countering them can be overcome. All these processes which take place in the cell would not have been possible had the protein existed in a single structure or was flipping between only two states, active and inactive. While there is a *single* conformation that the active enzyme should adopt for productive catalysis, there are multiple ways to inactivate it and thus many inactive states. The notion of a single structure bred the concept of the "lock-and-key" binding mechanism. This view was superseded by the "induced fit" mechanism which considered the presence of only two states, an active and an inactive state. In an induced fit scenario, the ligands bind to the single "open" protein structure and the interaction between a protein and a rigid binding partner induces a conformational change in the protein.[184] In contrast, the conformational selection mechanism[167−169,185] theorizes that the energy surface hosts a very large number of conformations, and the one that fits best is selected, with subsequent minor induced fit optimization, largely by side chains.

AlphaFold exploits AI to make template-free predictions of protein structures from their sequences, equipping biologists with structures with good resolution. The predictions that it yields, like those obtained by homology modeling, are rigid. Flexibility is implicitly captured by the absence of, or low confidence levels of predicted structure for certain regions, as in the case of intrinsically disordered proteins. Thus, computational methods once relegated to the periphery of biology, are now at the forefront, driving "the second molecular biology revolution". AlphaFold can drive breakthroughs in fundamental problems in the life sciences, including precision medicine, with promise to transform research and accelerate drug discovery. It is driven by deep learning innovations, which appear poised to transform MD simulations.

■ APPICATIONS OF ARTIFICIAL INTELLIGENCE AND MACHINE LANGUAGE

AI and Machine Language in Simulations. Machine learning for molecular simulations—tools, strategies, and principles—have been reviewed recently.[186] As can be seen from this excellent review, machine learning has already been making a significant impact on the development of approximate methods for complex atomic systems. The innovation in the development and integration of MD simulations with deep learning can reproduce, interpret, predict, and generate data relating to the behavior of biological macromolecules.[187−192] Deep learning methods can help MD simulations excel in their efficiency and scales, with AI bridging between deep learning technologies and simulations. Challenges toward broad usage include smooth connection of AI and MD and automation of workflows. These could popularize novel deep learning tools in MD simulations toward efficiently exploiting both powerful methods. The number of publications in this area has been skyrocketing, emphasizing the recognition of the potency of AI and machine learning in simulation. As an example, MD simulations need to perform extensive sampling of the conformational space that require long time scales. Deep

learning involving, e.g., variational autoencoders have been shown to be useful. The learned latent space in the variational autoencoders has been employed to generate unsampled protein conformations, and simulations starting from these conformations accelerated the sampling.[193] In another example, a deep learning framework with mixed classical and machine learning potentials (TorchMD) has been developed for molecular simulations.[188] The review cited above provides additional diverse examples. Deep learning has also been already exploited in structural modeling and design[60,194−196] and analysis[191] and linking these to function.[197]

AI and Machine Language in Prediction of Pathogen−Human Host PPIs. AI and deep learning are also being developed and applied to experimental determination and prediction of macromolecular structures,[198] as well as to PPIs.[199,200]

Applications of AI approaches to human−microbiome protein−protein interactions have also been reviewed recently.[66] These interactions play important roles in human health and disease. There is a rapid increase of data that microbes, bacteria, and viruses impact human health. They can modulate human signaling and immune response by interacting with the human proteins. To decipher this modulation, it is important to identify the specific interactions, the human host proteins that are involved, and the structure of the complex. Identification of the interactions along with their structural details at atomic resolution permit understanding the mechanisms involved in pathogen survival and assist in drug discovery targeting these interactions. The interactions help the pathogens to elude and bypass the immune defense, with the pathogens hijacking host signaling. Mechanistically, pathogen proteins can have surfaces which resemble those of the host, allowing them to mimic and compete with host protein interactions (Figure 3). They bind to the host protein and rewire its physiological signaling. Data, including structural details, are scarce and large-scale experimental detection is challenging. Efficient and robust computational strategies to predict the interactions is thus vital. We have developed an algorithm and server to predict these human host−microbial PPIs (HMIs) based on their protein structures, which can be experimental or modeled. In large scale applications, AlphaFold can now be used toward this aim. Machine learning permits both the large-scale efficient and generalizable application and addressing the complex dynamics of such relationships that the machine learning algorithms can decipher.

Challenges in machine learning for PPI prediction relate to both data and method. With limited microbial but not human data, microbial sample sizes are small. In sequence-based algorithms, the dimensionality problem can be pronounced, where the difficulty exponentially grows as the feature size increases. Principal component analysis (PCA), uniform manifold approximation and projection (UMAP), or autoencoders can be used to embed the samples into lower-dimensional spaces,[201,202] and preprocessing and postprocessing pipelines can be employed for other data. In structure-based methods the problems may relate to the quantity and diversity of the representation. Data relating to host-microbe PPIs with 3D structures are sparse, thus facing a problem in training and evaluating the computational methods. Additional problems involve lack of gold standard test data set. Evaluation metrics are also unclear, the PPI networks are sparse, and more.[66] DeepMind's AlphaFold2 success in sequence-based
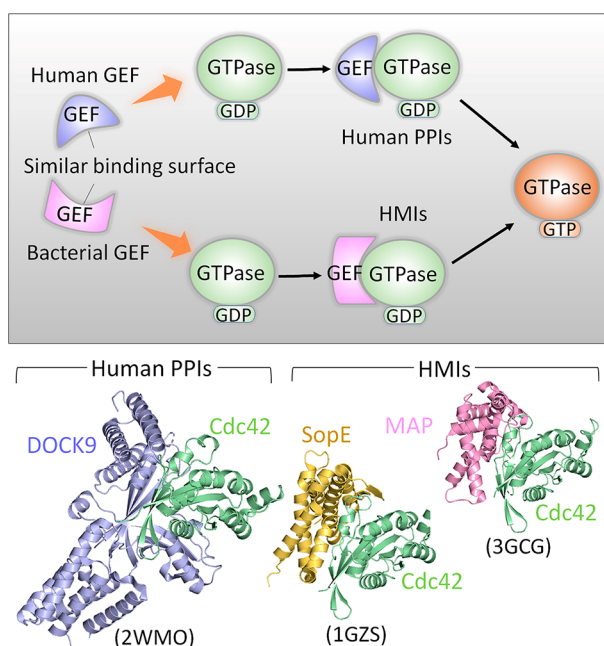
Figure 3. Human−microbiome PPIs promote GTPase activation. Human cell division control protein 42 homologue (Cdc42) is a small GTPase of the Rho family, involved in cell cycle. In human cells, it is activated by guanine-nucleotide exchange factors (GEFs), such as DOCK9 (PDB ID: 2WMO), by transforming the inactive GDP-bound to the active GTP-bound forms.[207] Bacterial secretes toxins or effectors mimicking the GEF proteins, such as SopE (PDB: 1GZS) from Salmonella[208] and MAP (PDB: 3GCG) from *Escherichia coli*,[209] can interact with Cdc42 and activate it. The interaction surfaces of these bacterial GEF mimicries resemble the host protein, allowing them to mimic and compete with the host protein interactions. PPIs, protein−protein interactions; HMIs, host−(or human−) microbiome interactions. Ongoing work incorporates AI into the HMI prediction algorithm. If the structures of the human or microbe are unable, AlphaFold can generate them.

protein structure prediction,[4] as well as the RoseTTAFold[70] open-source counterpart, and the publicly available Alpha-Fold2 prediction of all human proteins[203] are major steps that benefit the scientific community.

CONCLUSIONS

AI and machine learning are appending projects. They are applied in diverse applications, including biological networks.[204] They impact disease biology, drug discovery, microbiome research, and synthetic biology. They also evolved a machine learning pipeline for molecular complex detection in protein-interaction networks,[205] as well as the relevance of major signaling pathways in cancer survival.[46]

Here, we briefly overviewed the immense impact of AlphaFold, and of AI in structural biology, with some examples. We highlighted what AlphaFold can and cannot accomplish and why. Allosteric mechanisms fall into the latter category. Nevertheless, through MD simulations of models that AlphaFold produces, this aim can be accomplished as well. Still, however, even though simulations would address this dynamics problem, at such scales, the cost is prohibitive. A paradigm-shifting machine learning method is needed to model protein dynamics.

AlphaFold and its underlying deep learning innovations have opened up the next frontiers in protein science,[164] including precision medicine.[206] Protein structures connect to cell biology, chemistry, biophysics, and medicine. To date, over 180 000 protein structures are available, open to all researchers across the world in the PDB database. Still, structures of pathogens are not within them, and neither are many others, which are essential to human health. The resource is now there, and with growing computational power, eventually, these will be there as well. Nonetheless, the availability of the structures is insufficient. For us, as biophysicists, the key is what are the significant questions to ask. What should the focus of our research be, such that we do not repeat what was so well done but instead exploit the new capabilities to ask the really important questions.

■ AUTHOR INFORMATION

Corresponding Author

Ruth Nussinov − *Computational Structural Biology Section, Frederick National Laboratory for Cancer Research, Frederick, Maryland 21702, United States; Department of Human Molecular Genetics and Biochemistry, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel;* ⓘ orcid.org/0000-0002-8115-6415; Email: NussinoR@mail.nih.gov

Authors

Mingzhen Zhang − *Computational Structural Biology Section, Frederick National Laboratory for Cancer Research, Frederick, Maryland 21702, United States*

Yonglan Liu − *Cancer Innovation Laboratory, National Cancer Institute, Frederick, Maryland 21702, United States*

Hyunbum Jang − *Computational Structural Biology Section, Frederick National Laboratory for Cancer Research, Frederick, Maryland 21702, United States;* ⓘ orcid.org/0000-0001-9402-4051

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jpcb.2c04346

Notes

The authors declare no competing financial interest.

Biographies

Ruth Nussinov pioneered the "conformational selection and population shift" (1999) as an alternative to the "induced fit" textbook model to explain the molecular mechanism of recognition and posited that population shift underlies allosteric regulation. She extended this pre-existing ensemble model to catalysis (2000) and oncogenic activation, contributing to extraordinary advancements in understanding structure and function. Nussinov received her Ph.D. in 1977 from Rutgers University and did postdoctoral work in the Structural Chemistry Department of the Weizmann Institute. Subsequently, she was at the Chemistry Department at Berkeley, the Biochemistry Department at Harvard, and a visiting scientist at the NIH. In 1984, she joined the Department of Human Genetics, at the Medical School at Tel Aviv University. In 1985, she accepted a concurrent position at the National Cancer Institute of the NIH, Leidos Biomedical Research, where she is a Senior Principal Scientist and Principal Investigator heading the Computational Structural Biology Section at the NCI. She has authored over 700 scientific papers. She was the Editor-in-Chief of *PLoS Computational Biology*, is currently the Editor-in-Chief of *Current Opinion in Structural Biology*, and is Associate Editor and is on the Editorial Boards of a number of journals. She is a frequent speaker in Domestic and International meetings, symposia, and academic institutions, has won multiple

awards and has been elected fellow of several societies, and has been a highly cited researcher. Her National Cancer Institute website gives further details. https://ccr.cancer.gov/ruth-nussinov.

Mingzhen Zhang is a computational scientist in Computational Structural Biology Section at Frederick National Laboratory for Cancer Research in the Cancer Innovation Laboratory, National Cancer Institute (NCI), Frederick, MD. His research focuses on the activation mechanism of key oncogenic proteins in Ras signaling pathways. He uses computational tools to uncover the detailed high-resolution activation mechanisms of lipid kinase, PI3K, and protein kinase, Raf, the main downstream effectors of Ras, and to explore the structural principles for cancer drug discovery.

Yonglan Liu is a postdoctoral researcher in the Cancer Innovation Laboratory, National Cancer Institute, Frederick, MD. Her research interests include computational structural biology, biophysics, and biochemistry, specifically in cancer, neurodegenerative diseases, and biomaterials by using MD simulation and machine learning algorithms. Currently in NCI, she is studying the structural principles of signaling proteins, such as Bcr, c-Abl, and SHP2, and uncovering their mechanisms of protein–protein interaction, activation, and inhibition, for drug discovery.

Hyunbum Jang earned his Ph.D. in condensed matter physics studying thin magnetic films with the competing surface fields by performing a series of Monte Carlo (MC) simulations of the Heisenberg spin systems with anisotropy. In his early career, he studied membrane proteins using all-atom molecular dynamics (MD) simulations with the CHARMM and NAMD programs, focusing on a wide range of membrane systems from a membrane-bound small peptide to a large complex of ion channels in various bilayer systems. Currently, he is a senior scientist at Frederick National Laboratory for Cancer Research in the National Cancer Institute, Frederick, MD, working on the biological agents that are responsible for oncogenic signaling through the MAPK and PI3K/AKT pathways. He is a member of the American Chemical Society (ACS), American Association for Cancer Research (AACR), and the Biophysical Society (BPS).

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Hassoun, S.; Jefferson, F.; Shi, X.; Stucky, B.; Wang, J.; Rosa, E. Artificial intelligence for biology. *Integr Comp Biol.* **2022**, *61* (6), 2267–2275.

(2) Xu, Y.; Liu, X.; Cao, X.; Huang, C.; Liu, E.; Qian, S.; Liu, X.; Wu, Y.; Dong, F.; Qiu, C. W.; et al. Artificial intelligence: A powerful paradigm for scientific research. *Innovation* **2021**, *2* (4), 100179.

(3) Callaway, E. 'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures. *Nature* **2020**, *588* (7837), 203–204.

(4) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.;

Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589.

(5) Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; et al. AlphaFold protein structure database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **2022**, *50* (D1), D439–D444.

(6) Perrakis, A.; Sixma, T. K. AI revolutions in biology: The joys and perils of AlphaFold. *EMBO Rep* **2021**, *22* (11), No. e54046.

(7) Bouatta, N.; Sorger, P.; AlQuraishi, M. Protein structure prediction by AlphaFold2: Are attention and symmetries all you need? *Acta Crystallogr. D Struct Biol.* **2021**, *77* (8), 982–991.

(8) Saldano, T.; Escobedo, N.; Marchetti, J.; Zea, D. J.; Mac Donagh, J.; Velez Rueda, A. J.; Gonik, E.; Garcia Melani, A.; Novomisky Nechcoff, J.; Salas, M. N.; et al. Impact of protein conformational diversity on AlphaFold predictions. *Bioinformatics* **2022**, *38* (10), 2742–2748.

(9) Fowler, N. J.; Williamson, M. P. The accuracy of protein structures in solution determined by AlphaFold and NMR. *Structure* **2022**, *30*, 925.

(10) Binder, J. L.; Berendzen, J.; Stevens, A. O.; He, Y.; Wang, J.; Dokholyan, N. V.; Oprea, T. I. AlphaFold illuminates half of the dark human proteins. *Curr. Opin Struct Biol.* **2022**, *74*, 102372.

(11) Buel, G. R.; Walters, K. J. Can. AlphaFold2 predict the impact of missense mutations on structure? *Nat. Struct Mol. Biol.* **2022**, *29* (1), 1–2.

(12) Strodel, B. Energy landscapes of protein aggregation and conformation switching in intrinsically disordered proteins. *J. Mol. Biol.* **2021**, *433* (20), 167182.

(13) Outeiral, C.; Nissley, D. A.; Deane, C. M. Current structure predictors are not learning the physics of protein folding. *Bioinformatics* **2022**, *38*, 1881.

(14) *AlphaFold protein structure database*; https://alphafold.ebi.ac.uk/ (accessed July 22, 2021).

(15) *Open source, AlphaFold protein structure database*; https://www.deepmind.com/open-source/alphafold-protein-structure-database (accessed July 22, 2021).

(16) Brookes, E.; Rocco, M. A database of calculated solution parameters for the AlphaFold predicted protein structures. *Sci. Rep* **2022**, *12* (1), 7349.

(17) Simpkin, A. J.; Thomas, J. M. H.; Keegan, R. M.; Rigden, D. J. MrParse: finding homologues in the PDB and the EBI AlphaFold database for molecular replacement and more. *Acta Crystallogr. D Struct Biol.* **2022**, *78* (5), 553–559.

(18) Lomize, A. L.; Schnitzer, K. A.; Todd, S. C.; Cherepanov, S.; Outeiral, C.; Deane, C. M.; Pogozheva, I. D. Membranome 3.0: Database of single-pass membrane proteins with AlphaFold models. *Protein Sci.* **2022**, *31* (5), No. e4318.

(19) McGreig, J. E.; Uri, H.; Antczak, M.; Sternberg, M. J. E.; Michaelis, M.; Wass, M. N. 3DLigandSite: Structure-based prediction of protein-ligand binding sites. *Nucleic Acids Res.* **2022**, *50*, W13–W20.

(20) Aderinwale, T.; Bharadwaj, V.; Christoffer, C.; Terashi, G.; Zhang, Z.; Jahandideh, R.; Kagaya, Y.; Kihara, D. Real-time structure search and structure classification for AlphaFold protein models. *Commun. Biol.* **2022**, *5* (1), 316.

(21) Chen, A. Y.; Lee, J.; Damjanovic, A.; Brooks, B. R. Protein pKa prediction by tree-based machine learning. *J. Chem. Theory Comput* **2022**, *18* (4), 2673–2686.

(22) Porta-Pardo, E.; Ruiz-Serra, V.; Valentini, S.; Valencia, A. The structural coverage of the human proteome before and after AlphaFold. *PLoS Comput. Biol.* **2022**, *18* (1), No. e1009818.

(23) Ramalli, S. G.; Miles, A. J.; Janes, R. W.; Wallace, B. A. The PCDDB (protein circular dichroism data bank): A bioinformatics resource for protein characterisations and methods development. *J. Mol. Biol.* **2022**, *434*, 167441.

(24) Rossi Sebastiano, M.; Ermondi, G.; Hadano, S.; Caron, G. AI-based protein structure databases have the potential to accelerate rare

diseases research: AlphaFoldDB and the case of IAHSP/Alsin. *Drug Discov Today* **2022**, *27* (6), 1652−1660.

(25) Wheeler, R. J. A resource for improved predictions of Trypanosoma and Leishmania protein three-dimensional structure. *PLoS One* **2021**, *16* (11), No. e0259871.

(26) Nussinov, R.; Jang, H.; Tsai, C. J.; Cheng, F. Precision medicine review: Rare driver mutations and their biophysical classification. *Biophys Rev.* **2019**, *11* (1), 5−19.

(27) Nussinov, R.; Tsai, C. J. 'Latent drivers' expand the cancer mutational landscape. *Curr. Opin Struct Biol.* **2015**, *32*, 25−32.

(28) Brown, A. L.; Li, M.; Goncearenco, A.; Panchenko, A. R. Finding driver mutations in cancer: Elucidating the role of background mutational processes. *PLoS Comput. Biol.* **2019**, *15* (4), No. e1006981.

(29) Nussinov, R.; Jang, H.; Tsai, C. J.; Cheng, F. Review: Precision medicine and driver mutations: Computational methods, functional assays and conformational principles for interpreting cancer drivers. *PLoS Comput. Biol.* **2019**, *15* (3), No. e1006658.

(30) Nussinov, R.; Tsai, C. J.; Jang, H. Why are some driver mutations rare? *Trends Pharmacol. Sci.* **2019**, *40* (12), 919−929.

(31) Martinez-Jimenez, F.; Muinos, F.; Sentis, I.; Deu-Pons, J.; Reyes-Salazar, I.; Arnedo-Pac, C.; Mularoni, L.; Pich, O.; Bonet, J.; Kranas, H.; et al. A compendium of mutational cancer driver genes. *Nat. Rev. Cancer* **2020**, *20* (10), 555−572.

(32) Pon, J. R.; Marra, M. A. Driver and passenger mutations in cancer. *Annu. Rev. Pathol* **2015**, *10*, 25−50.

(33) Wang, H.; Wang, T.; Zhao, X.; Wu, H.; You, M.; Sun, Z.; Mao, F. AI-Driver: An ensemble method for identifying driver mutations in personal cancer genomes. *NAR Genom Bioinform* **2020**, *2* (4), lqaa084.

(34) Nussinov, R.; Tsai, C. J.; Jang, H. Autoinhibition can identify rare driver mutations and advise pharmacology. *FASEB J.* **2020**, *34* (1), 16−29.

(35) Park, H. M.; Park, Y.; Vankerschaver, J.; Van Messem, A.; De Neve, W.; Shim, H. Rethinking protein drug design with highly accurate structure prediction of anti-CRISPR proteins. *Pharmaceuticals (Basel)* **2022**, *15* (3), 310.

(36) Badaczewska-Dawid, A. E.; Garcia-Pardo, J.; Kuriata, A.; Pujols, J.; Ventura, S.; Kmiecik, S. A3D database: Structure-based predictions of protein aggregation for the human proteome. *Bioinformatics* **2022**, *38*, 3121.

(37) Song, B.; Luo, X.; Luo, X.; Liu, Y.; Niu, Z.; Zeng, X. Learning spatial structures of proteins improves protein-protein interaction prediction. *Brief Bioinform* **2022**, *23* (2), bbab558.

(38) Diwan, G. D.; Gonzalez-Sanchez, J. C.; Apic, G.; Russell, R. B. Next generation protein structure predictions and genetic variant interpretation. *J. Mol. Biol.* **2021**, *433* (20), 167180.

(39) AlQuraishi, M. AlphaFold at CASP13. *Bioinformatics* **2019**, *35* (22), 4862−4865.

(40) LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521* (7553), 436−444.

(41) Nussinov, R.; Tsai, C. J.; Jang, H. Allostery, and how to define and measure signal transduction. *Biophys Chem.* **2022**, *283*, 106766.

(42) CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences *Deep learning on cell signaling networks establishes AI for single-cell biology*; https://phys.org/news/2020-08-deep-cell-networks-ai-single-cell.html (accessed Aug 4, 2020).

(43) Fortelny, N.; Bock, C. Knowledge-primed neural networks enable biologically interpretable deep learning on single-cell sequencing data. *Genome Biol.* **2020**, *21* (1), 190.

(44) Bostan, B.; Greiner, R.; Szafron, D.; Lu, P. Predicting homologous signaling pathways using machine learning. *Bioinformatics* **2009**, *25* (22), 2913−2920.

(45) Henriques, D.; Villaverde, A. F.; Rocha, M.; Saez-Rodriguez, J.; Banga, J. R. Data-driven reverse engineering of signaling pathways using ensembles of dynamic models. *PLoS Comput. Biol.* **2017**, *13* (2), No. e1005379.

(46) Feng, J.; Zhang, H.; Li, F. Investigating the relevance of major signaling pathways in cancer survival using a biologically meaningful deep learning model. *BMC Bioinformatics* **2021**, *22* (1), 47.

(47) Neumann, J.; Schwierz, N. Artificial intelligence resolves kinetic pathways of magnesium binding to RNA. *J. Chem. Theory Comput* **2022**, *18* (2), 1202−1212.

(48) Hindley, J. W.; Zheleva, D. G.; Elani, Y.; Charalambous, K.; Barter, L. M. C.; Booth, P. J.; Bevan, C. L.; Law, R. V.; Ces, O. Building a synthetic mechanosensitive signaling pathway in compartmentalized artificial cells. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (34), 16711−16716.

(49) Zhang, H.; Chen, Y.; Li, F. Predicting anticancer drug response with deep learning constrained by signaling pathways. *Front Bioinform* **2021**, DOI: 10.3389/fbinf.2021.639349.

(50) Englander, S. W.; Mayne, L. The nature of protein folding pathways. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (45), 15873−15880.

(51) Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R. The protein folding problem. *Annu. Rev. Biophys* **2008**, *37*, 289−316.

(52) Dill, K. A.; MacCallum, J. L. The protein-folding problem, 50 years on. *Science* **2012**, *338* (6110), 1042−1046.

(53) Baker, D.; Sali, A. Protein structure prediction and structural genomics. *Science* **2001**, *294* (5540), 93−96.

(54) Fersht, A. R. Nucleation mechanisms in protein folding. *Curr. Opin Struct Biol.* **1997**, *7* (1), 3−9.

(55) Dobson, C. M. Protein folding and misfolding. *Nature* **2003**, *426* (6968), 884−890.

(56) Mirny, L.; Shakhnovich, E. Protein folding theory: from lattice to all-atom models. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 361−396.

(57) Shea, J. E.; Brooks, C. L., 3rd. From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annu. Rev. Phys. Chem.* **2001**, *52*, 499−535.

(58) Kuhlman, B.; Bradley, P. Advances in protein structure prediction and design. *Nat. Rev. Mol. Cell Biol.* **2019**, *20* (11), 681−697.

(59) Deng, H.; Jia, Y.; Zhang, Y. Protein structure prediction. *Int. J. Mod. Phys. B* **2018**, *32* (18), 1840009.

(60) Gao, W.; Mahajan, S. P.; Sulam, J.; Gray, J. J. Deep learning in protein structural modeling and design. *Patterns (N Y)* **2020**, *1* (9), 100142.

(61) Wu, F.; Xu, J. Deep template-based protein structure prediction. *PLoS Comput. Biol.* **2021**, *17* (5), No. e1008954.

(62) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46* (W1), W296−W303.

(63) Khor, B. Y.; Tye, G. J.; Lim, T. S.; Choong, Y. S. General overview on structure prediction of twilight-zone proteins. *Theor Biol. Med. Model* **2015**, *12*, 15.

(64) Yang, J.; Anishchenko, I.; Park, H.; Peng, Z.; Ovchinnikov, S.; Baker, D. Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (3), 1496−1503.

(65) Bhattacharya, S.; Roche, R.; Shuvo, M. H.; Bhattacharya, D. Recent advances in protein homology detection propelled by inter-residue interaction map threading. *Front Mol. Biosci* **2021**, *8*, 643752.

(66) Lim, H.; Cankara, F.; Tsai, C. J.; Keskin, O.; Nussinov, R.; Gursoy, A. Artificial intelligence approaches to human-microbiome protein-protein interactions. *Curr. Opin Struct Biol.* **2022**, *73*, 102328.

(67) Moreno-Indias, I.; Lahti, L.; Nedyalkova, M.; Elbere, I.; Roshchupkin, G.; Adilovic, M.; Aydemir, O.; Bakir-Gungor, B.; Santa Pau, E. C.; D'Elia, D.; et al. Statistical and machine learning techniques in human microbiome studies: Contemporary challenges and solutions. *Front Microbiol* **2021**, *12*, 635781.

(68) Goncearenco, A.; Li, M.; Simonetti, F. L.; Shoemaker, B. A.; Panchenko, A. R. Exploring protein-protein interactions as drug

targets for anti-cancer therapy with in silico workflows. *Methods Mol. Biol.* **2017**, *1647*, 221−236.

(69) Chang, C. K.; Lin, S. M.; Satange, R.; Lin, S. C.; Sun, S. C.; Wu, H. Y.; Kehn-Hall, K.; Hou, M. H. Targeting protein-protein interaction interfaces in COVID-19 drug discovery. *Comput. Struct Biotechnol J.* **2021**, *19*, 2246−2255.

(70) Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R. D.; et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373* (6557), 871−876.

(71) Pittala, S.; Bailey-Kellogg, C. Learning context-aware structural representations to predict antigen and antibody binding interfaces. *Bioinformatics* **2020**, *36* (13), 3996−4003.

(72) Dai, B.; Bailey-Kellogg, C. Protein interaction interface region prediction by geometric deep learning. *Bioinformatics* **2021**, *37*, 2580.

(73) Cao, Y.; Shen, Y. Energy-based graph convolutional networks for scoring protein docking models. *Proteins* **2020**, *88* (8), 1091−1099.

(74) Lim, J.; Ryu, S.; Park, K.; Choe, Y. J.; Ham, J.; Kim, W. Y. Predicting drug-target interaction using a novel graph neural network with 3D structure-embedded graph representation. *J. Chem. Inf Model* **2019**, *59* (9), 3981−3988.

(75) Wang, X.; Terashi, G.; Christoffer, C. W.; Zhu, M.; Kihara, D. Protein docking model evaluation by 3D deep convolutional neural networks. *Bioinformatics* **2020**, *36* (7), 2113−2118.

(76) Guven-Maiorov, E.; Hakouz, A.; Valjevac, S.; Keskin, O.; Tsai, C. J.; Gursoy, A.; Nussinov, R. HMI-PRED: A web server for structural prediction of host-microbe interactions based on interface mimicry. *J. Mol. Biol.* **2020**, *432* (11), 3395−3403.

(77) Guven-Maiorov, E.; Tsai, C. J.; Nussinov, R. Oncoviruses can drive cancer by rewiring signaling pathways through interface mimicry. *Front Oncol* **2019**, *9*, 1236.

(78) Guven-Maiorov, E.; Tsai, C. J.; Ma, B.; Nussinov, R. Interface-based structural prediction of novel host-pathogen interactions. *Methods Mol. Biol.* **2019**, *1851*, 317−335.

(79) Guven-Maiorov, E.; Tsai, C. J.; Ma, B.; Nussinov, R. Prediction of host-pathogen interactions for Helicobacter pylori by interface mimicry and implications to gastric cancer. *J. Mol. Biol.* **2017**, *429* (24), 3925−3941.

(80) Guven-Maiorov, E.; Tsai, C. J.; Nussinov, R. Structural host-microbiota interaction networks. *PLoS Comput. Biol.* **2017**, *13* (10), No. e1005579.

(81) Pushpakom, S.; Iorio, F.; Eyers, P. A.; Escott, K. J.; Hopper, S.; Wells, A.; Doig, A.; Guilliams, T.; Latimer, J.; McNamee, C.; et al. Drug repurposing: progress, challenges and recommendations. *Nat. Rev. Drug Discov* **2019**, *18* (1), 41−58.

(82) Jarada, T. N.; Rokne, J. G.; Alhajj, R. A review of computational drug repositioning: strategies, approaches, opportunities, challenges, and directions. *J. Cheminform* **2020**, *12* (1), 46.

(83) Pulley, J. M.; Rhoads, J. P.; Jerome, R. N.; Challa, A. P.; Erreger, K. B.; Joly, M. M.; Lavieri, R. R.; Perry, K. E.; Zaleski, N. M.; Shirey-Rice, J. K.; et al. Using what we already have: Uncovering new drug repurposing strategies in existing omics data. *Annu. Rev. Pharmacol Toxicol* **2020**, *60*, 333−352.

(84) Zhou, Y.; Wang, F.; Tang, J.; Nussinov, R.; Cheng, F. Artificial intelligence in COVID-19 drug repurposing. *Lancet Digit Health* **2020**, *2* (12), e667−e676.

(85) Sahoo, B. M.; Ravi Kumar, B. V. V.; Sruti, J.; Mahapatra, M. K.; Banik, B. K.; Borah, P. Drug repurposing strategy (DRS): Emerging approach to identify potential therapeutics for treatment of novel coronavirus infection. *Front Mol. Biosci* **2021**, *8*, 628144.

(86) Fang, J.; Pieper, A. A.; Nussinov, R.; Lee, G.; Bekris, L.; Leverenz, J. B.; Cummings, J.; Cheng, F. Harnessing endophenotypes and network medicine for Alzheimer's drug repurposing. *Med. Res. Rev.* **2020**, *40* (6), 2386−2426.

(87) Begley, C. G.; Ashton, M.; Baell, J.; Bettess, M.; Brown, M. P.; Carter, B.; Charman, W. N.; Davis, C.; Fisher, S.; Frazer, I.; et al. Drug

repurposing: Misconceptions, challenges, and opportunities for academic researchers. *Sci. Transl Med.* **2021**, *13* (612), No. eabd5524.

(88) Zeng, X.; Zhu, S.; Hou, Y.; Zhang, P.; Li, L.; Li, J.; Huang, L. F.; Lewis, S. J.; Nussinov, R.; Cheng, F. Network-based prediction of drug-target interactions using an arbitrary-order proximity embedded deep forest. *Bioinformatics* **2020**, *36* (9), 2805−2812.

(89) Roessler, H. I.; Knoers, N.; van Haelst, M. M.; van Haaften, G. Drug repurposing for rare diseases. *Trends Pharmacol. Sci.* **2021**, *42* (4), 255−267.

(90) Zeng, X.; Zhu, S.; Lu, W.; Liu, Z.; Huang, J.; Zhou, Y.; Fang, J.; Huang, Y.; Guo, H.; Li, L.; et al. Target identification among known drugs by deep learning from heterogeneous networks. *Chem. Sci.* **2020**, *11* (7), 1775−1797.

(91) Zeng, X.; Zhu, S.; Liu, X.; Zhou, Y.; Nussinov, R.; Cheng, F. DeepDR: A network-based deep learning approach to in silico drug repositioning. *Bioinformatics* **2019**, *35* (24), 5191−5198.

(92) Webb, S. Deep learning for biology. *Nature* **2018**, *554* (7693), 555−557.

(93) Nature Editorial Staff. Artificial intelligence in structural biology is here to stay. *Nature* **2021**, *595* (7869), 625−626.

(94) Xu, C.; Jackson, S. A. Machine learning and complex biological data. *Genome Biol.* **2019**, *20* (1), 76.

(95) Emmert-Streib, F. Grand challenges for artificial intelligence in molecular medicine. *Front Mol. Med.* **2021**, DOI: 10.3389/fmmed.2021.734659.

(96) He, J.; Baxter, S. L.; Xu, J.; Xu, J.; Zhou, X.; Zhang, K. The practical implementation of artificial intelligence technologies in medicine. *Nat. Med.* **2019**, *25* (1), 30−36.

(97) Zou, J.; Huss, M.; Abid, A.; Mohammadi, P.; Torkamani, A.; Telenti, A. A primer on deep learning in genomics. *Nat. Genet.* **2019**, *51* (1), 12−18.

(98) Zeng, T.; Sun, S. Y.; Wang, Y.; Zhu, H.; Chen, L. Network biomarkers reveal dysfunctional gene regulations during disease progression. *FEBS J.* **2013**, *280* (22), 5682−5695.

(99) Martin, W.; Sheynkman, G.; Lightstone, F. C.; Nussinov, R.; Cheng, F. Interpretable artificial intelligence and exascale molecular dynamics simulations to reveal kinetics: Applications to Alzheimer's disease. *Curr. Opin Struct Biol.* **2022**, *72*, 103−113.

(100) Nussinov, R.; Jang, H.; Nir, G.; Tsai, C. J.; Cheng, F. A new precision medicine initiative at the dawn of exascale computing. *Signal Transduct Target Ther* **2021**, *6* (1), 3.

(101) Biddle, J. W.; Martinez-Corral, R.; Wong, F.; Gunawardena, J. Allosteric conformational ensembles have unlimited capacity for integrating information. *Elife* **2021**, *10*, e65498.

(102) Campitelli, P.; Modi, T.; Kumar, S.; Ozkan, S. B. The role of conformational dynamics and allostery in modulating protein evolution. *Annu. Rev. Biophys* **2020**, *49*, 267−288.

(103) Guo, J.; Zhou, H. X. Protein allostery and conformational dynamics. *Chem. Rev.* **2016**, *116* (11), 6503−6515.

(104) Nussinov, R.; Tsai, C. J. Allostery without a conformational change? Revisiting the paradigm. *Curr. Opin Struct Biol.* **2015**, *30*, 17−24.

(105) 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* (4), 566−578.

(106) Liu, J.; Nussinov, R. Allostery: An overview of Its history, concepts, methods, and applications. *PLoS Comput. Biol.* **2016**, *12* (6), No. e1004966.

(107) Verkhivker, G. M.; Agajanian, S.; Hu, G.; Tao, P. Allosteric regulation at the crossroads of new technologies: Multiscale modeling, networks, and machine learning. *Front Mol. Biosci* **2020**, *7*, 136.

(108) Di Cera, E. Mechanisms of ligand binding. *Biophys Rev.* **2020**, *1* (1), No. 011303.

(109) Wang, J.; Jain, A.; McDonald, L. R.; Gambogi, C.; Lee, A. L.; Dokholyan, N. V. Mapping allosteric communications within individual proteins. *Nat. Commun.* **2020**, *11* (1), 3862.

(110) Lopez, A.; Dahiya, V.; Delhommel, F.; Freiburger, L.; Stehle, R.; Asami, S.; Rutz, D.; Blair, L.; Buchner, J.; Sattler, M. Client

binding shifts the populations of dynamic Hsp90 conformations through an allosteric network. *Sci. Adv.* **2021**, 7 (51), No. eabl7295.

(111) Greener, J. G.; Sternberg, M. J. Structure-based prediction of protein allostery. *Curr. Opin Struct Biol.* **2018**, *50*, 1−8.

(112) Guarnera, E.; Berezovsky, I. N. Allosteric sites: remote control in regulation of protein activity. *Curr. Opin Struct Biol.* **2016**, *37*, 1−8.

(113) Morgan, B. R.; Massi, F. A computational study of RNA binding and specificity in the tandem zinc finger domain of TIS11d. *Protein Sci.* **2010**, *19* (6), 1222−1234.

(114) Akimoto, M.; Martinez Pomier, K.; VanSchouwen, B.; Byun, J. A.; Khamina, M.; Melacini, G. Allosteric pluripotency: challenges and opportunities. *Biochem. J.* **2022**, *479* (7), 825−838.

(115) Tee, W. V.; Tan, Z. W.; Guarnera, E.; Berezovsky, I. N. Conservation and diversity in allosteric fingerprints of proteins for evolutionary-inspired engineering and design. *J. Mol. Biol.* **2022**, 167577.

(116) Tee, W. V.; Tan, Z. W.; Lee, K.; Guarnera, E.; Berezovsky, I. N. Exploring the allosteric territory of protein function. *J. Phys. Chem. B* **2021**, *125* (15), 3763−3780.

(117) Guarnera, E.; Berezovsky, I. N. Allosteric drugs and mutations: chances, challenges, and necessity. *Curr. Opin Struct Biol.* **2020**, *62*, 149−157.

(118) Nussinov, R.; Ma, B.; Tsai, C. J.; Csermely, P. Allosteric conformational barcodes direct signaling in the cell. *Structure* **2013**, *21* (9), 1509−1521.

(119) Nussinov, R.; Tsai, C. J.; Liu, J. Principles of allosteric interactions in cell signaling. *J. Am. Chem. Soc.* **2014**, *136* (51), 17692−17701.

(120) Nussinov, R.; Tsai, C. J. The design of covalent allosteric drugs. *Annu. Rev. Pharmacol Toxicol* **2015**, *55*, 249−267.

(121) Nussinov, R.; Zhang, M.; Maloney, R.; Liu, Y.; Tsai, C. J.; Jang, H. Allostery: Allosteric cancer drivers and innovative allosteric drugs. *J. Mol. Biol.* **2022**, 167569.

(122) de Vries, R.; Meijer, F. A.; Doveston, R. G.; Leijten-van de Gevel, I. A.; Brunsveld, L. Cooperativity between the orthosteric and allosteric ligand binding sites of RORgammat. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118* (6), e2021287118.

(123) Chen, X.; Ali, Y. I.; Fisher, C. E.; Arribas-Bosacoma, R.; Rajasekaran, M. B.; Williams, G.; Walker, S.; Booth, J. R.; Hudson, J. J.; Roe, S. M.; et al. Uncovering an allosteric mode of action for a selective inhibitor of human Bloom syndrome protein. *Elife* **2021**, *10*, e65339.

(124) Mingione, V. R.; Foda, Z.; Paung, Y.; Philipose, H.; Rangwala, A. M.; Shan, Y.; Seeliger, M. A. Validation of an allosteric binding site of Src kinase identified by unbiased ligand binding simulations. *J. Mol. Biol.* **2022**, 167628.

(125) Li, L.; Meyer, C.; Zhou, Z. W.; Elmezayen, A.; Westover, K. Therapeutic targeting the allosteric cysteinome of RAS and kinase families. *J. Mol. Biol.* **2022**, 167626.

(126) Nussinov, R.; Tsai, C. J. Unraveling structural mechanisms of allosteric drug action. *Trends Pharmacol. Sci.* **2014**, *35* (5), 256−264.

(127) Abyzov, A.; Blackledge, M.; Zweckstetter, M. Conformational dynamics of intrinsically disordered proteins regulate biomolecular condensate chemistry. *Chem. Rev.* **2022**, *122* (6), 6719−6748.

(128) Gopal, S. M.; Wingbermuhle, S.; Schnatwinkel, J.; Juber, S.; Herrmann, C.; Schafer, L. V. Conformational preferences of an intrinsically disordered protein domain: A case study for modern force fields. *J. Phys. Chem. B* **2021**, *125* (1), 24−35.

(129) Santambrogio, C.; Natalello, A.; Brocca, S.; Ponzini, E.; Grandori, R. Conformational characterization and classification of intrinsically disordered proteins by native mass spectrometry and charge-state distribution analysis. *Proteomics* **2019**, *19* (6), No. 1800060.

(130) Jin, F.; Grater, F. How multisite phosphorylation impacts the conformations of intrinsically disordered proteins. *PLoS Comput. Biol.* **2021**, *17* (5), No. e1008939.

(131) Delaforge, E.; Cordeiro, T. N.; Bernadó, P.; Sibille, N. Conformational characterization of intrinsically disordered proteins

and its biological significance. In *Modern Magnetic Resonance*; Springer: Cham, 2018; p 381; .

(132) Bondos, S. E.; Dunker, A. K.; Uversky, V. N. Intrinsically disordered proteins play diverse roles in cell signaling. *Cell Commun. Signal* **2022**, *20* (1), 20.

(133) Phillips, A. H.; Kriwacki, R. W. Intrinsic protein disorder and protein modifications in the processing of biological signals. *Curr. Opin Struct Biol.* **2020**, *60*, 1−6.

(134) Ruan, H.; Sun, Q.; Zhang, W.; Liu, Y.; Lai, L. Targeting intrinsically disordered proteins at the edge of chaos. *Drug Discov Today* **2019**, *24* (1), 217−227.

(135) Schuler, B.; Borgia, A.; Borgia, M. B.; Heidarsson, P. O.; Holmstrom, E. D.; Nettels, D.; Sottini, A. Binding without folding - the biomolecular function of disordered polyelectrolyte complexes. *Curr. Opin Struct Biol.* **2020**, *60*, 66−76.

(136) Gupta, A.; Dey, S.; Zhou, H.-X. Artificial intelligence guided conformational mining of intrinsically disordered proteins. *bioRxiv* 2021; DOI: 10.1101/2021.1111.1121.469457.

(137) Ramanathan, A.; Ma, H.; Parvatikar, A.; Chennubhotla, S. C. Artificial intelligence techniques for integrative structural biology of intrinsically disordered proteins. *Curr. Opin Struct Biol.* **2021**, *66*, 216−224.

(138) Zhao, B.; Kurgan, L. Deep learning in prediction of intrinsic disorder in proteins. *Comput. Struct Biotechnol J.* **2022**, *20*, 1286−1294.

(139) Chen, J.; Stites, W. E. Packing is a key selection factor in the evolution of protein hydrophobic cores. *Biochemistry* **2001**, *40* (50), 15280−15289.

(140) Haber, E.; Anfinsen, C. B. Side-chain interactions governing the pairing of half-cystine residues in ribonuclease. *J. Biol. Chem.* **1962**, *237*, 1839−1844.

(141) Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **1973**, *181* (4096), 223−230.

(142) Anson, M. L.; Mirsky, A. E. Protein coagulation and Its reversal: The preparation of insoluble globin, soluble globin and heme. *J. Gen Physiol* **1930**, *13* (4), 469−476.

(143) Weingand-Ziade, A.; Ribes, F.; Renault, F.; Masson, P. Pressure- and heat-induced inactivation of butyrylcholinesterase: evidence for multiple intermediates and the remnant inactivation process. *Biochem. J.* **2001**, *356* (2), 487−493.

(144) Anfinsen, C. B.; Haber, E.; Sela, M.; White, F. H., Jr. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. U. S. A.* **1961**, *47*, 1309−1314.

(145) Anfinsen, C. B.; Scheraga, H. A. Experimental and theoretical aspects of protein folding. *Adv. Protein Chem.* **1975**, *29*, 205−300.

(146) Lazaridis, T.; Karplus, M. Effective energy functions for protein structure prediction. *Curr. Opin Struct Biol.* **2000**, *10* (2), 139−145.

(147) Karplus, M. The Levinthal paradox: yesterday and today. *Fold Des* **1997**, *2* (4), S69−75.

(148) Levitt, M.; Warshel, A. Computer simulation of protein folding. *Nature* **1975**, *253* (5494), 694−698.

(149) Bujnicki, J. M. Protein-structure prediction by recombination of fragments. *Chembiochem* **2006**, *7* (1), 19−27.

(150) Leaver-Fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Thompson, J.; Jacak, R.; Kaufman, K.; Renfrew, P. D.; Smith, C. A.; Sheffler, W.; et al. ROSETTA3: An object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol* **2011**, *487*, 545−574.

(151) Yang, J. S.; Chen, W. W.; Skolnick, J.; Shakhnovich, E. I. All-atom ab initio folding of a diverse set of proteins. *Structure* **2007**, *15* (1), 53−63.

(152) Levinthal, C. How to fold graciously. In *Mossbauer spectroscopy in biological systems*; University of Illinois Press: 1969.

(153) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **1995**, *21* (3), 167−195.

(154) Braitbard, M.; Schneidman-Duhovny, D.; Kalisman, N. Integrative structure modeling: Overview and assessment. *Annu. Rev. Biochem.* **2019**, *88*, 113−135.

(155) Bantegnie, B. Commentary: Integrative modeling and the role of neural constraints. *Front Psychol* **2017**, *8*, 1531.

(156) Ozdemir, E. S.; Nussinov, R.; Gursoy, A.; Keskin, O. Developments in integrative modeling with dynamical interfaces. *Curr. Opin Struct Biol.* **2019**, *56*, 11−17.

(157) AlQuraishi, M. Machine learning in protein structure prediction. *Curr. Opin Chem. Biol.* **2021**, *65*, 1−8.

(158) Noe, F.; De Fabritiis, G.; Clementi, C. Machine learning for protein folding and dynamics. *Curr. Opin Struct Biol.* **2020**, *60*, 77−84.

(159) Mardt, A.; Pasquali, L.; Wu, H.; Noe, F. VAMPnets for deep learning of molecular kinetics. *Nat. Commun.* **2018**, *9* (1), 5.

(160) Ribeiro, J. M. L.; Bravo, P.; Wang, Y.; Tiwary, P. Reweighted autoencoded variational Bayes for enhanced sampling (RAVE). *J. Chem. Phys.* **2018**, *149* (7), No. 072301.

(161) Bhowmik, D.; Gao, S.; Young, M. T.; Ramanathan, A. Deep clustering of protein folding simulations. *BMC Bioinformatics* **2018**, *19* (S18), 484.

(162) Tsai, S. T.; Kuo, E. J.; Tiwary, P. Learning molecular dynamics with simple language model built upon long short-term memory neural network. *Nat. Commun.* **2020**, *11* (1), 5115.

(163) Smith, Z.; Ravindra, P.; Wang, Y.; Cooley, R.; Tiwary, P. Discovering protein conformational flexibility through artificial-intelligence-aided molecular dynamics. *J. Phys. Chem. B* **2020**, *124* (38), 8221−8229.

(164) Miller, J. L. Deep learning opens up protein science's next frontiers. *Physics Today* **2021**, *74*, 14.

(165) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **1991**, *254* (5038), 1598−1603.

(166) Nussinov, R.; Wolynes, P. G. A second molecular biology revolution? The energy landscapes of biomolecular function. *Phys. Chem. Chem. Phys.* **2014**, *16* (14), 6321−6322.

(167) Tsai, C. J.; Ma, B.; Nussinov, R. Folding and binding cascades: shifts in energy landscapes. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (18), 9970−9972.

(168) Ma, B.; Kumar, S.; Tsai, C. J.; Nussinov, R. Folding funnels and binding mechanisms. *Protein Eng.* **1999**, *12* (9), 713−720.

(169) Kumar, S.; Ma, B.; Tsai, C. J.; Sinha, N.; Nussinov, R. Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci.* **2000**, *9* (1), 10−19.

(170) Gianni, S.; Freiberger, M. I.; Jemth, P.; Ferreiro, D. U.; Wolynes, P. G.; Fuxreiter, M. Fuzziness and frustration in the energy landscape of protein folding, function, and assembly. *Acc. Chem. Res.* **2021**, *54* (5), 1251−1259.

(171) Miskei, M.; Gregus, A.; Sharma, R.; Duro, N.; Zsolyomi, F.; Fuxreiter, M. Fuzziness enables context dependence of protein interactions. *FEBS Lett.* **2017**, *591* (17), 2682−2695.

(172) Ferreiro, D. U.; Komives, E. A.; Wolynes, P. G. Frustration in biomolecules. *Q. Rev. Biophys.* **2014**, *47* (4), 285−363.

(173) Christensen, N. G.; Sandberg, S.; Ness, K. O.; Jacobsen, H. [Glucose analysis in general practice. An examination of 4 instruments designed for general practice and self-monitoring]. *Tidsskr Nor Laegeforen* **1989**, *109* (16), 1770−1772.

(174) Bosshard, B. [Histogenesis and morphology of the tissues of the rabbit ear chamber−current light and electron microscopy findings]. *Morphol Med.* **1983**, *3* (2), 71−80.

(175) Nussinov, R.; Tsai, C. J.; Jang, H. Dynamic protein allosteric regulation and disease. *Adv. Exp. Med. Biol.* **2019**, *1163*, 25−43.

(176) Nussinov, R.; Tsai, C. J.; Jang, H. Does Ras activate Raf and PI3K allosterically? *Front Oncol* **2019**, *9*, 1231.

(177) Zhang, M.; Jang, H.; Li, Z.; Sacks, D. B.; Nussinov, R. B-Raf autoinhibition in the presence and absence of 14−3-3. *Structure* **2021**, *29* (7), 768−777.e2.

(178) Joerger, A. C.; Ang, H. C.; Fersht, A. R. Structural basis for understanding oncogenic p53 mutations and designing rescue drugs. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (41), 15056−15061.

(179) Liu, J.; Nussinov, R. Allosteric effects in the marginally stable von Hippel-Lindau tumor suppressor protein and allostery-based rescue mutant design. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (3), 901−906.

(180) You, D.; Zhang, J.; Xie, J.; Chen, B.; Ma, S. COAST: Controllable arbitrary-sampling network for compressive sensing. *IEEE Trans Image Process* **2021**, *30*, 6066−6080.

(181) May, M. Mutations to the rescue. *Nat. Med.* **2011**, *17* (4), 405−407.

(182) Filteau, M.; Hamel, V.; Pouliot, M. C.; Gagnon-Arsenault, I.; Dube, A. K.; Landry, C. R. Evolutionary rescue by compensatory mutations is constrained by genomic and environmental backgrounds. *Mol. Syst. Biol.* **2015**, *11* (10), 832.

(183) Bar, D. Z.; Arlt, M. F.; Brazier, J. F.; Norris, W. E.; Campbell, S. E.; Chines, P.; Larrieu, D.; Jackson, S. P.; Collins, F. S.; Glover, T. W.; et al. A novel somatic mutation achieves partial rescue in a child with Hutchinson-Gilford progeria syndrome. *J. Med. Genet* **2017**, *54* (3), 212−216.

(184) Csermely, P.; Palotai, R.; Nussinov, R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem. Sci.* **2010**, *35* (10), 539−546.

(185) Boehr, D. D.; Nussinov, R.; Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **2009**, *5* (11), 789−796.

(186) Noe, F.; Tkatchenko, A.; Muller, K. R.; Clementi, C. Machine learning for molecular simulation. *Annu. Rev. Phys. Chem.* **2020**, *71*, 361−390.

(187) Greener, J. G.; Jones, D. T. Differentiable molecular simulation can learn all the parameters in a coarse-grained force field for proteins. *PLoS One* **2021**, *16* (9), No. e0256990.

(188) Doerr, S.; Majewski, M.; Perez, A.; Kramer, A.; Clementi, C.; Noe, F.; Giorgino, T.; De Fabritiis, G. TorchMD: A deep learning framework for molecular simulations. *J. Chem. Theory Comput* **2021**, *17* (4), 2355−2363.

(189) Verkhivker, G.; Spiwok, V.; Gervasio, F. L. Editorial: Machine learning in biomolecular simulations. *Front Mol. Biosci* **2019**, *6*, 76.

(190) Gupta, C.; Cava, J. K.; Sarkar, D.; Wilson, E.; Vant, J.; Murray, S.; Singharoy, A.; Karmaker, S. K. Mind reading of the proteins: Deep-learning to forecast molecular dynamics. *bioRxiv* 2020; DOI: 10.1101/2020.1107.1128.225490.

(191) Kaptan, S.; Vattulainen, I. Machine learning in the analysis of biomolecular simulations. *Adv. Phys-X* **2022**, *7* (1), 2006080.

(192) Bai, Q. F.; Liu, S.; Tian, Y. N.; Xu, T. Y.; Banegas-Luna, A. J.; Perez-Sanchez, H.; Huang, J. Z.; Liu, H. X.; Yao, X. J. Application advances of deep learning methods for de novo drug design and molecular dynamics simulation. *Wires Comput. Mol. Sci.* **2022**, *12* (3), 1581.

(193) 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.

(194) Ding, W.; Nakai, K.; Gong, H. Protein design via deep learning. *Brief Bioinform* **2022**, *23* (3), bbac102.

(195) Pearce, R.; Zhang, Y. Deep learning techniques have significantly impacted protein structure prediction and protein design. *Curr. Opin Struct Biol.* **2021**, *68*, 194−207.

(196) Ingraham, J.; Riesselman, A.; Sander, C.; Marks, D. Learning protein structure with a differentiable simulator. https://openreview.net/forum?id=Byg3y3C9Km (accessed Sep 27, 2019).

(197) 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. U. S. A.* **2021**, *118* (48), e2104878118.

(198) Thorn, A. Artificial intelligence in the experimental determination and prediction of macromolecular structures. *Curr. Opin Struct Biol.* **2022**, *74*, 102368.

(199) Lee, D.; Xiong, D.; Wierbowski, S.; Li, L.; Liang, S.; Yu, H. Deep learning methods for 3D structural proteome and interactome modeling. *Curr. Opin Struct Biol.* **2022**, *73*, 102329.

(200) Li, S.; Wu, S.; Wang, L.; Li, F.; Jiang, H.; Bai, F. Recent advances in predicting protein-protein interactions with the aid of artificial intelligence algorithms. *Curr. Opin Struct Biol.* **2022**, *73*, 102344.

(201) Chen, H. M.; Shen, J.; Wang, L.; Song, J. N. A framework towards data analytics on host-pathogen protein-protein interactions. *J. Amb Intel Hum Comp* **2020**, *11* (11), 4667−4679.

(202) Oh, M.; Zhang, L. DeepMicro: Deep representation learning for disease prediction based on microbiome data. *Sci. Rep* **2020**, *10* (1), 6026.

(203) Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Zidek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon, A.; et al. Highly accurate protein structure prediction for the human proteome. *Nature* **2021**, *596* (7873), 590−596.

(204) Camacho, D. M.; Collins, K. M.; Powers, R. K.; Costello, J. C.; Collins, J. J. Next-generation machine learning for biological networks. *Cell* **2018**, *173* (7), 1581−1592.

(205) Palukuri, M. V.; Marcotte, E. M. Super.Complex: A supervised machine learning pipeline for molecular complex detection in protein-interaction networks. *PLoS One* **2021**, *16* (12), No. e0262056.

(206) Zhang, S.; Bamakan, S. M. H.; Qu, Q.; Li, S. Learning for personalized medicine: A comprehensive review from a deep learning perspective. *IEEE Rev. Biomed Eng.* **2019**, *12*, 194−208.

(207) Yang, J.; Zhang, Z.; Roe, S. M.; Marshall, C. J.; Barford, D. Activation of Rho GTPases by DOCK exchange factors is mediated by a nucleotide sensor. *Science* **2009**, *325* (5946), 1398−1402.

(208) Buchwald, G.; Friebel, A.; Galan, J. E.; Hardt, W. D.; Wittinghofer, A.; Scheffzek, K. Structural basis for the reversible activation of a Rho protein by the bacterial toxin SopE. *EMBO J.* **2002**, *21* (13), 3286−3295.

(209) Huang, Z.; Sutton, S. E.; Wallenfang, A. J.; Orchard, R. C.; Wu, X.; Feng, Y.; Chai, J.; Alto, N. M. Structural insights into host GTPase isoform selection by a family of bacterial GEF mimics. *Nat. Struct Mol. Biol.* **2009**, *16* (8), 853−860.