PCCP

ROYAL SOCIETY OF CHEMISTRY

PERSPECTIVE

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Phys. Chem. Chem. Phys.*, 2022, **24**, 26371

Thirty years of molecular dynamics simulations on posttranslational modifications of proteins

Austin T. Weigle,[a] Jiangyan Feng[b] and Diwakar Shukla*[bcde]

Received 24th June 2022,
Accepted 13th October 2022

DOI: 10.1039/d2cp02883b

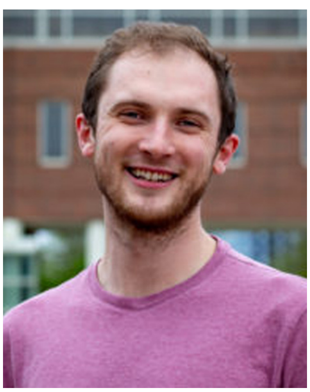
rsc.li/pccp

Posttranslational modifications (PTMs) are an integral component to how cells respond to perturbation. While experimental advances have enabled improved PTM identification capabilities, the same throughput for characterizing how structural changes caused by PTMs equate to altered physiological function has not been maintained. In this Perspective, we cover the history of computational modeling and molecular dynamics simulations which have characterized the structural implications of PTMs. We distinguish results from different molecular dynamics studies based upon the timescales simulated and analysis approaches used for PTM characterization. Lastly, we offer insights into how opportunities for modern research efforts on *in silico* PTM characterization may proceed given current state-of-the-art computing capabilities and methodological advancements.

1. Introduction

The basis of cellular function is rooted in atomistic movement. Irrespective of the molecular form, atomistic interactions resulting from these movements signal the need for necessary biophysical actions. DNA replication, RNA transcription, protein translation, and related signal transduction are all fundamental molecular processes which are dynamically recruited to satiate cellular need. Accordingly, an organism's genomic architecture equips its cells with the tools necessary to maintain homeostasis. The genome encodes genetic and proteinogenic responses to cellular perturbation. However, genomic architecture fixture

[a] *Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*
[b] *Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*
[c] *Center for Biophysics and Quantitative Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*
[d] *Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*
[e] *Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. E-mail: diwakar@illinois.edu*



Austin T. Weigle

Austin Weigle obtained his BS in Plant Biology (2017) with a concentration in biochemical and molecular physiology from Southern Illinois University Carbondale (SIU). He then joined the Chemistry PhD program at the University of Illinois at Urbana-Champaign (UIUC), specializing in Chemical Biology. Austin is currently a PhD candidate in Prof. Diwakar Shukla's research group at UIUC. His graduate research involves using molecular modeling, simulation, and informatic approaches to investigate chemical selectivity and regulation principles in membrane transporters. Austin also works on acyltransferase enzyme evolution.

Jiangyan Feng

Jiangyan Feng is a research scientist in the Computational Science and Protein Engineering Group at Eli Lilly and Company. Jiangyan received her PhD in Chemical and Biomolecular Engineering from the University of Illinois at Urbana–Champaign under Prof. Diwakar Shukla. Her PhD research was focused on employing molecular dynamics simulations and machine learning to understand protein dynamics and functions. She is currently focused on understanding antibody develop ability and antibody-antigen interactions by leveraging a variety of computational tools such as AlphaFold2 and protein language modeling.

occurs on evolutionary timescales. Thus any *a priori* information available to cellular machinery *via* the genome may not always possess the precision nor efficiency for any given stimulus occurring along physiological and developmental timescales. DNA and protein synthesis are essential provisions for maintaining homeostasis and cellular feedback mechanisms, but additional forms of spatiotemporal regulation are required to guarantee that the cell can acclimate to emergent complexity.[1] In turn, cells developed a versatile biochemical arsenal through posttranslational modifications (PTMs).[2]

PTM refers to either reversible or irreversible chemical changes in proteins after translation, which often occur on specific amino acid side chains.[3] PTM addition or removal is executed by means of a covalent modification by bonding or hydrolytic cleavage, typically occurring on side chains which may behave as nucleophiles (Cys, Ser, Thr, Tyr, Lys, His, Arg, Asp, Glu).[4] Hydroxylation of Pro and Asn act as exceptions.[4] Approximating almost 700 unique entries in the UniProt database to date,[5] the most common additive PTMs are phosphorylation, acylation (includes ubiquitination), alkylation, glycosylation, and oxidation; less common ones include hydroxylation/carboxylation, the addition of peptide moieties to preexisting protein residues, and sulfur-sulfur transfers, among others.[1,6,7] Proteolysis, deamidation, and eliminylation are examples of nonadditive PTMs.[8] Some PTMs for commonly modified residues, which can be computationally modeled, are shown in Fig. 1. Each PTM serves a role in altering the target protein in some way, perhaps by introducing stability or redirecting localization and subsequent interactions post-installment.[9] In effect, target protein structure and function are inherently changed upon PTM introduction, as the PTM fundamentally changes the physicochemical properties of its host amino acid.[4,6]

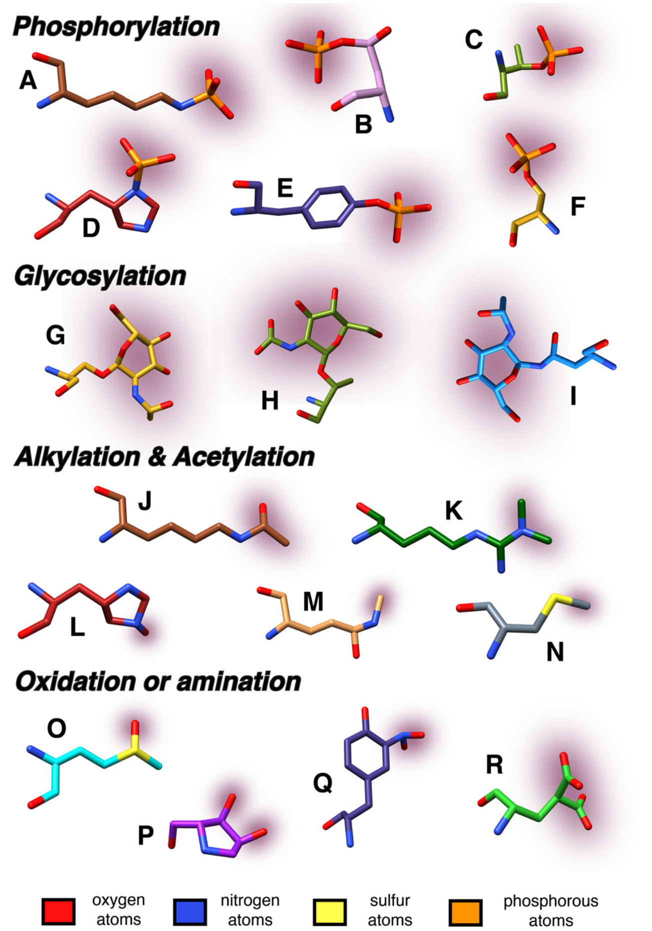


Fig. 1 Display of modeling capabilities for common PTMs on commonly modified residues. Modifications are shown for residues which can be readily modeled using the Vienna-PTM server.[10] PTMs are highlighted with a black raspberry glow. Structures resulting from the server are only parameterized with GROMOS force fields. Phosphorylation is shown for (A) lysine, (B) aspartate, (C) threonine, (D) histidine, (E) tyrosine, and (F) serine. Glycosylation is shown for (G) serine, (H) threonine, and (I) asparagine. Acetylation is shown for (J) lysine. Alkylation is shown for (K) arginine, (L) histidine, (M) glutamine, and (N) cysteine. Oxidation is shown for (O) methionine, (P) proline, and (R) glutamate. Amination is shown for (Q) tyrosine. A color key is provided at the bottom of the image for distinction of oxygen, nitrogen, sulfur, and phosphorous atoms. Colors not mentioned in the key represent carbon for each of the amino acids.

Diwakar Shukla

Diwakar Shukla is an associate professor in the Department of Chemical and Biomolecular Engineering at the University of Illinois at Urbana-Champaign. He is affiliated with the Center for Biophysics and Quantitative Biology, Department of Plant Biology and Department of Bioengineering. His research work is focused on understanding complex biological processes using physics-based models and techniques. He received his BTech and MTech degrees in Chemical Engineering at the Indian Institute of Technology (IIT) Bombay, India. He then joined Massachusetts Institute of Technology where he received his MS and PhD degrees in Chemical Engineering for his work on solution biochemistry with Prof. Bernhardt Trout. Before joining Illinois, he worked with Prof. Vijay Pande at Stanford University on developing distributed computing approaches understanding protein dynamics.

Thousands of proteins exist within individual proteomes, comprising a set of machinery capable of "reading", "writing", and "erasing" PTMs.[11–13] Reader proteins recognize PTMs installed by writers (kinases, ubiquitin ligases, acetyltransferases), through molecular interactions for the preparation of an adequate biological response to cellular conditions.[13] For reversible PTMs, eraser proteins (phosphatases, deubiquitinases, deacetylases) terminate molecular signal propagation by removing PTMs.[13] Furthermore, target proteins may contain multiple PTM sites, resulting in combinatoric PTM possibilities which may redefine expected cellular outcomes that would otherwise occur from just a single PTM instance.[1] Although PTMs may instantiate a sense of proteomic crosstalk, they are also capable of intra-protein crosstalk, communicating allosteric information about what kind of conformation the target protein should adopt in preparation for its next role in cellular signaling and stress response.[14] Interrogating how PTMs communicate such intra- and inter-protein changes thus becomes central to understanding how protein function is regulated throughout any physiological state of the cell.

As experimental advances in mass spectrometry and chemical biology have facilitated PTM discovery and methodologies for characterization,[13,15,16] the opportunity to systematically describe PTM functional outcomes gains clarity. Still, barriers exist to an exhaustive experimental interrogation of how PTMs influence protein function. Systematic studies using mass spectrometry often consider identification for few modification types of a single PTM rather than considering the existence or functional consequences of multiple PTMs.[3] Exploration of multisite PTM characterization poses similar challenges as seen for exploring higher-order mutations in variant effect prediction with respect to throughput.[17] Although cellular assays can be performed to examine PTM effects on protein expression, function, and downstream effects, dissecting how these modifications relate to individual protein structure and conformational changes requires more effort. Throughput for spectroscopy and structure determination experiments is limited by large scale efforts of *in vitro* heterologous expression for both soluble and membrane proteins, as well as specialized training for probe/label design or preventing protein precipitation. However, *in silico* approaches to studying PTMs offer an alternative for challenging experimental characterization efforts.

In this Perspective, we highlight the foundations laid by PTM *in silico* research spanning the last 30 years to offer insights towards future PTM characterization using computation. Recent advances in computing power and simulation methodology have rendered *in silico* molecular modeling and molecular dynamics (MD) simulations an important, and increasingly accessible, tool to understand biological macromolecules.[18–22] Here we compare what MD simulations of PTMs have achieved in the past using less computational power and fewer methodological advances with what is now capable with more modern resources and simulation techniques. Our goal is to illuminate the eventual quality of PTM functional understanding so that it may rival the increasing quantity of discovered PTMs.

2. Modeling posttranslational modifications for molecular dynamics simulations

2.1. Introducing molecular dynamics simulations

MD simulations require two inputs: initial positions extracted from an experimentally determined (or computationally predicted) structure and a force field describing the interactions between atoms. A typical force field consists of the following terms: covalent bonds, bond angles, dihedral angles, electrostatic interactions, and other non-bonded interactions. According to Newton's second law of motion, the acceleration of each atom can be computed. Using numerical integration techniques, the positions and velocities of every atom in the system can be updated as a function of time. To ensure numerical stability, the time steps in an MD simulation must be short – typically only a few femtoseconds – so that each time step is shorter than the timescale of the fastest chemical bond vibration. The output of an MD simulation is a trajectory file containing the position of each atom at each time step, describing the motion of the system at the atomic level during the simulated time period.[23] Preliminary stages of simulation involving heating and restraint relaxation are applied first, ensuring thermostability of modeled molecules and their starting conformations. After undergoing short simulations where restraints have been removed and the model system has converged to a stable state, data can then be reliably recorded from trajectories known as "production runs" with minimal fear of simulation artifacts.[24]

Before committing to any computational study design, the greatest bottleneck to conducting MD studies is the availability of computing resources. This caveat is especially important, as the simulation of a given biomolecular process must occur over a specified timescale to accurately sample the conformational ensemble for the dynamics of interest.[19,25,26] Scientific computing costs have been alleviated as more powerful hardware has been produced, but simulation throughput has historically not been as efficient nor parallelizable as can be seen today. Within the first decade of the 21st Century, the acceleration of MD simulations *via* porting to graphical processing units (GPUs) was retarded by a few lurking variables – namely scaling across multiple GPUs, changing how memory was accessed on GPUs *versus* central processing units (CPUs), and interfacing with compilers which would allow for MD software to run on GPU source code.[27] By 2010, these challenges were eventually overcome by software developers, where MD simulation work performed on GPUs became routine and continuous efforts were made to increase simulation efficiency.[28–31] Naturally, the outcomes of GPU porting achievements were amplified by the integration of MD workflows onto large distributed computing as well as supercomputing platforms (*e.g.*, private clusters, Folding@Home, D.E. Shaw's Anton, the National Science Foundation's Frontera, or National Center for Supercomputing Resources like BlueWaters and XSEDE initiatives).[32–37]

MD methodologies can be applied for studying the evolution of specific biomolecular processes, evaluating how these processes may be perturbed following a controlled change, or simply demonstrating conformational flexibility and stability along some molecular structure.[20] Technical advances for MD simulation can be divided based on whether they address phase space sampling or interpretation of resulting simulation data. While we focus on MD study design with respect to PTMs later in this Perspective, the different types of MD simulation techniques, their application and development, have been extensively reviewed elsewhere.[38] On the other hand, the large deluge of results and calculations derived from MD simulations has prompted the use of machine learning for improved data representation and analysis.[39–41]

Regardless of the method used for sampling, measures for uncertainty quantification and sampling quality have been communicated as best practices when designing and conducting MD simulation studies.[42] A variety of simulation techniques used to calculate binding free energies, thermostability, or kinetic rates of transition often come in exact agreement, or occur within error, of experimental observations.[43–48] Minimally, MD simulation can offer qualitative agreement with experimental results. Simulation accuracy is extremely sensitive to the initially modeled conditions, but can be strengthened by repeated and systematic sampling of the phase space for a given process. Outside of the modeled structure and selected parameterization set quality, the predictive power of MD simulations and the minimization of error lies in the attempt for approximating statistical convergence of ensemble-averaged properties.[49] Today, this is more easily achieved by running longer and greater trajectories thanks to better computing and ensemble simulation-based approaches.[49]

2.2. In silico installation of PTMs for preparing simulation studies

Proteins are highly dynamical systems with functions governed essentially by their dynamical behaviors.[50] MD simulations thus become a powerful tool for studying PTMs. Notably, MD simulations can capture the dynamic behavior of biological systems at atomistic resolution in a label-free manner. They can therefore offer a unique structure-dynamics perspective in deciphering the roles of PTMs on proteins without any form of perturbation introduced by probes.[51,52] Furthermore, the simulation conditions can be carefully controlled by the simulation practitioner. So by comparing simulations performed under different PTM conditions, one can identify the effects of a wide variety of PTMs, which is usually experimentally unfeasible.[53] Accurate PTM modeling is still reliant on appropriate information related to the PTM-bearing amino acid site and tools to introduce and parameterize these changes to the protein structure.

Databases and informatic methods cataloguing general PTM information,[54–66] as well as resources depicting experimentally-validated or computationally-predicted sites specific for phosphorylation,[67–88] glycosylation,[80,89–98] and acylation,[80,99–119] among other modification forms,[80,120–132] exist. These resources can be used to identify PTM sites on target proteins for improved molecular modeling. Traditionally, if a PTM was not already present on a resolved crystal structure, covalent modifications were made to input coordinate and topology files to reflect the correct modification using modeling software or through manual text editing of the input files. Such hands-on changes can be introduced using either molecular visualization software[133–138] or system preparation modules found in popular MD simulation engines (*e.g.* AMBER, CHARMM, GROMACS, NAMD).[29,30,139,140] Eventually, standalone- and web-tools were developed to streamline PTM installation onto protein structures.

As a historiographic example, glycosylation efforts starting from the late 1990s and continuing into the mid 2010s were focused on modeling carbohydrate structures, although none of these tools could directly merge the resulting sugars onto protein coordinates and generate glycoprotein topologies necessary for simulation.[141–145] That is, these software offer limited accessibility to PTM modeling by making simulation topology generation exceedingly nontrivial for nonexperts. In 2005, GLYCAM emerged as a tool developed specifically for modeling glycoproteins and solution carbohydrates as AMBER and CHARMM simulation inputs.[146] CHARMM developers then introduced glycan reading and modeling functionalities into the CHARMM-GUI webserver for glycoprotein structure preparation.[147–149] Similarly, GROMACS developers made *doGlycans* for modeling GROMACS-compatible glycoprotein structures.[150] Although each of these tools enabled glycoprotein structure preparation, preference for one MD engine over another confound progress on PTM modeling. Features available in some tools for one engine – *i.e.*, sugar library chemical and conformational diversity – may not be available for another, requiring nontrivial file conversion and topology preparation strategies using additional or third-party tools. Indeed, juxtaposed singularity between different software and computational tools is an inherent obstacle in molecular modelling!

By 2017, several MD engine camps had developed their own respective tools for glycoprotein structure generation, inspiring force field-independent efforts to make carbohydrate structures available and amenable for MD input file preparation.[151–153] Outside of common PTMs like phosphorylation and ubiquitination, additional tools were developed to facilitate modeling. The PyMOL plugin PyTM can model phosphorylation, acetylation, carbamylation, citrullination, nitration, methylation, hydroxylation, aldehyde adduct formation, as well as any combination of the above.[154] The Vienna-PTM server offers residue templates for over 256 (non)enzymatic modifications, although modified output structures are designed to be used within a GROMACS framework.[10] Despite these advances, CHARMM-GUI maintains popularity because of its ease of access; its phosphorylating, lipid anchoring (acylate), glycosylating, and peptide stapling capabilities; and its file output schemes that maximize compatibility with diverse MD engines (Fig. 2).[155–157] Another reason for the popularity of CHARMM-GUI is its ability to archive structures of interest for modelers to download and immediately use in their own research. A great example of this is the recent availability of a fully glycosylated, full-length SARS-CoV-2 spike protein model in a realistic viral membrane, which has been used directly for subsequent computational study.[158–164] Regardless of the methods used for PTM installation, modified amino side chain orientations may be further optimized using rotamer libraries or through equilibration simulations.[165]

With protein structure files covalently modified, PTM parameterization is required to ensure that simulations accurately depict biophysical behaviors as expected from experiment. Initially, PTM models did not necessitate their own specific force fields. MD practitioners could then argue that accuracy of PTM-containing simulations was not significantly compromised, as atom parameters could be assumed by analogy from preexisting libraries. Such practice was acceptable at the turn of the 21st Century, before more PTM-specific force fields were developed. For common PTMs, phosphorylation moieties could borrow nucleotide parameters to describe their phosphate groups; likewise, glycosylation, ubiquitination, and acylation additions shared atoms with protein and small molecule force

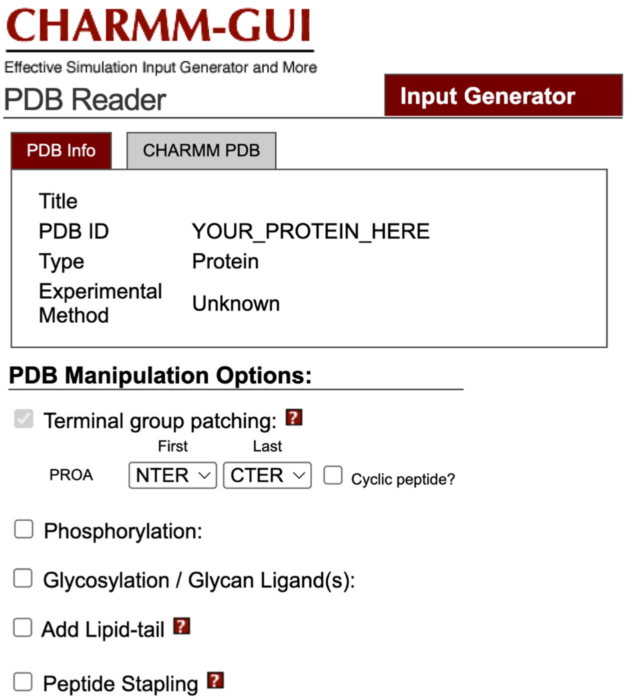


Fig. 2 Modified screenshot of CHARMM-GUI PDB Reader function highlighting specific PTM options for input protein structure manipulation.

fields. Missing parameters were otherwise developed as needed by specific research groups without immediately being publicly available. Specific force field developments for particular PTMs, such as phosphorylation and glycosylation, were later expanded into addons and were incorporated into libraries commonly used for each of the popular MD engines.[166–176] It is worth noting that the majority of discussed modeling and parameterization efforts were first dedicated to characterize biologically common PTMs, suggesting that a need for improved modeling of less common PTMs still remains.

3. Designing molecular dynamics approaches for the study of posttranslational modifications

Perhaps the beauty of molecular simulation lies in its unfettered power to describe molecular systems in almost unlimited detail. Extent of available computational resources and methodologies will always be a bottleneck to data procurement, but the onus of appropriate research design and meaningful analyses falls to the MD practitioner.

Generally, trajectory analyses assume either a time-dependent or time-independent form. Time-dependent data collection regimes incorporate analyses that project some feature against the order parameter of time, following observations exactly as they transpired along individual trajectories. In contrast, time-independent data collection regimes instead employ analyses that project data against order parameters as a timeless ensemble, where data points occupy positions within a dimensional phase space with respect to the selected order parameters.[177–180] Removing reliance on time allows for data to be viewed from a perspective of population density as opposed to mere frequency of occurrence within a trajectory. Such a style of data projection can be seen in free energy landscapes (FELs), where aggregate data points occupying highly frequent regions in the projected dimensional space can be identified as either metastable or transitionary macrostates.

Each analysis and data collection approach has its strengths and weaknesses. Time-dependent data collection and analyses are typically employed to focus on specific interactions given a starting structure and with or without a target structure in mind. These types of results follow "single" or "few" trajectory simulation schemes where coordinate and velocity information are retained to maintain the existence of a long trajectory. As such, time-dependent data representations offer a simplified perspective reinforced by multiple replicates of production runs observing similar molecular events. Because data output by individual time-dependent trajectories is not aggregated, enhanced sampling methods are often used instead of just classical MD simulations to capture molecular transitions occurring along longer timescales. Approaches like accelerated, targeted, or steered MD are examples of such techniques.[181–183] Although studies following a time-dependent data collection regime can utilize time-independent analyses, these studies tend to lack sufficient sampling to offer comprehensive conclusions beyond the direct research question in focus.

On the other hand, time-independent data collection and analyses are employed to observe a desired transitionary path from one macrostate to another along order parameters which are not based in time. Data representation can then naturally take the form of FELs. Studies designed using time-dependent approaches are still capable of capturing a FEL but may not be able to do so for longer timescale processes. Time-independent approaches achieve this goal by combining the results from an aggregation of "brute force" classical MD simulations or by biasing potentials along selected order parameter(s). Such a biasing method includes metadynamics, where the choice of order parameters directly affects how easily transitions may be sampled.[184] Results from brute force simulation of unbiased parallel trajectories need to be "stitched" together using kinetic frameworks like Markov state models (MSMs) or milestoning,[185–188] while path search and optimization protocols can otherwise be biased along potentials using approaches like umbrella sampling, replica exchange, or the string method.[189–191] Although nontrivial to implement, results from unbiased and biased simulations can be augmented into a single multiensemble kinetic framework for time-independent data representation.[48] Depending on the extent of collective variable exploration, time-independent data representation can demand extremely more resources than research designs relying on a time-dependent data collection and analysis approach.

Herein, we describe the evolution of MD studies on PTMs and their effect on protein structure-function. We begin by discussing methodologies and results from studies observing how PTMs influence protein structure and function along shorter timescales. In these studies, the dominant styles of molecular simulation include time-dependent data collection and analysis regimes, as well as path optimization methods like replica exchange MD (REMD) given available experimental structures. Subsequently, we transition to studies employing FELs for detailing how PTMs impact the range of protein function. Thanks to increased computing capabilities and methodological advances, PTM MD studies wielding a FEL-based approach have recently become more widespread within the literature.

4. Foundational insights on posttranslational modifications revealed by studies using trajectory-specific approaches

During the mid-1990s, MD practitioners were first challenged with the task of modeling and simulating covalently modified proteins. At first, it was questioned whether results from PTM simulations could even concur with experiments. The initial quest to address how PTM dynamics regulated protein function was confounded by other lurking variables, too. As described in Section 2, few existing tools to easily modify protein structures were available; therefore, initial works were constrained to studying proteins whose resolved structures already contained PTMs. Additionally, GPU-based acceleration of MD simulations was not implemented until the late 2000s, meaning that approaches to modeling PTMs needed to be carefully designed so that they could efficiently offer unique insights with minimal data and computing time requirements. Many of these initial simulation studies between 1990 and 2010 were therefore performed along picosecond to nanosecond timescales, focusing on specific regions of protein structures or interactions with covalently modified peptides. Short timescale MD studies were capable of modeling intramolecular protein sidechain contacts and whether target protein or peptide structures were altered by PTM installation. Identifying how hydrogen bonding networks could be altered after PTM introduction was a worthwhile contribution for the time, as NMR experiments could not always distinguish hydrogen bonding interactions involving PTMs which lacked hydrogens. Despite how early computing capabilities limited MD to the study of short timescale transitions like changes in hydrogen bonding, any atomistic details offered from experiments alone were still incomplete without the help of molecular modeling.

4.1. Characterization of active and binding site rearrangement through PTM interactions

What was first of interest included the structural consequences of PTMs on sidechain orientation and peptide secondary structure within binding pockets and recognition interfaces. In 1995, the basis of aldolase inhibition by a synthetic peptide matching the first 15 residues of human erythrocyte peptide band 3 (B3P) was characterized using MD simulation of NMR-resolved aldolase structures.[192] The structure of the peptide in complex with aldolase was determined *via* NMR, and a phosphorylated version of the peptide was then docked into the aldolase binding pocket and simulated for 5 ps. An unphosphorylated hydroxyl moiety of the peptide's tyrosine group was able to experience favorable interactions within the docking site, but pTyr resulted in unfavorable interactions and inhibited certain rotational ranges of motion because of electrostatic repulsion. These findings offered an explanation as to how B3P phosphorylation inhibited glycolytic function of aldolase off phosphate electrostatics alone.[192] Using computing capabilities available before the year 2000, the binding of unmodified, P_i-bound and phosphopeptide-bound phospholipase SH2 domain MD simulations were modeled in a separate study.[193] To give perspective on the available computing power, SH2 was simulated for 250 ps at a rate of 3 ps day^{-1} in this study.[193] These phospholipase simulations demonstrated how PTMs recruit mostly the same residues as their solution-based counterparts (*e.g.*, P_i *versus* pTyr), as P_i-bound and pTyr simulations highlighted the importance of a conserved arginine triad in stabilizing phosphate group positioning within the SH2 binding pocket (Fig. 3).[193] Here, short simulations could suggest an evolutionarily conserved role within binding pockets for select residues when handling PTMs by maintaining the hydrogen bonding network of the active site.

In other cases, there exists a possibility for active site rearrangement following PTMs, which need not necessarily be a result of direct residue-residue interactions with the covalently modified PTM site. Comparison between catalytic domains seen in the phosphorylated structures of inactive (closed) Src tyrosine kinase and active (open) Lck kinase suggested that phosphorylation-induced conformational change along a polypeptide linker was required to reflect kinase active state (Fig. 4).[194] To study this dynamic hinge motion hypothesized for maintaining active *versus* inactive states upon phosphorylation, a targeted MD approach was used where increasing amounts of harmonic restraints were applied across increments of 60 ps to the Src kinase hinge region Cα atoms.[194] From this, targeted MD

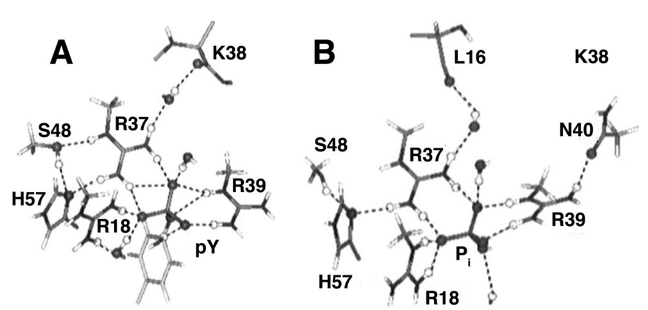


Fig. 3 SH2 domain phosphate binding pocket hydrogen bonding network. (A) Hydrogen-bonding network shown for phosphopeptide binding to the phosphate binding site. (B) Hydrogen-bonding network for inorganic phosphate to the phosphate binding site. Both binding simulations depict central role of conserved arginine triad in stabilizing phosphomoieties. Adapted with permission from M.-H. Feng, M. Philippopoulos, A. D. MacKerell and C. Lim, *J. Am. Chem. Soc.*, 1996, **118**, 11265–11277.[193] Copyright 2022 American Chemical Society.

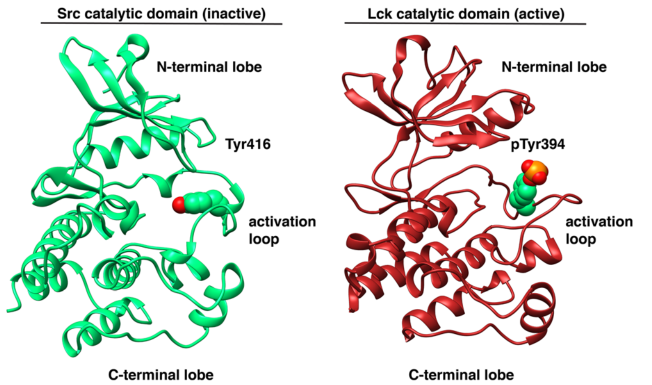


Fig. 4 Comparison of Src kinase inactive structure (left) *versus* Lck active structure (right). A common strategy for exploring PTM-induced conformational change at reduced computational cost was to use covalently modified starting structures in different conformations. This case study used the Lck kinase phosphorylated active structure to create a homology model for the active Src kinase structure. Adapted with permission from J. Mendieta and G. Gago, *J. Mol. Graph. Model.*, 2004, **23**, 189–198.[194] Copyright 2022 Elsevier.

converted pTyr416 Src kinase from the resolved closed state to a target pseudoactive open state based off the Lck crystal structure.[194] Targeted MD could not exactly mimic an Lck-like extended conformation until bound ATP was replaced with ADP, which then forced immediate conformational change within 800 ps in support of experimentally proposed intramolecular autophosphorylation of Tyr416. Thus, in this 2004 study, Src kinase could only achieve the active conformation when the active site reorganized around pTyr416 and ADP interactions.[194] Similar perturbations to cofactor coordination in cyclin-dependent kinase 2 (CDK2) active site residues were seen during 60 ps classical MD simulations of inhibitory phosphorylation along the G-loop.[195] Still, the position for phosphorylation is critical for determining the effect on active site rearrangement. For CDK2, it was reported from classical simulations ranging between 3–10 nanoseconds that pThr160 activation occurred by constraining the active site T-loop for phosphotransfer; meanwhile pThr14 and pTyr15 introduced enough flexibility in the G-loop to widen the active site, destabilize cofactor placement, and inhibit phosphotransfer.[196,197] Thus, PTM installation can rearrange loop conformations to alter protein function by generating new cavities for accommodating protein-specific activation functions.[198]

Active site reorganization can also occur through solvation effects. Phosphorylation of human eukaryotic translation initiation factor 4E, or eIF4E, found that pSer209 narrowed the dinucleotide cap binding pocket after shortening nearby backbone Cα-Cα distances by about 10 Å throughout an 800 ps simulation performed in 2003.[199] However, this tightening of the dinucleotide binding pocket could be accredited to dianionic phosphate recruiting a large solvation shell as a consequence of the pSer209 phosphate group's electronegative charge. This localized electronegativity thus created a force that pushed active site residues away and further enclosed the dinucleotide binding site.[199] Either by electropositive amino acid composition or hydration, positive charge is needed in order for phosphopeptide binding or phosphorylation to be supported.[200] In the case of phosphorylation, constrained conformational change is likely the means by which proteins accommodate the introduction of negative charge.

4.2. Influence of PTMs on protein structure, thermostability, and allosteric communication

Outside of active site studies, other short peptide simulations would proceed to computationally characterize the effects of phosphorylation by highlighting the importance of electronegativity introduced by phosphorylation. Phosphorylation was seen to induce conformational changes in a 10-residue region of tyrosine hydroxylase where most metastable states depicted peptide compression after 5 ps of simulation, unlike what was seen in the unphosphorylated structure.[201] Salt bridges seen in the unphosphorylated tyrosine hydroxylase peptide were broken, leaving many charged residues solvent-exposed as the pSer moiety hydrogen bonded with terminal residues. New interactions and conformations caused by Ser phosphorylation at one tyrosine hydroxylase site were argued to be the cause for cooperative increases seen in experimental phosphorylation at additional sites along the peptide region.[201] Peptide compression in response to phosphorylation has been reported in additional studies spanning from 2002 to 2012, too, as well as the breaking of native salt bridges in order to generate hydrogen bonds for mitigating the negative charge of phosphate groups.[202,203] Other forms of conformational change for Ser phosphorylation have been reported, such as stabilized alpha-helical secondary structure formation based on phosphate-backbone electrostatic interactions,[204] as well as preferred dihedral orientations introduced at the phosphorylation site.[205] Thr-specific phosphorylation was also seen to increase helicity in 100 ns classical simulations of p53 and p73 proteins, although decreases in helical content have been seen for other proteins.[206,207] Even as recent as 2019 has the interplay of individual PTM forms been found to stabilize N-terminal helices through phosphate charge neutralization and backbone-sidechain hydrogen bonding *via* advanced REMD simulations.[208] From the large body of phosphorylation modeling and simulation research, modification along specific amino acids and in different local environments will alter the probabilities of distinct resulting conformations. Work from the Jacobson group using umbrella sampling and quantum mechanical studies found in 2007 that hydrogen bond strength with phosphate groups noticeably changes in response to the geometry assumed by different phosphorylated residues based on their amino acid identity and their local environments.[209] Similarly, REMD simulations on glycosylated proteins also showed in 2010 that protein secondary structure can be altered in a site-specific manner based on shifted backbone dihedral preferences after glycosylation.[210]

Intrinsic disorder and repositioning of flexible loops serve as strategies for proteins to rapidly change their local environments in response to dynamic processes. PTM installation along disordered regions can instead exacerbate cellular perturbation response strategies of proteins by altering thermostability.[211] In 2008, the Levy group designed an elegant study by systematically modeling glycan groups onto an SH3 domain protein to distinguish the entropic *versus* enthalpic effects of glycosylation on protein function (Fig. 5A).[212,213] Coarse-grained modeling of the glycosylated SH3 structures enabled this systematic study design, where 63 different glycosylated SH3 domains could be efficiently simulated at different temperatures and in replicates. By varying temperature for the different simulations, Arrhenius-style free energy projections showed how increased extent of glycosylation increased protein thermal stability, where both enthalpy and entropy were reduced by the addition of glycans (Fig. 5B–D). The addition of glycans to SH3 was found to rigidify the protein structure, thereby lowering the enthalpy as the glycan moieties restrict protein dynamics. Similar results and effects on dynamics were seen on a variety of glycosylated protein structures by the Im group in a 2015 study.[214] Given the breadth of collected simulation data for the time, Shental-Bechor and Levy's study could offer direct insights towards SH3 stabilization that could not be offered by experiments alone.[212,213] Even the restriction of potential dynamics by PTM installation could impose significant biological implications, as the PTM offers other contributions to the target protein structure. For example, the Woods group found through 1 μs simulations of various SARS-CoV-2 spike protein glycoforms that glycosylation shields on average approximately 40% of the glycoprotein surface from antibody recognition.[159] In this case, the evolutionary advent of a PTM enables the SARS-CoV-2 virus to better bypass human immune systems and promote host infection.[164]

Even if PTMs may not principally operate through electrostatic perturbations, the entropic and enthalpic contributions offered by a specific covalent modification contribute to how the effects of a PTM manifest in a site- and modification-specific manner. Modeling and simulation by the Levy group have accurately predicted glycosylation-induced destabilization of WW-domain protein in regions where large numbers of contacts in native topology models preexist, as glycan installation would disrupt otherwise stabilizing interactions.[215] However, the opposite is not necessarily true when comparing simulation to experiment. The same 2010 study found that just because a glycosylation site along the WW-domain scaffold appears as if it can sterically accommodate a glycan does not universally imply that such a modification will result in

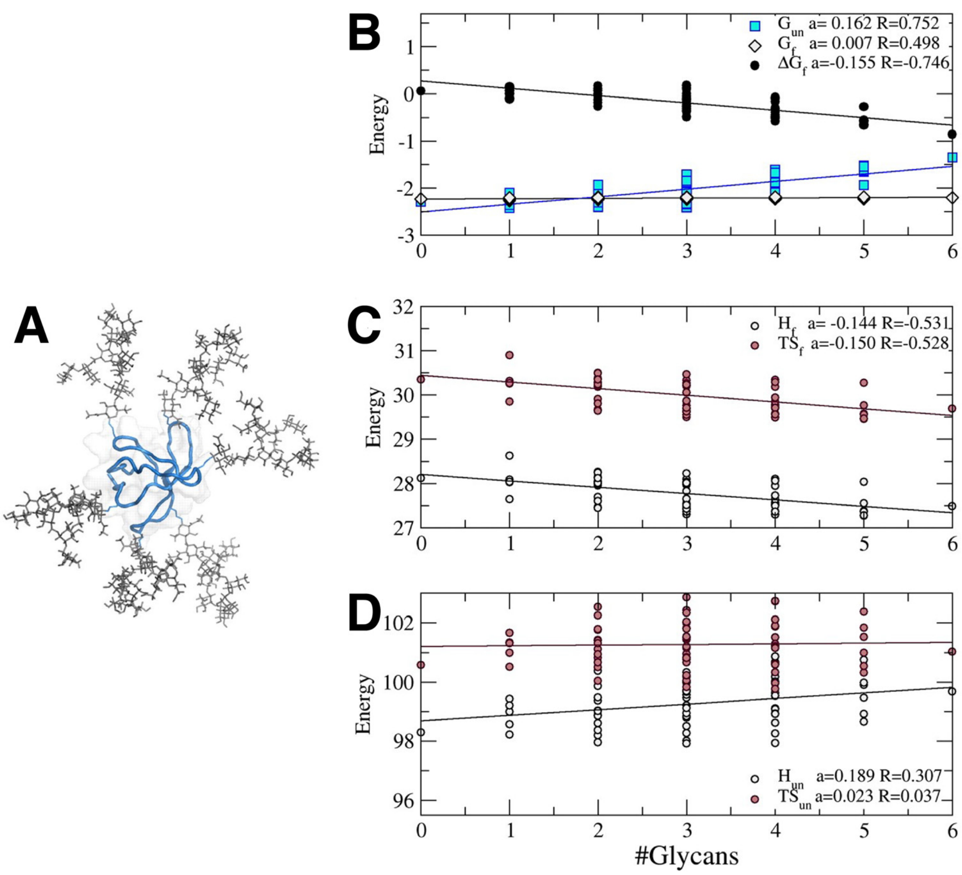


Fig. 5 Glycosylation effect on thermodynamic stability of SH3 domain protein. (A) SH3 domain represented as a blue ribbon, with glycosylation moieties represented as grey sticks. This modified protein was simulated using a coarse-grained native topology model rather than the all-atom version presented here. (B) Folding free energy (ΔG_f) of the SH3 domain protein. (C) SH3 domain protein folded state thermodynamics, broken down by enthalpic and entropic contributions. (D) SH3 domain protein unfolded state thermodynamics, broken down by enthalpic and entropic contributions. Panels B–D represent free energy as a function of the length of the glycan chains added. Adapted with permission from D. Shental-Bechor and Y. Levy, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 8256–8261.[212] Copyright 2022 National Academy of Sciences.

stabilization.[215] Entropic stabilization could instead arise through PTM-based dehydration, as WW-domain protein PEGylation simulations suggest that solvent removal from the protein surface by PEGylation improves solubility and folded state preference.[216] Thus, it can be expected that entropy-based PTM effects are tied to the hydrophobicity and stericity inherent to that modification. In 2016, the Jacobson group found that the entropic effects of ubiquitination on a human kinase could not be replicated when performing acetylation at the same site, as an active-like conformation caused by ubiquitination was not observed across 10 ns simulations of an acetylated structure.[217] The difference in effect between acetylation *versus* ubiquitination could be attributed to the difference in size and bulk between the two types of covalent modification, a finding which likely applies to other forms of PTMs as well. For instance, a 2020 modeling and MD simulation of diverse PTMs onto the central coiled-coil connector of fibrinogen varied from having zero to severe structural effects.[218]

The impacts of PTMs on protein conformation extend beyond local environments and into long-range contacts and allosteric communications. As the Levy group helped identify how loss in natively stabilizing interactions upon glycosylation underscored structural destabilization in 2015, they further showed how interactions lost from glycosylation also include long-range contacts in the case of simulations of glycosylated MM1 protein.[219] Additionally, long-range communication between covalently-modified residues is also site-specific, as only modifications on certain regions of a protein may induce conformational change in regions distinct from the site of modification.[220] For example, terminal phosphorylation only increased cross-correlated motion between the SH2 and SH3 peptide binding domains in c-Src kinase, but not for other parts of the protein during simulations from 2001.[221] In the case of multiple PTM sites, only acetylation on a specific H4 histone tail lysine residue uniquely altered the simulated conformational ensemble in 100 ns REMD simulations, although acetylation at any lysine along the tail increased long-range backbone contacts.[222] Synchronous residue-residue contacts can also been influenced by PTM crosstalk. Different REMD system setups modeling combinatoric acetylation of High Mobility Group Box (HMGB) protein lysines found that contact networks in HMBG:DNA complexes differed despite each combination being minimally comprised of the same two to four lysine site acetylations.[223]

4.3. Summary of PTM simulations using "single" or "few" trajectory approaches

Despite using short trajectories analyzed in time-dependent manners, most of the works discussed above illustrated key findings with respect to how PTMs impact protein dynamics and function. Binding of covalently modified peptide substrates indicated direct interactions between protein binding interfaces, where side chain reorientation accommodated new electrostatic or steric forces brought by the PTM. When installed onto part of the target protein structure, PTMs were seen to alter secondary structure of disordered regions by affecting backbone hydrogen bonding preferences and again calling for side chain reorientation to offset any destabilizing interactions with stabilizing ones. Conformational preference thus reciprocates the nature by which a protein structurally responds to a particular PTM. Subversive changes in local side chain environments in response to changed hydrophobicity or electrostatic potential nearby the PTM site are perhaps the driving forces which help guide a modified protein towards a metastable state. Systematic modeling of PTMs indicative of yeast stress responses have shown that covalently modified structures simulated for 20 ns classical MD simulations do demonstrate conformational change, but that the extent of conformational change largely persisted as side chain fluctuations instead of backbone movements.[224] Site-specificity underlies all PTM effects on the ability for a protein to bind and function. In an enormous study on the effect of phosphorylation on G-protein-coupled receptor activation and arrestin binding in 2020, the Dror group performed 1 ms of independent simulations to show that site-specificity of phosphorylation patterns results in different extents of activation and arrestin binding.[225] In reference to other studies mentioned so far, each PTM position and type were shown to display unique effects per protein, which may not necessarily be confined to the neighborhood of the PTM site. This point was emphasized by the Shukla group in their 2022 study on the effects of N-linked glycosylation on SLC6 transporters.[226] From over 1 ms of independent simulations using combinatoric glycosylation patterns across four different SLC6 transporters, Chan and Shukla concluded that the impacts of glycosylation should be characterized at a per-protein level regardless of shared function within a family.[226] And amazingly, across all studies implementing the "single" or "few" trajectories approach, decent extents of conformational change, and the conversion between different states, could be observed in very short amounts of simulation time.

What one should notice about this style of MD simulation and analysis is that its popularity as a design strategy has remained popular throughout the entire history of PTM simulations. That is, sufficient observations could be made from either classical or biased MD performed along shorter timescales. Conversely, these types of studies still have their limitations. Thermodynamic and kinetic insights are often not obtained from studies utilizing a "single" or "few" trajectory approach. However, the descriptive power of these studies is enabled because of the dramatic conformational changes caused by introduction of steric bulk or electrostatics by the specified PTM(s). Because of computational expense, a number of studies enumerated within this section simulated regions of proteins as peptides, which may affect how their findings translate to the effects PTMs have on their respective protein's global structure. In most of the exemplary literature cited in this section, a chief complaint written in their discussion sections was that not enough simulation could be performed because of available computing resources. Given how substantive conformational change could be seen even along the pico- to nano-second timescales, these foundational studies underscore the impact PTMs have on their local and distal environments. Still, one cannot help but wonder what could be seen from PTM simulations ranging from micro- to milli-second timescale trajectories, generating aggregate datasets, and/or implementing more advanced enhanced sampling methodology.

5. Using aggregate datasets for the modulation of conformational free energy by posttranslational modifications

Free energy landscape (FEL) representations of MD datasets are useful in that they represent a probability distribution as to how often a particular protein macrostate is observed given a finite extent of sampling. Because FELs represent distributions, they involve sampling to observe a given process. The extent of sampling required for a given process depends on its timescale, where rare events require enhanced sampling because they are higher energy. Rare events occur on slower timescales and are therefore less likely to record during unbiased simulation. Sufficient FEL coverage offers a comprehensive view of all possible conformations related to a protein's function (as described with a given order parameter set). Sampling rare events is then necessary to fully understand how a protein uses intermediate states as transitions between metastable states; note that these transitionary conformations are very difficult to resolve experimentally and can often only be observed computationally. Thus, time-independent data collection regimes using FELs of datasets gathered by enhanced sampling methods offer the unique opportunity to describe how PTMs affect entire protein conformational ensembles. The timeframe for PTM studies implementing FEL-based approaches is primarily split across two time periods: the turn of the 21st Century and after 2015.

5.1. Ensemble-based sampling of short timescale processes using classical molecular dynamics

Depending on the timescale of the modeled process, PTM studies implementing a time-dependent research design can offer useful FEL-based insights. Examples of this include demonstrating how phosphorylation of a peptide shifts backbone torsion or dihedral angle conformational preferences.[204,205,207,227,228] Another process with a relatively short timescale includes dynamics of reasonably sized loops. Comparing 2.5 μs classical MD data of loop movement *versus* solvent accessible surface area of an E2 ubiquitin ligase enzyme showed that phosphorylated variants could only occupy an open, but not a closed, conformation because of the increased solvation required to satiate electronegativity along the installed phosphate groups.[229] FEL projection of principal component analyses for 1 μs WNK1 kinase simulations were also able to isolate specific activation loop conformations, demonstrating that a $\sim$3 kcal mol^{-1} free energy barrier prevented the unphosphorylated activation loop from adopting an extended conformation readily seen in phosphorylated WNK1 simulations.[230] More stringent FEL analyses using dihedral angles further showed that unphosphorylated WNK1 data hardly had any conformational overlap with structures seen during phosphorylated simulations (Fig. 6). However, the free energy difference between phosphorylated and unphosphorylated WNK1 in regions of shared dihedral angle conformational sampling still approximated $\sim$3 kcal mol^{-1}.[230]

5.2. Use of biasing potentials for simulating PTM dynamics

When following single trajectory analyses, sometimes ''vanilla'' MD simulations performed at standard temperature and pressure conditions are insufficient for capturing processes associated with PTM-based conformational change. High temperature

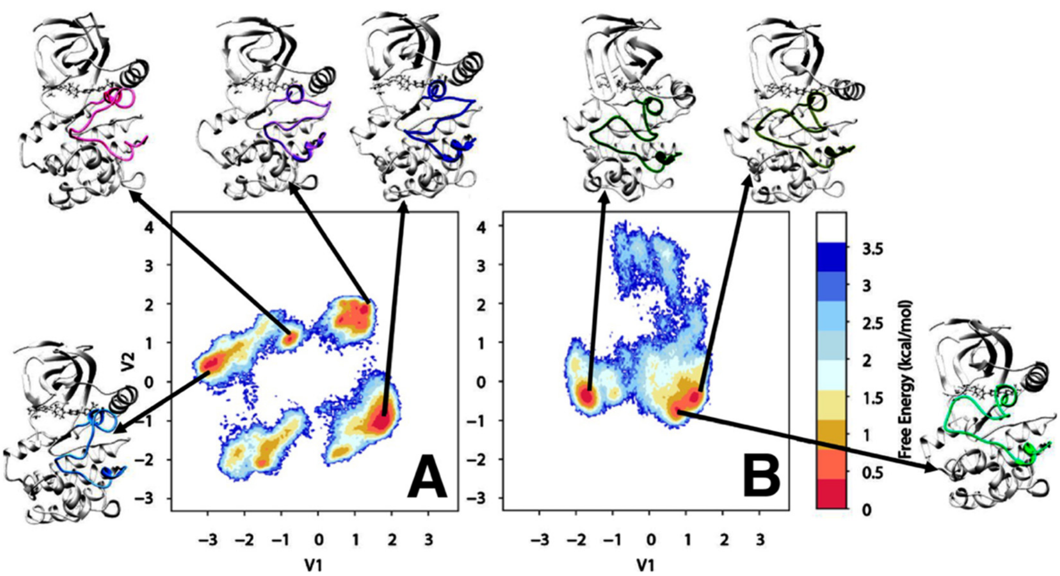


Fig. 6 Free energy landscapes for WNK1 kinase activation loop dynamics projected using order parameters derived from dihedral principal component analysis. (A) Unphosphorylated WNK1. (B) Phosphorylated WNK1. Note that the minima for one proteoform's landscape is distinct from the other's, where any regions of overlapped sampling experience opposite favorability depending on phosphorylation state. Adapted with permission from N. A. Jonniya, M. F. Sk, and P. Kar, *ACS Omega*, 2019, **4**, 17404–17416.[230] Copyright 2022 American Chemical Society.

simulations have been used to facilitate unfolding of covalently modified proteins within the nanosecond timescale, suggesting how different helices stabilize binding interactions involving covalent modification by projecting helical unfolding FELs.[231] High temperature simulations can also facilitate FEL exploration by lowering transition barriers. Work from the Gervasio group done in 2017 combined high temperature unbiased MD with parallel tempering metadynamics to obtain more than 115 μs of total simulation data on p38α kinase activation upon phosphorylation.[232] Metadynamics simulations of up to 25 μs each for unphosphorylated, apo and holo dually phosphorylated p38α resolved conformational FELs were performed, where the order parameters used were distances distinguishing the active from inactive conformations (Fig. 7). These p38α kinase FEL landscapes demonstrated that the unphosphorylated apo protein is unable to sample the active conformation, although phosphorylated apo protein can but at a much higher energetic cost. However, introduction of ATP-Mg^{2+} encouraged deeper minima formation for phosphorylated p38α kinase surrounding an active conformation, whereas binding of ATP-Mg^{2+} and the MK2 docking peptide drastically smoothened barriers between and deepened minima for phosphorylated intermediate and active states. That is, p38α kinase phosphorylation was shown to decrease the probability of inactive state conformations during metadynamics simulations. To confirm the thermo-stability or -instability of the metastable states seen for each system setup during metadynamics simulations, Gervasio and coworkers ran ten independent 1 μs-long classical MD simulations at 380 K from the respective p38α kinase crystal structures, extending simulation durations for any runs which happened to cross barriers into minima observed from metadynamics simulations. The independent unbiased runs showed strong agreement with the metadynamics results, as projecting the time-based progression of the independent data runs over the FELs showed that classical simulations resided in the aforementioned minima for longer periods of time.[231]

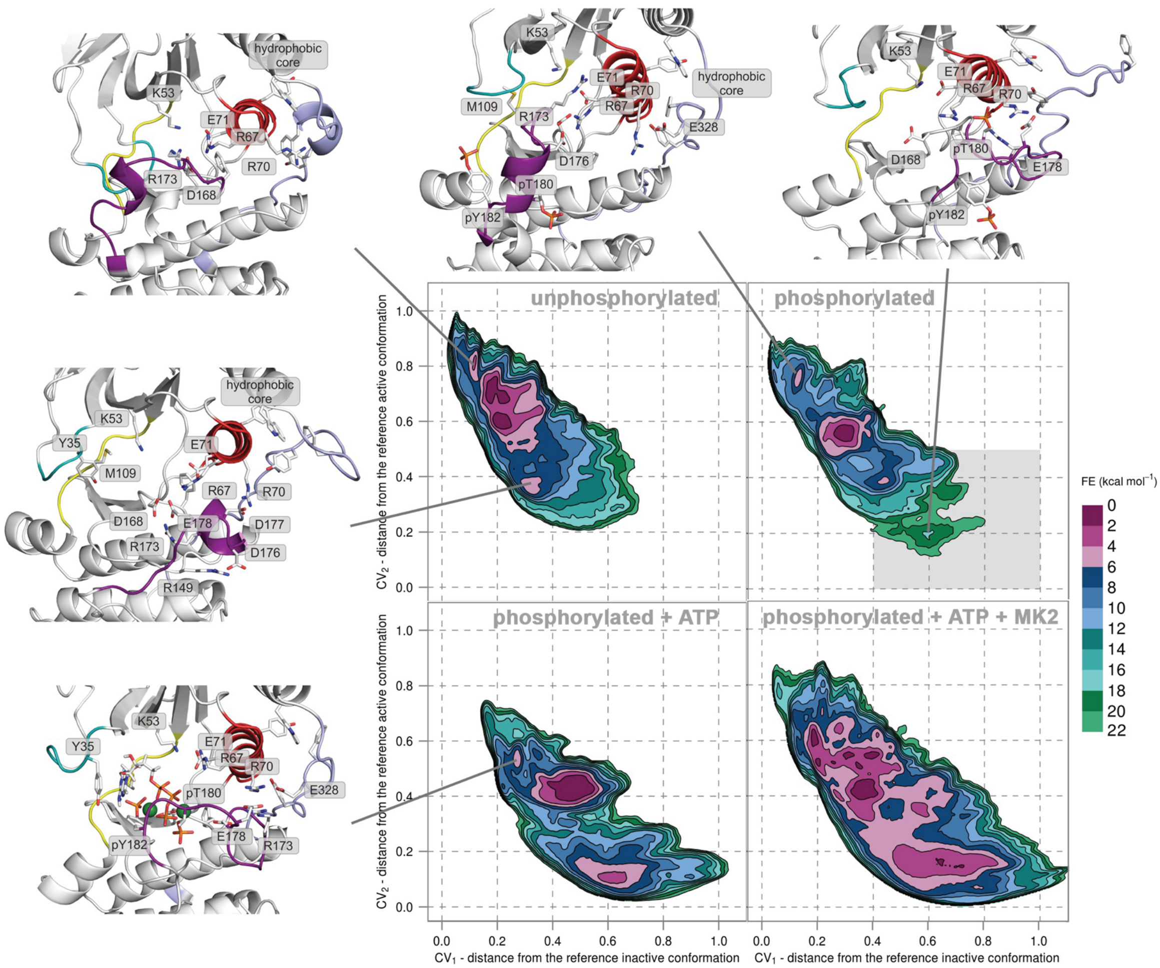


Fig. 7 Free energy surface for differentially phosphorylated p38α proteoforms. The shared axes indicate functions representing the distance from the reference inactive (CV_1) and active (CV_2) structures for p38α. Representative structures from the reference publication are included to visualize differences between active- and inactive-like p38α states. Reproduced with permission from A. Kuzmanic, L. Sutto, G. Saladino, A. R. Nebreda, F. L. Gervasio and M. Orozco, *eLife*, 2017, **6**, e22175.[232] This article was originally published and distributed under the terms of the Creative Commons Attribution License, permitting unrestricted reproduction and adaptation provided proper crediting to author and source. Copyright 2022 eLife Sciences Publications Ltd.

While there was agreement between the biased and unbiased datasets, the conformational requirements for existing in a relatively active *versus* inactive conformation were further supported by FEL exploration from longer metadynamics trajectories. Otherwise, reliance on just the respective crystal structure conformations may have made comparisons between simulated states too stringent, possibly limiting how well phosphorylation was understood to impact extent of p38α activation.[231]

Other biased methods can be used to explore how PTMs affect transitions between metastable states. The impact of seven different combinatoric phosphorylation states on the activation and ATP-binding of plant receptor kinase BAK1 was explored *via* independent 500 ns Gaussian accelerated MD trajectories in 2020.[233] Comparing these phosphorylated proteoforms to unphosphorylated BAK1, the activation loop in unphosphorylated BAK1 was found to have the largest computed root-mean-square fluctuations. By contrast, the activation loop root-mean-square fluctuations for each of the other phosphorylated BAK1 proteoforms was reduced, suggesting a stabilizing role. Specifically, Thr455 phosphorylation appeared to be of secondary importance, further stabilizing the activation loop only when Thr450 was also phosphorylated. Comparison of each of the FELs, simulations involving pThr450 in any PTM combination always resulted in free energy minima surrounding an active BAK1 conformation. Thus, phosphorylation of Thr450 was suggested to be essential for stabilizing the BAK1 activation loop in an active conformation. Estimates of unbiased potentials of mean force through multistate Bennett acceptance ratio reweighting demonstrated that pThr450 uniquely engaged in salt bridges in each BAK1 phosphorylation scenario.[233]

Using the phosphorylated and unphosphorylated cystatin and NtrC crystal structures as case study proteins, the Wolynes group implemented an associated memory Hamiltonian for transferable structure prediction to model the effect of phosphorylation on each protein's folding FEL.[234] Here, the authors mutated the residue occupying the phosphorylation site into glutamic acid, where the Hamiltonian-based method enabled ''supercharging'' of the phosphorylation mimic. For cystatin, these Hamiltonian simulations predicted that phosphorylation makes transitions between the unfolded and folded states more energetically favorable for a simple two-state folding model (Fig. 8A and B). Principal component analysis of cystatin snapshots from both the phosphorylated and unphosphorylated simulation datasets showed that transitionary semi-unfolded states were most similar, suggesting that cystatin is likely phosphorylated when occupying an intermediary structure somewhere closer to the unfolded state (Fig. 8C). Meanwhile, NtrC FELs derived from the Hamiltonian simulations suggested that NtrC folding proceeds through a three-state folding model with a well-defined intermediate structure, regardless of phosphorylation status (Fig. 6D and E). For NtrC, the largest extent of overlap for principal component analysis of simulation snapshots showed that the unfolded states for phosphorylated *versus* unphosphorylated states are most similar, suggesting that phosphorylation would likely occur on unfolded NtrC (Fig. 8F).[234] Here, FELs enabled direct comparison between phosphorylated and unphosphorylated dynamics, where overlap in accessible

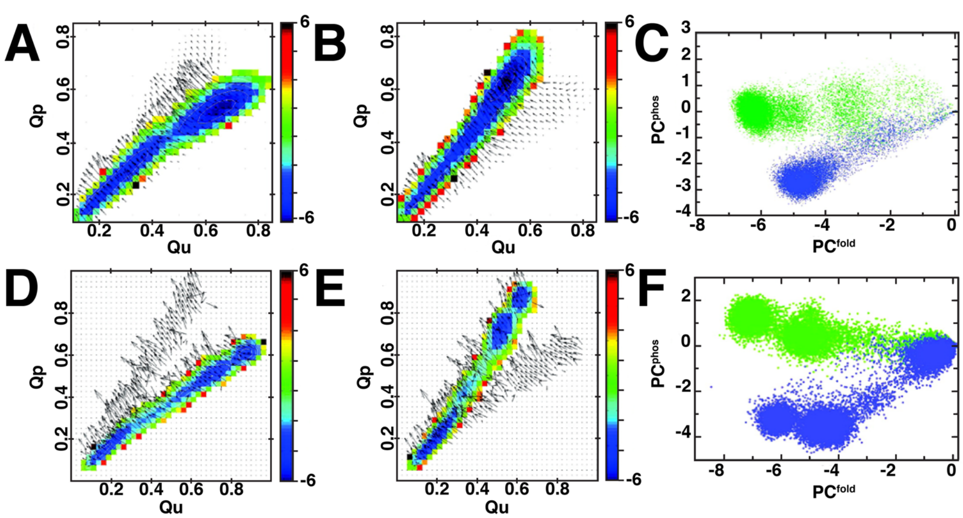


Fig. 8 Structure prediction Hamiltonian free energy landscapes of cystatin folding (A–C) and NrtC folding (D–F) in the phosphorylated and unphosphorylated forms. Global order parameters reflecting how similar simulated structures are to the crystal structures for the phosphorylated (Qp) or unphosphorylated (Qu) proteoforms (A, B, D and E). Principal component analysis (PCA) was performed to project the simulation data with respect to folded state (PC^{fold}) and phosphorylation state (PC^{phos}) (C and F). For PCA plots, green data describes the unphosphorylated proteoform while purple data describes the phosphorylated proteoform. Cystatin folding was found to exhibit two-state behavior for the unphosphorylated (A) and phosphorylated (B) forms. Cystatin PCA suggested that phosphorylation could occur along a slightly unfolded transitionary state, due to the overlap in data sampled (C). NrtC folding was found to exhibit three-state behavior for the unphosphorylated (D) and phosphorylated (E) forms. NrtC PCA suggested that phosphorylation must occur in a completely unfolded state based on overlap seen in the landscape (F). Adapted with permission from J. Latzer, T. Shen, and P. G. Wolynes, *Biochemistry*, 2008, **47**, 2110–2122.[234] Copyright 2022 American Chemical Society.

conformations suggested which kinetic states could undergo phosphorylation *in cellulo*.

5.3. Aggregate sampling for elucidating the impact of PTMs on conformational dynamics

Another way to summarize a biomolecular process through FEL coverage is to run multiple parallel trajectories and then project the aggregate data against appropriate order parameters. Note that kinetic reweighting schemes (*e.g.*, MSMs, milestoning) should accompany aggregate dataset projections, as even the selection of seed structures for classical MD trajectories does bias the resulting FEL.

5.3.1. Ensemble-based sampling of short timescale processes using classical molecular dynamics. Aggregate sampling approaches have been applied to study the effects of PTMs. One flavor of aggregate sampling studies is to run several trajectories from any given point and to then evaluate the total data based on some order parameters. For example, up to 5 μs of aggregate simulation for site-specific and additive villin headpiece carbonylation simulations were performed to examine the effects of carbonylation on villin thermostability in 2011.[235] Although time-dependent analyses demonstrated consistent increases in root-mean-square deviation; increased solvent accessible surface area within the hydrophobic core; and decreases in alpha-helical content as a result of carbonylation, the exact conformational effects differentiating villin headpiece native topology from carbonylated variations of villin could best be discerned using FEL projections.[235] Specifically, FELs comparing molecular hydrophobicity potential *versus* the sum of distances between phenylalanines occupying the villin hydrophobic core demonstrated that combinatoric carbonylations experienced a distinct metastable state, but could sample molecular hydrophobicity potentials and hydrophobic core compactness indicative of native or fully carbonylated minima. However, the opposite is not true, as the native and fully carbonylated villin structures predominantly occupied metastable states at either end of the distribution.[235]

5.3.2. Markov state models on distinguishing PTM-based thermodynamic and kinetic changes. Another flavor of aggregate sampling study designs is to use adaptive sampling to explore a conformational landscape. In adaptive sampling, the MD practitioner guides FEL exploration by running multiple, short, and parallel, unbiased trajectories in tandem as part of a "round" of simulation.[236–239] Seed structures for subsequent rounds are selected based on certain selection criteria that help facilitate efficient exploration of a landscape based on some order parameters. Methodology has been proposed for making seed state selection as statistically unbiased as possible, as adaptive sampling can use any frame from a previous trajectory as a starting seed for the next round.[240–244] Because adaptive sampling regimes incorporate selection of multiple trajectories per round based on preexisting landscape coverage, they can improve statistical characterization of rare high energy regions between minima; traditionally, long trajectories struggle to capture rare events in more than one transition. Adaptive sampling workflows – and their representation through FELs – are easily amenable to Markov state model (MSM) generation.[41,185–187,245–251]

MSMs and adaptive sampling have been used to study the structural and regulatory effects of PTM on protein function. What matters most in these types of studies is that there is sufficient coverage of the FEL. In a 2017 study on the effect of *S*-glutathionylation on the plant receptor kinase BAK1, the Shukla group first used accelerated MD to accumulate ~22 μs of simulation data, from which starting structures were taken to seed an aggregate 132 μs of classical MD through an adaptive sampling workflow.[252] For this study, four simulation systems were constructed and compared: nonglutathionylated BAK1 core kinase domain (BAK1-SH) as a reference system, and then three singly glutathionylated on each possible glutathionylation site. After MSM construction, the authors quantified the global effects of *S*-glutathionylation on BAK1 conformational dynamics using the Kullback–Leibler divergence, a measure of similarity between two probability distributions.[253] These analyses suggested that one of the glutathionylation sites, Cys408, may serve as an inhibitory *S*-glutathionylation site responsible for the decrease of activity in ABK1. Firstly, glutathionylation on C408 caused dramatic effects on BAK1 structure throughout the kinase domain, whereas little changes were observed during the other glutathionylation simulation datasets. Furthermore, C408 glutathionyation shifted the free energy landscapes away from the active-like state, while modifications of other sites retained most features of the unmodified protein. Here, adaptive sampling-based exploration of the BAK1 kinase activation FEL provided sufficient input data for explaining how *S*-glutathionylation allosterically controlled kinase activity.[252]

The Shukla group constructed an MSM from ~320 μs of aggregate MD data on plant abscisic acid (ABA) receptors to delineate the role of tyrosine nitration ABA signaling inhibition during 2019.[254] Comparing the conformational FELs of the PYL5 receptor with and without nitration revealed that ABA cannot bind to PYL5 after tyrosine nitration (Fig. 9). Instead, ABA was stabilized by multiple hydrogen bonds formed by residues surrounding one of the nitrotyrosine residues. Tyrosine nitration was found to significantly shift nearby residue backbone positions when compared to unmodified PYL5. Collectively, tyrosine nitration of the PYL5 receptor rearranged the binding pocket to prevent ABA from reaching the final binding site and therefore inhibiting ABA signaling.[254] Although MSMs offer a kinetic framework based on the aggregate stitching of individual trajectories, their resulting transition probabilities calculated for interconversion between different clustered microstates can be used to instantiate a single trajectory. Often sampled using Monte Carlo simulations along the MSM, a single unbiased trajectory can then be generated which directs a biomolecular system from one conformation to another, all while simultaneously capturing each of the transitions identified along the specified path. The Shukla group's PYL5 receptor study used Monte Carlo sampling to generate a 3 ms trajectory depicting unmodified PYL5 activation.[254] While the same analysis was not performed on nitrated PYL5 because its FEL did not sample an activated state, trajectories formed from MSM-based Monte Carlo sampling could be used to consolidate data from aggregate sampling for time-dependent analyses.

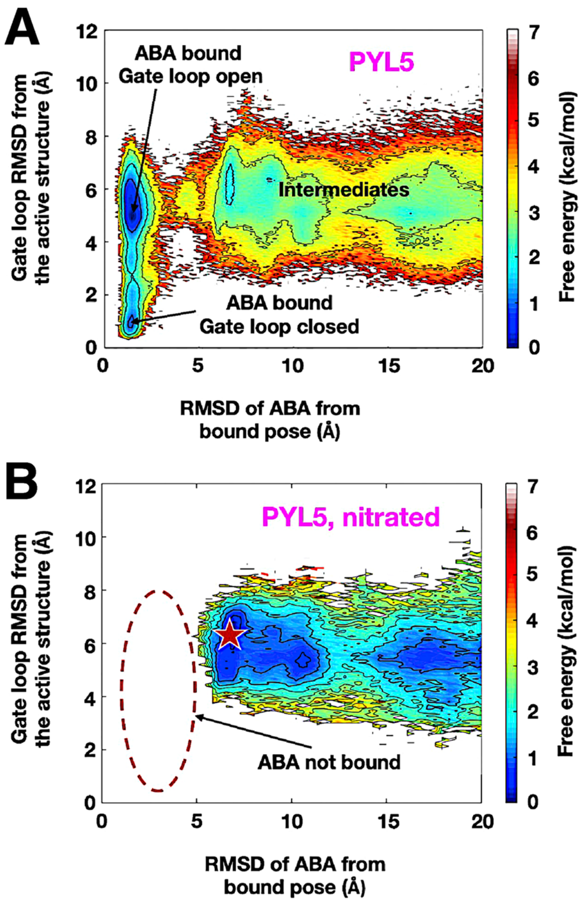


Fig. 9 The effects of tyrosine nitration on ABA binding. The free energy landscapes of ABA binding to the (A) unmodified and (B) nitrated PYL5 receptors. The structure closest to the active ABA-bound pose, labeled as a red star in (B). Adapted with permission from S. Shukla, C. Zhao, and D. Shukla, *Structure*, 2019, **27**, 692–702.e3.[254] Copyright 2022 Elsevier.

In addition to demonstrating how PTMs perturb the likelihood of a certain conformation being accessed, MSMs can also provide information about the timescale for kinetic transition between different macrostates observed along a FEL. The Shukla group did so by implementing transition path theory in their NRT1.1 phosphorylation study.[255,256] The plant nitrate transporter NRT1.1 is regulated by phosphorylation. With affinity inverse to the levels of environmental nitrate, NRT1.1 is either a low-affinity transporter during saturating nitrate conditions or a high-affinity transporter during desaturating nitrate conditions.[257] Additionally, the phosphorylation state determines the oligomeric state of NRT1.1, as the transporter is known to dimerize when in the unphosphorylated state.[258,259] To explore the effects of phosphorylation on enhanced transport and transporter oligomerization, the Shukla group performed unbiased MD simulations for four systems: unphosphorylated NRT1.1 (UnpNRT1.1) ($\sim$9 μs) and phosphorylated (pNRT1.1) ($\sim$5 μs) dimer, as well as UnpNRT1.1 ($\sim$142 μs) and pNRT1.1 ($\sim$63 μs) monomer.[255] Cross-correlation analyses among all Cα atoms within each system demonstrated that dynamic coupling disappeared at the dimeric interface for pNRT1.1 *versus* UnpNRT1.1 dimer (Fig. 10A–C). Meanwhile, dynamic correlation within each monomer in pNRT1.1 was enhanced compared with the UnpNRT1.1 dimer. Thus, phosphorylation appeared to decouple the dimer and allow the two monomers to behave independently. Increasing independent behavior by each participating monomer enhanced their respective structural flexibility, leading to a higher transport rate. Using transition path theory calculations, the average time required for one complete cycle from inward-facing to outward-facing conformations was 5 $\pm$ 2 μs and 18 $\pm$ 3 μs for pNRT1.1 and UnpNRT1.1 monomer, respectively. FELs further reflect these differences in transport rate (Fig. 10D and E). UnpNRT1.1 and pNRT1.1 monomer FELs show how phosphorylation stabilized the outward-facing state while eliminating unnecessary intermediate state, thereby facilitating transport by hastening the transition from inward-facing to outward-facing states. Overall, this study suggested that phosphorylation accelerates necessary conformational transitions, resulting in a higher transport rate.[255]

5.4. Summary of PTM simulations using ensemble-based approaches for free energy landscape construction

Studies implementing aggregate sampling approaches to cover a FEL still fundamentally provide similar insights as studies strictly following single trajectories and their time-dependent analyses.

As the Shukla group constructed MSMs on over 600 μs of aggregate simulation data to demonstrate that Thr276-phosphorylated human serotonin transporter experienced lower relative free energy barriers for transitions between occluded to inward-facing states, the basis for this lower free energy barrier was due to rearrangements in nearby hydrogen-bond networks lining the intracellular gate.[260] Time-dependent approaches to studying PTM with MD simulation have indeed declared similar results; changes in hydrogen bonding networks or sidechain orientations are indeed the immediate response mechanism to local PTM installation, where these changes may allosterically communicate the need for long-range structural rearrangements as well. But what separates previous studies from those implementing aggregate sampling and/or MSMs is that the latter works are able to define what constitutes a specific conformational state, as well as identify the path by which one conformational state transitions into another. That is, aggregate simulation studies can evaluate the thermostability of different conformations through more rigorous statistical sampling. Based on the ergodic assumption that all MD trajectories will eventually sample the same phase space for a given biomolecular process, aggregate sampling just uncovers a greater extent of the stationary density for the process under inspection. MSMs, however, also provide kinetic insights on the evolution of a given molecular process. A specific PTM may not alter the phase space but instead change the kinetics required to traverse across FELs. Trajectory-driven results can offer a relative perspective on the residence time for a given molecular event, but are otherwise unable to provide kinetic rates for transitioning between metastable states separated by larger extents of conformational change. Said calculations do come at a cost, requiring more involved simulation and analyses

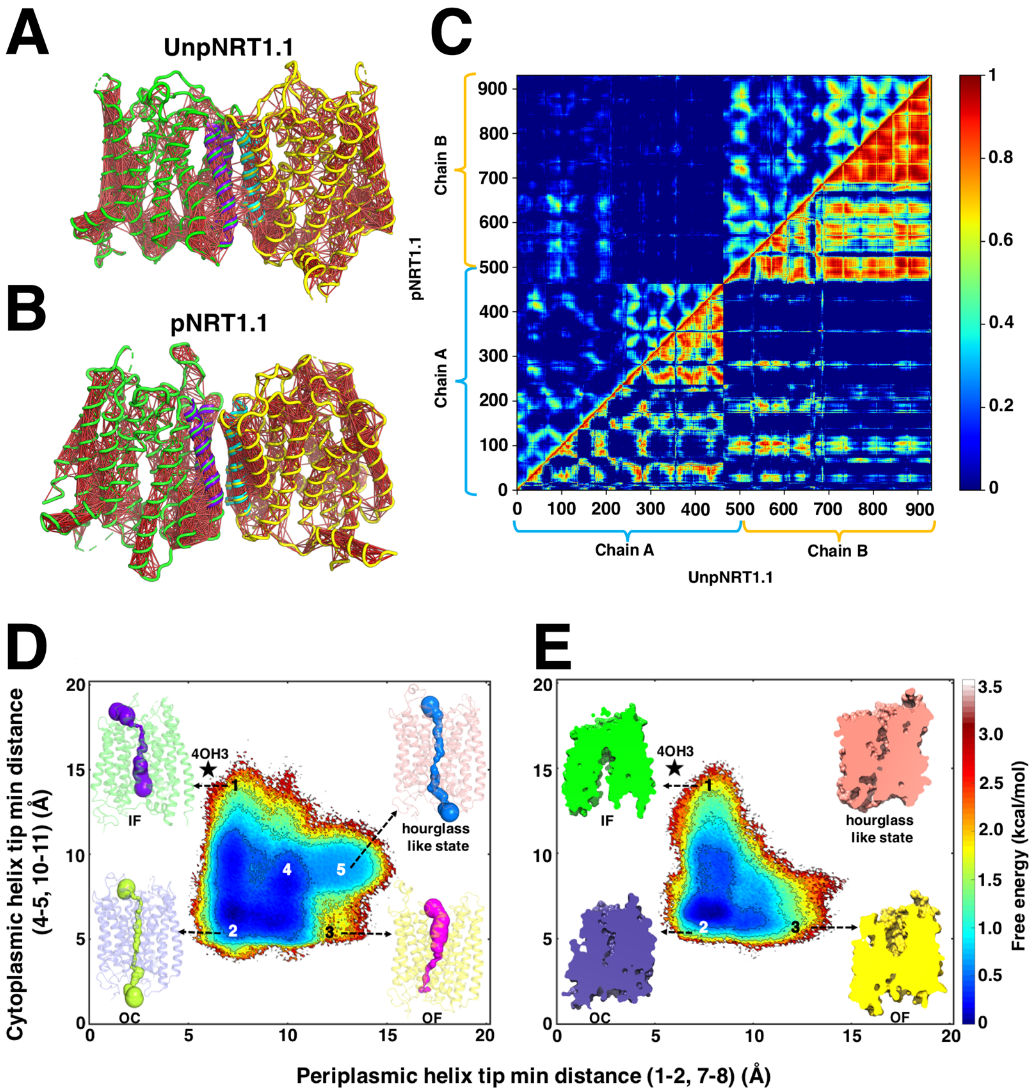


Fig. 10 The effects of phosphorylation on NRT1.1. (A and C) Dynamic cross-correlation matrix (DCCM) analysis for all alpha carbons in UnpNRT1.1 and (B and C) pNRT1.1 dimer. DCCM results displayed in loop representation for (A) and (B). Dimer interface is colored purple and cyan for Chain A and B, while the remaining parts are colored in green and yellow, respectively. Red lines linking residue pairs represent strong dynamic pairwise coupling between residues. (D) Conformational free energy landscapes for UnpNRT1.1 monomer. (E) Conformational free energy landscapes for pNRT1.1 monomer. Intermediate states are labeled as (1) inward-facing (IF), (2) occluded (OC), (3) outward-facing (OF), (4) partial IF-OF (inward-outward facing), and (5) hourglass-like state, respectively. Adapted with permission from B. Selvam, J. Feng, and D. Shukla, *bioRxiv*, 2022.[255] Permission directly granted from authors for use of preprint figures.

workflows, often accompanied with greater computational resource needs. Additionally, uncertainty estimates and other postprocessing for statistical validation are necessary to ensure convergence around ensemble averages.[42,49]

6. Outlooks on PTM modeling and simulations

It has proven difficult to say exactly how any given PTM universally affects some proteins' structures and functions. Experimental throughput makes characterization at atomistic resolution difficult, leaving much of the burden to molecular modeling for explaining cause-and-effect relationships between site-specific PTMs, structural changes, and altered protein function. Luckily, molecular modeling and simulation are perfectly poised to embrace the challenge of PTM characterization laid ahead. Structural and informatic modeling of PTMs in kinases has shown their phosphorylation sites to be relatively conserved, suggesting a potential for *in silico* PTM characterizations to at least universally apply to proteins within the same family.[261] Still, proteins and their covalently modified proteoforms need to be systematically studied so that the relationship between regulatory roles and structural effects of PTMs can be fundamentally understood. Much of the advances seen in machine learning capabilities have been applied towards the prediction and identification of PTMs and their respective sites for covalent modification. As the input data for PTM prediction and identification is based on mass spectrometry results, machine learning-based predictions on functional characterization will require larger datasets focused on characterizing PTM function. In this sense, molecular dynamics simulations are unrivaled in their ability to generate large swaths of data from which distributions of molecular descriptors can be compared to covalent modification status using regression-based models.

Changing focus from simulation outputs, innovation in computing capabilities means that simulation inputs and their parameterization can also be expected to improve. PTM chemical diversity must be appropriately covered by all force field sets for more amenable simulated applications to any research problem. The GROMOS force field library offers tremendous coverage of PTMs. Still, MD practitioners gravitate towards specific force field libraries when tackling specific research problems. For instance, plans to simulate membrane proteins may result in CHARMM force field parameterization given its fantastic coverage of lipid diversity with over 2000 lipid molecules in different starting conformations for lipid bilayer building.[262,263] Still, CHARMM parameterization with compatible force fields for less common PTMs are not necessarily automatically, nor easily, incorporated into existing computational tools as is seen per the GROMOS library.[10,264] As observed in the development of a membrane protein-specific OPLS-AA force field,[176] it is only a matter of time before force field libraries recognize their application-specific deficiencies and offer the same research opportunities for modeling PTMs with equivalent ease of access. That is, until efforts like the Open Force Field Initiative and machine learning-based force field libraries can consistently offer the accuracy of *ab initio* methods without incurring the same computational cost.[265–267] The same argument applies to the efficient use of polarizable force fields when dealing with PTM protonation state.[268,269]

Aside from parameterization improvements, advances in structure prediction methodology raise the possibility of predicting covalently modified proteoform conformations. Software like DeepMind's AlphaFold, along with the Baker lab's RoseTTaFold, have now revolutionized protein structure prediction with their machine learning-based template-free methods that, in most cases, can rival experimental structure quality.[270–273] How these software are learning from numerous multiple sequence alignments and pairwise evolutionary constraints to generate a distribution of probable models suggests that incorporating the effects of PTMs in state-of-the-art conformer prediction is not so farfetched. Previous efforts using molecular mechanics showed that it was possible to predict protein conformational change upon covalent modification given an unmodified structure.[274] Now, tools like Privateer have already been developed to graft glycan PTMs directly onto AlphaFold predicted structures.[275] Inter-residue impacts of PTM-containing structures in the Protein Data Bank could be learned and included as a stage within these state-of-the-art structure prediction pipelines.[276] Given how a majority of PTMs (excluding ubiquination) occur along intrinsically disordered regions of proteins, integrated modeling of PTMs onto even difficult-to-predict protein structures is more feasible than ever because of AlphaFold.[277] Including post-predictional modifications could offer a unique opportunity for seeding MD simulations from more unique starting structures for more efficient PTM characterization.

Regardless of the inputs provided, future PTM simulation studies that will capture vast free energy landscape coverage should strive to standardize analyses as to leverage their extensive sampling of macrostate population densities. As PTM installment can mirror the effects of mutation in shifting free energy landscapes, performing routine allosteric analyses on vast PTM simulation datasets could make the interpretation of long-range effects of PTMs more systematic across different proteins.[278–281] PTM studies incorporating MSMs should report the exact transition probability matrix weights and calculate macrostate kinetic flux rates *via* transition path theory.[185–187,256] In doing so, how often a PTM proteoform exists in a specific conformation (*e.g.*, active *versus* inactive) can be ascertained. Offering probability distributions relative to the specific activity of a protein after PTM installation could drastically improve the accuracy of systems-scale kinetic analyses for multiscale modeling endeavors.[1,282–284] State-of-the-art structure predictions and faster computing hardware now unleash an array of *in silico* opportunities to realistically characterize PTMs for proteins that lack experimentally resolved structures or are simply understudied.[285] Together, modern MD simulations can join theory and experiments in informing us as to how the structure-dynamic perturbations caused by PTMs impact whole-cell physiology.[283,284]

Despite the opportunities they provide in characterizing PTMs, it is worth recognizing the caveats of molecular simulations. A major caveat to molecular simulations is that the process of covalent bond breakage and formation is extremely expensive to model. This reality is a major driver for the development of machine learning-based force fields so that molecular mechanics simulations can be performed with parameterization done at the accuracy level of quantum mechanics simulations. Thus, simulations are likely unable to provide as valuable insights as experiments when it comes to dynamical modeling for the direct instance of PTM, such as the rate of covalent modification or cellular trafficking. Experimental characterization can further delineate spatio-temporal factors underlying PTM effects which are not easily, if entirely possible to be, modeled using MD simulations. What is a simple task for computational research is often an arduous one for experiments, and *vice versa*. Still, the level of control and atomistic resolution offers a glimpse as to what could be happening given explicit conditions and sufficient sampling.

Throughout this perspective, we have highlighted almost 30 years' worth of PTM research using *in silico* molecular modeling and simulation techniques. Much of this work has demonstrated how molecular dynamics simulations can characterize the structural consequences induced by PTMs. With recent computing developments and simulation methodologies enabling MD to be more efficient than ever before, the future of PTM-based molecular dynamics research can more easily employ currently established enhanced sampling and structure prediction methodology. Computational research efforts surrounding PTMs and their modeling will shift towards identifying new conformations caused by PTMs, and the frequency at which these conformations are sampled within the stationary density for the specified proteoform. Systematic computational characterization leveraging larger extents of sampling will lead to systematic translational gains in experimental efforts surrounding therapeutic developments and bioengineering. The predictive power of MD simulations is limited insofar as the realism used to model the system. As efforts have been cast to make the modeling of lipid bilayers and membrane proteins match experiments with improved realism,[286] so too shall efforts surrounding PTM simulation strive to make our biophysical understanding of processes regulating protein function more complete.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgements

The authors would like to thank Matthew Chan for providing comments on earlier drafts of this manuscript. Other Shukla Group members – Prateek Bansal, Jiming Chen, Tanner Dean, Soumajit Dutta – are thanked for recommending additional references for the completion of this manuscript. D. S. acknowledges support from the Foundation for Food and Agriculture Research *via* the New Innovator Award and NSF Early CAREER Award (NSF-MCB-1845606), as well as the National Institute of General Medical Sciences (NIGMS; National Institutes of Health, NIH) (Award No. R35GM142745).

References

1 S. Prabakaran, G. Lippens, H. Steen and J. Gunawardena, Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding, *Wiley Interdiscip. Rev. Syst. Biol. Med.*, 2012, **4**, 565–583.
2 D. Bradley, The evolution of post-translational modifications, *Curr. Opin. Genet. Dev.*, 2022, **76**, 101956.
3 A. H. Millar, J. L. Heazlewood, C. Giglione, M. J. Holdsworth, A. Bachmair and W. X. Schulze, The scope, functions, and dynamics of posttranslational protein modifications, *Annu. Rev. Plant Biol.*, 2019, **70**, 119–151.
4 C. T. Walsh, *Posttranslational Modification of Proteins: Expanding Nature's Inventory*, Roberts and Company Publishers, Greenwood Village, Colorado, 2005.
5 Uniprot Consortium, ptmlist.txt, https://www.uniprot.org/docs/ptmlist, (accessed May 9, 2022).
6 C. T. Walsh, S. Garneau-Tsodikova and G. J. Gatto Jr., Protein posttranslational modifications: the chemistry of proteome diversifications, *Angew. Chem., Int. Ed*, 2005, **44**, 7342–7372.
7 S. Ramazi and J. Zahiri, Post-translational modifications in proteins: resources, tools and prediction methods, *Database*, 2021, **2021**, baab012.
8 S. H. Spoel, Orchestrating the proteome with post-translational modifications, *J. Exp. Bot.*, 2018, **69**, 4499–4503.
9 B. Macek, K. Forchhammer, J. Hardouin, E. Weber-Ban, C. Grangeasse and I. Mijakovic, Protein post-translational modifications in bacteria, *Nat. Rev. Microbiol.*, 2019, **17**, 651–664.
10 C. Margreitter, D. Petrov and B. Zagrovic, Vienna-PTM web server: a toolkit for MD simulations of protein post-translational modifications, *Nucleic Acids Res.*, 2013, **41**, W422–W426.
11 W. A. Lim and T. Pawson, Phosphotyrosine Signaling: evolving a new cellular communication system, *Cell*, 2010, **142**, 661–667.
12 A. S. Venne, L. Kollipara and R. P. Zahedi, The next level of complexity: crosstalk of posttranslational modifications, *Proteomics*, 2014, **14**, 513–524.
13 M. Leutert, S. W. Entwisle and J. Villén, Decoding post-translational modification crosstalk with proteomics, *Mol. Cell. Proteomics*, 2021, **20**, 100129.
14 R. Nussinov, C.-J. Tsai, F. Xin and P. Radivojac, Allosteric post-translational modification codes, *Trends Biochem. Sci.*, 2012, **37**, 447–455.
15 M. R. Larsen, M. B. Trelle, T. E. Thingholm and O. N. Jensen, Analysis of posttranslational modifications of proteins by tandem mass spectrometry, *Biotechniques*, 2006, **40**, 790–798.
16 A. Yang, K. Cho and H.-S. Park, Chemical biology approaches for studying posttranslational modifications, *RNA Biol.*, 2018, **15**, 427–440.
17 J. Horne and D. Shukla, Recent advances in machine learning variant effect prediction tools for protein engineering, *Ind. Eng. Chem. Res.*, 2022, **61**, 6235–6245.
18 M. Karplus and J. A. McCammon, Molecular dynamics simulations of biomolecules, *Nat. Struct. Mol. Biol.*, 2002, **9**, 646–652.
19 R. O. Dror, R. M. Dirks, J. P. Grossman, H. Xu and D. E. Shaw, Biomolecular simulation: a computational microscope for molecular biology, *Annu. Rev. Biophys.*, 2012, **41**, 429–452.
20 S. A. Hollingsworth and R. O. Dror, Molecular dynamics simulation for all, *Neuron*, 2018, **99**, 1129–1143.
21 M. Reif and M. Zacharias, in *Biomolecular and Bioanalytical Techniques*, ed. V. Ramesh, John Wiley & Sons, Ltd, 2019, pp. 501–535.
22 T. Jiang, P.-C. Wen, N. Trebesch, Z. Zhao, S. Pant, K. Kapoor, M. Shekhar and E. Tajkhorshid, Computational dissection of membrane transport at a microscopic level, *Trends Biochem. Sci.*, 2020, **45**, 202–216.
23 J. Feng, J. Chen, B. Selvam and D. Shukla, Computational microscopy: revealing molecular mechanisms in plants using molecular dynamics simulations, *Plant Cell*, 2019, **31**, tpc.119.tt1219.
24 E. Braun, J. Gilmer, H. B. Mayes, D. L. Mobley, J. I. Monroe, S. Prasad and D. M. Zuckerman, Best Practices for

Foundations in Molecular Simulations [Article v1.0], *Living J. Comput. Mol. Sci.*, 2019, **1**(1), 5957.

25 M. C. Zwier and L. T. Chong, Reaching biological timescales with all-atom molecular dynamics simulations, *Curr. Opin. Pharmacol.*, 2010, **10**, 745–752.

26 S. Kalyaanamoorthy and Y.-P. P. Chen, Modelling and enhanced molecular dynamics to steer structure-based drug discovery, *Prog. Biophys. Mol. Biol.*, 2014, **114**, 123–136.

27 M. S. Friedrichs, P. Eastman, V. Vaidyanathan, M. Houston, S. Legrand, A. L. Beberg, D. L. Ensign, C. M. Bruns and V. S. Pande, Accelerating molecular dynamic simulation on graphics processing units, *J. Comput. Chem.*, 2009, **30**, 864–872.

28 A. W. Götz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand and R. C. Walker, Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born, *J. Chem. Theory Comput.*, 2012, **8**, 1542–1555.

29 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX*, 2015, **1–2**, 19–25.

30 J. C. Phillips, D. J. Hardy, J. D. C. Maia, J. E. Stone, J. V. Ribeiro, R. C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, R. McGreevy, M. C. R. Melo, B. K. Radak, R. D. Skeel, A. Singharoy, Y. Wang, B. Roux, A. Aksimentiev, Z. Luthey-Schulten, L. V. Kalé, K. Schulten, C. Chipot and E. Tajkhorshid, Scalable molecular dynamics on CPU and GPU architectures with NAMD, *J. Chem. Phys.*, 2020, **153**, 044130.

31 P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks and V. S. Pande, OpenMM 7: rapid development of high performance algorithms for molecular dynamics, *PLoS Comput. Biol.*, 2017, **13**, e1005659.

32 M. Shirts and V. S. Pande, Screen savers of the world unite!, *Science*, 2000, **290**, 1903–1904.

33 A. L. Beberg, D. L. Ensign, G. Jayachandran, S. Khaliq and V. S. Pande, *2009 IEEE International Symposium on Parallel & Distributed Processing*, 2009, pp. 1–8.

34 B. Bode, M. Butler, T. Dunning, W. Gropp, T. Hoefler, W. Hwu and W. Kramer, in *Contemporary HPC Architectures*, ed. J. Vetter, Chapman and Hall, London, England, 2013.

35 J. Towns, T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, R. Roskies, J. R. Scott and N. Wilkens-Diehr, XSEDE: accelerating scientific discovery, *Comput. Sci. Eng.*, 2014, **16**, 62–74.

36 D. Stanzione, J. West, R. T. Evans, T. Minyard, O. Ghattas and D. K. Panda, Frontera, The evolution of leadership computing at the National Science Foundation, *Practice and Experiience in Advanced Research Computing*, 2020, pp. 106–111.

37 D. E. Shaw, J. P. Adams, A. Azaria, J. A. Bank, B. Batson, A. Bell, M. Bergdorf, J. Bhatt, J. A. Butts, T. Correia, R. M. Dirks, R. O. Dror, M. P. Eastwood, B. Edwards, A. Even, P. Feldmann, M. Fenn, C. H. Fenton, A. Forte, J. Gagliardo, G. Gill, M. Gorlatova, B. Greskamp, J. P. Grossman, J. Gullingsrud, A. Harper, W. Hasenplaugh, M. Heily, B. C. Heshmat, J. Hunt, D. J. Ierardi, L. Iserovich, B. L. Jackson, N. P. Johnson, M. M. Kirk, J. L. Klepeis, J. S. Kuskin, K. M. Mackenzie, R. J. Mader, R. McGowen, A. McLaughlin, M. A. Moraes, M. H. Nasr, L. J. Nociolo, L. O'Donnell, A. Parker, J. L. Peticolas, G. Pocina, C. Predescu, T. Quan, J. K. Salmon, C. Schwink, K. S. Shim, N. Siddique, J. Spengler, T. Szalay, R. Tabladillo, R. Tartler, A. G. Taube, M. Theobald, B. Towles, W. Vick, S. C. Wang, M. Wazlowski, M. J. Weingarten, J. M. Williams and K. A. Yuh, Anton 3: twenty microseconds of molecular dynamics simulation before lunch, Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, 2021, vol. 1, pp. 1–11.

38 S. Decherchi and A. Cavalli, Thermodynamics and kinetics of drug-target binding by molecular simulation, *Chem. Rev.*, 2020, **120**, 12788–12833.

39 F. Noé, A. Tkatchenko, K.-R. Müller and C. Clementi, Machine learning for molecular simulation, *Annu. Rev. Phys. Chem.*, 2020, **71**, 361–390.

40 A. Glielmo, B. E. Husic, A. Rodriguez, C. Clementi, F. Noé and A. Laio, Unsupervised learning methods for molecular simulation data, *Chem. Rev.*, 2021, **121**, 9722–9758.

41 K. A. Konovalov, I. C. Unarta, S. Cao, E. C. Goonetilleke and X. Huang, Markov state models to study the functional dynamics of proteins in the wake of machine learning, *JACS Au*, 2021, **1**, 1330–1341.

42 A. Grossfield, P. N. Patrone, D. R. Roe, A. J. Schultz, D. Siderius and D. M. Zuckerman, Best practices for quantification of uncertainty and sampling quality in molecular simulations [Article v1.0], *Living J. Comput. Mol. Sci.*, 2019, **1**, 5067.

43 J. C. Gumbart, B. Roux and C. Chipot, Efficient determination of protein–protein standard binding free energies from first principles, *J. Chem. Theory Comput.*, 2013, **9**, 3789–3798.

44 L. Wang, Y. Wu, Y. Deng, B. Kim, L. Pierce, G. Krilov, D. Lupyan, S. Robinson, M. K. Dahlgren, J. Greenwood, D. L. Romero, C. Masse, J. L. Knight, T. Steinbrecher, T. Beuming, W. Damm, E. Harder, W. Sherman, M. Brewer, R. Wester, M. Murcko, L. Frye, R. Farid, T. Lin, D. L. Mobley, W. L. Jorgensen, B. J. Berne, R. A. Friesner and R. Abel, Accurate and reliable prediction of relative ligand binding Potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field, *J. Am. Chem. Soc.*, 2015, **137**, 2695–2703.

45 T. B. Steinbrecher, M. Dahlgren, D. Cappel, T. Lin, L. Wang, G. Krilov, R. Abel, R. Friesner and W. Sherman, Accurate binding free energy predictions in fragment optimization, *J. Chem. Inf. Model.*, 2015, **55**, 2411–2420.

46 K. Lindorff-Larsen, S. Piana, R. O. Dror and D. E. Shaw, How fast-folding proteins fold, *Science*, 2011, **334**, 517–520.
47 E. Suárez, J. L. Adelman and D. M. Zuckerman, Accurate estimation of protein folding and unfolding times: beyond Markov state models, *J. Chem. Theory Comput.*, 2016, **12**, 3473–3481.
48 H. Wu, F. Paul, C. Wehmeyer and F. Noé, Multiensemble Markov models of molecular thermodynamics and kinetics, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, E3221–E3230.
49 S. Wan, R. C. Sinclair and P. V. Coveney, Uncertainty quantification in classical molecular dynamics, *Philos. Trans. R. Soc., A*, 2021, **379**, 20200082.
50 L.-Q. Yang, P. Sang, Y. Tao, Y.-X. Fu, K.-Q. Zhang, Y.-H. Xie and S.-Q. Liu, Protein dynamics and motions in relation to their functions: several case studies and the underlying mechanisms, *J. Biomol. Struct. Dyn.*, 2014, **32**, 372–393.
51 A. S. Moffett and D. Shukla, Using molecular simulation to explore the nanoscale dynamics of the plant kinome, *Biochem. J.*, 2018, **475**, 905–921.
52 E. Papaleo, Integrating atomistic molecular dynamics simulations, experiments, and network analysis to study protein dynamics: strength in unity, *Front. Mol. Biosci.*, 2015, **2**, 28.
53 A. C. Conibear, Deciphering protein post-translational modifications using chemical biology tools, *Nat. Rev. Chem.*, 2020, **4**, 674–695.
54 Z. Li, S. Li, M. Luo, J.-H. Jhong, W. Li, L. Yao, Y. Pang, Z. Wang, R. Wang, R. Ma, J. Yu, Y. Huang, X. Zhu, Q. Cheng, H. Feng, J. Zhang, C. Wang, J. B.-K. Hsu, W.-C. Chang, F.-X. Wei, H.-D. Huang and T.-Y. Lee, dbPTM in 2022: an updated database for exploring regulatory networks and functional associations of protein post-translational modifications, *Nucleic Acids Res.*, 2022, **50**, D471–D479.
55 R. Oughtred, J. Rust, C. Chang, B.-J. Breitkreutz, C. Stark, A. Willems, L. Boucher, G. Leung, N. Kolas, F. Zhang, S. Dolma, J. Coulombe-Huntington, A. Chatr-aryamontri, K. Dolinski and M. Tyers, The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions, *Protein Sci.*, 2021, **30**, 187–200.
56 P. V. Hornbeck, J. M. Kornhauser, S. Tkachev, B. Zhang, E. Skrzypek, B. Murray, V. Latham and M. Sullivan, PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse, *Nucleic Acids Res.*, 2012, **40**, D261–D270.
57 P. Minguez, I. Letunic, L. Parca, L. Garcia-Alonso, J. Dopazo, J. Huerta-Cepas and P. Bork, PTMcode v2: a resource for functional associations of post-translational modifications within and between proteins, *Nucleic Acids Res.*, 2015, **43**, D494–D502.
58 K. Yu, Q. Zhang, Z. Liu, Q. Zhao, X. Zhang, Y. Wang, Z.-X. Wang, Y. Jin, X. Li, Z.-X. Liu and R.-H. Xu, qPhos: a database of protein phosphorylation dynamics in humans, *Nucleic Acids Res.*, 2019, **47**, D451–D458.
59 H. Xu, J. Zhou, S. Lin, W. Deng, Y. Zhang and Y. Xue, PLMD: an updated data resource of protein lysine modifications, *J. Genet. Genomics*, 2017, **44**, 243–250.
60 Z. Liu, Y. Wang, T. Gao, Z. Pan, H. Cheng, Q. Yang, Z. Cheng, A. Guo, J. Ren and Y. Xue, CPLM: a database of protein lysine modifications, *Nucleic Acids Res.*, 2014, **42**, D531–D536.
61 L. Ledesma, E. Sandoval, U. Cruz-Martínez, A. M. Escalante, S. Mejía, P. Moreno-Álvarez, E. Ávila, E. García, G. Coello and F. Torres-Quiroz, YAAM: yeast amino acid modifications database, *Database*, 2018, **2018**, bax099.
62 T. S. Keshava Prasad, R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venugopal, L. Balakrishnan, A. Marimuthu, S. Banerjee, D. S. Somanathan, A. Sebastian, S. Rani, S. Ray, C. J. Harrys Kishore, S. Kanth, M. Ahmed, M. K. Kashyap, R. Mohmood, Y. L. Ramachandra, V. Krishna, B. A. Rahiman, S. Mohan, P. Ranganathan, S. Ramabadran, R. Chaerkady and A. Pandey, Human protein reference database—2009 update, *Nucleic Acids Res.*, 2009, **37**, D767–D772.
63 F. Gnad, J. Gunawardena and M. Mann, PHOSIDA 2011: the posttranslational modification database, *Nucleic Acids Res.*, 2011, **39**, D253–D260.
64 P. Craveur, J. Rebehmed and A. G. de Brevern, PTM-SD: a database of structurally resolved and annotated posttranslational modifications in proteins, *Database*, 2014, **2014**, bau041.
65 Y. Xu, S. Zhang, S. Lin, Y. Guo, W. Deng, Y. Zhang and Y. Xue, WERAM: a database of writers, erasers and readers of histone acetylation and methylation in eukaryotes, *Nucleic Acids Res.*, 2017, **45**, D264–D270.
66 D. Plewczynski, A. Tkacz, L. S. Wyrwicz and L. Rychlewski, AutoMotif server: prediction of single residue post-translational modifications in proteins, *Bioinformatics*, 2005, **21**, 2525–2527.
67 S. Lin, C. Wang, J. Zhou, Y. Shi, C. Ruan, Y. Tu, L. Yao, D. Peng and Y. Xue, EPSD: a well-annotated data resource of protein phosphorylation sites in eukaryotes, *Brief. Bioinform.*, 2021, **22**, 298–307.
68 S. Ullah, S. Lin, Y. Xu, W. Deng, L. Ma, Y. Zhang, Z. Liu and Y. Xue, dbPAF: an integrative database of protein phosphorylation in animals and fungi, *Sci. Rep.*, 2016, **6**, 23534.
69 H. Cheng, W. Deng, Y. Wang, J. Ren, Z. Liu and Y. Xue, dbPPT: a comprehensive database of protein phosphorylation in plants, *Database*, 2014, **2014**, bau121.
70 Kinexus|PhosphoNET, https://www.phosphonet.ca/, (accessed May 10, 2022).
71 K.-Y. Huang, H.-Y. Wu, Y.-J. Chen, C.-T. Lu, M.-G. Su, Y.-C. Hsieh, C.-M. Tsai, K.-I. Lin, H.-D. Huang, T.-Y. Lee and Y.-J. Chen, RegPhos 2.0: an updated resource to explore protein kinase–substrate phosphorylation networks in mammals, *Database*, 2014, **2014**, bau034.
72 H. Dinkel, C. Chica, A. Via, C. M. Gould, L. J. Jensen, T. J. Gibson and F. Diella, Phospho.ELM: a database of phosphorylation sites—Update 2011, *Nucleic Acids Res.*, 2011, **39**, D261–D267.
73 A. Zanzoni, D. Carbajo, F. Diella, P. F. Gherardini, A. Tramontano, M. Helmer-Citterich and A. Via, Phospho3D 2.0: an enhanced database of three-dimensional

structures of phosphorylation sites, *Nucleic Acids Res.*, 2011, **39**, D268–D271.
74 Y. Shi, Y. Zhang, S. Lin, C. Wang, J. Zhou, D. Peng and Y. Xue, dbPSP 2.0, an updated database of protein phosphorylation sites in prokaryotes, *Sci. Data*, 2020, **7**, 164.
75 C. Wang, H. Xu, S. Lin, W. Deng, J. Zhou, Y. Zhang, Y. Shi, D. Peng and Y. Xue, GPS 5.0: an update on the prediction of kinase-specific phosphorylation sites in proteins, *Genom. Proteom. Bioinform.*, 2020, **18**, 72–80.
76 M. Carrascal, D. Ovelleiro, V. Casas, M. Gay and J. Abian, Phosphorylation analysis of primary human T lymphocytes using sequential IMAC and titanium oxide enrichment, *J. Proteome Res.*, 2008, **7**, 5167–5176.
77 Q. Yao, H. Ge, S. Wu, N. Zhang, W. Chen, C. Xu, J. Gao, J. J. Thelen and D. Xu, P3DB 3.0: from plant phosphorylation sites to protein networks, *Nucleic Acids Res.*, 2014, **42**, D1206–D1213.
78 M. Zulawski, R. Braginets and W. X. Schulze, PhosPhAt goes kinases—Searchable protein kinase target information in the plant phosphorylation site database PhosPhAt, *Nucleic Acids Res.*, 2013, **41**, D1176–D1184.
79 K. J. van Wijk, T. Leppert, Q. Sun, S. S. Boguraev, Z. Sun, L. Mendoza and E. W. Deutsch, The Arabidopsis PeptideAtlas: harnessing worldwide proteomics data to create a comprehensive community proteomics resource, *Plant Cell*, 2021, **33**, 3421–3453.
80 D. Wang, D. Liu, J. Yuchi, F. He, Y. Jiang, S. Cai, J. Li and D. Xu, MusiteDeep: a deep-learning based webserver for protein post-translational modification site prediction and visualization, *Nucleic Acids Res.*, 2020, **48**, W140–W146.
81 F. Li, C. Li, T. T. Marquez-Lago, A. Leier, T. Akutsu, A. W. Purcell, A. Ian Smith, T. Lithgow, R. J. Daly, J. Song and K.-C. Chou, Quokka: a comprehensive tool for rapid and accurate prediction of kinase family-specific phosphorylation sites in the human proteome, *Bioinformatics*, 2018, **34**, 4223–4231.
82 N. Blom, S. Gammeltoft and S. Brunak, Sequence and structure-based prediction of eukaryotic protein phosphorylation sites, *J. Mol. Biol.*, 1999, **294**, 1351–1362.
83 N. Blom, T. Sicheritz-Pontén, R. Gupta, S. Gammeltoft and S. Brunak, Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence, *Proteomics*, 2004, **4**, 1633–1649.
84 Y.-H. Wong, T.-Y. Lee, H.-K. Liang, C.-M. Huang, T.-Y. Wang, Y.-H. Yang, C.-H. Chu, H.-D. Huang, M.-T. Ko and J.-K. Hwang, KinasePhos 2.0: a web server for identifying protein kinase-specific phosphorylation sites based on sequences and coupling patterns, *Nucleic Acids Res.*, 2007, **35**, W588–W594.
85 R. Patrick, K.-A. Lê Cao, B. Kobe and M. Bodén, PhosphoPICK: modelling cellular context to map kinase-substrate phosphorylation events, *Bioinformatics*, 2015, **31**, 382–389.
86 J. Song, H. Wang, J. Wang, A. Leier, T. Marquez-Lago, B. Yang, Z. Zhang, T. Akutsu, G. I. Webb and R. J. Daly, PhosphoPredict: a bioinformatics tool for prediction of human kinase-specific phosphorylation substrates and sites by integrating heterogeneous feature selection, *Sci. Rep.*, 2017, **7**, 6862.
87 J. C. Obenauer, L. C. Cantley and M. B. Yaffe, Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs, *Nucleic Acids Res.*, 2003, **31**, 3635–3641.
88 R. Chaudhuri, A. Sadrieh, N. J. Hoffman, B. L. Parker, S. J. Humphrey, J. Stöckli, A. P. Hill, D. E. James and J. Y. H. Yang, PhosphOrtholog: a web-based tool for cross-species mapping of orthologous protein post-translational modifications, *BMC Genom.*, 2015, **16**, 617.
89 H. Zhang, P. Loriaux, J. Eng, D. Campbell, A. Keller, P. Moss, R. Bonneau, N. Zhang, Y. Zhou, B. Wollscheid, K. Cooke, E. C. Yi, H. Lee, E. R. Peskind, J. Zhang, R. D. Smith and R. Aebersold, UniPep—a database for human N-linked glycosites: a resource for biomarker discovery, *Genome Biol.*, 2006, **7**, R73.
90 D. Baycin-Hizal, Y. Tian, I. Akan, E. Jacobson, D. Clark, J. Chu, K. Palter, H. Zhang and M. J. Betenbaugh, GlycoFly: a database of Drosophila N-linked glycoproteins identified using SPEG–MS techniques, *J. Proteome Res.*, 2011, **10**, 2777–2784.
91 D. Baycin-Hizal, Y. Tian, I. Akan, E. Jacobson, D. Clark, A. Wu, R. Jampol, K. Palter, M. Betenbaugh and H. Zhang, GlycoFish: a database of zebrafish N-linked glycoproteins identified using SPEG method coupled with LC/MS, *Anal. Chem.*, 2011, **83**, 5296–5303.
92 J. S. Chauhan, A. H. Bhat, G. P. S. Raghava and A. Rao, GlycoPP: a webserver for prediction of *N*- and *O*-glycosites in prokaryotic protein sequences, *PLoS One*, 2012, **7**, e40155.
93 K. Julenius, A. Mølgaard, R. Gupta and S. Brunak, Prediction, conservation analysis, and structural characterization of mammalian mucin-type *O*-glycosylation sites, *Glycobiology*, 2005, **15**, 153–164.
94 R. Gupta, E. Jung, A. A. Gooley, K. L. Williams, S. Brunak and J. Hansen, Scanning the available Dictyostelium discoideum proteome for O-linked GlcNAc glycosylation sites using neural networks, *Glycobiology*, 1999, **9**, 1009–1022.
95 C.-H. Chien, C.-C. Chang, S.-H. Lin, C.-W. Chen, Z.-H. Chang and Y.-W. Chu, *N*-GlycoGo: predicting protein *N*-glycosylation sites on imbalanced data sets by using heterogeneous and comprehensive strategy, *IEEE Access*, 2020, **8**, 165944–165950.
96 T. Pitti, C.-T. Chen, H.-N. Lin, W.-K. Choong, W.-L. Hsu and T.-Y. Sung, N-GlyDE: a two-stage N-linked glycosylation site prediction incorporating gapped dipeptides and pattern-based encoding, *Sci. Rep.*, 2019, **9**, 15975.
97 G. Taherzadeh, A. Dehzangi, M. Golchin, Y. Zhou and M. P. Campbell, SPRINT-Gly: predicting N- and O-linked glycosylation sites of human and mouse proteins by using sequence and predicted structural properties, *Bioinformatics*, 2019, **35**, 4140–4146.
98 A. H. Bhat, H. Mondal, J. S. Chauhan, G. P. S. Raghava, A. Methi and A. Rao, ProGlycProt: a repository of experimentally characterized prokaryotic glycoproteins, *Nucleic Acids Res.*, 2012, **40**, D388–D393.

99 H. Fu, Y. Yang, X. Wang, H. Wang and Y. Xu, DeepUbi: a deep learning framework for prediction of ubiquitination sites in proteins, *BMC Bioinform.*, 2019, **20**, 86.

100 Z. Chen, Y.-Z. Chen, X.-F. Wang, C. Wang, R.-X. Yan and Z. Zhang, Prediction of ubiquitination sites by using the composition of k-spaced amino acid pairs, *PLoS One*, 2011, **6**, e22930.

101 W. Qiu, C. Xu, X. Xiao and D. Xu, Computational prediction of ubiquitination proteins using evolutionary profiles and functional domain annotation, *Curr. Genomics*, 2019, **20**, 389–399.

102 T. Chen, T. Zhou, B. He, H. Yu, X. Guo, X. Song and J. Sha, mUbiSiDa: a comprehensive database for protein ubiquitination sites in mammals, *PLoS One*, 2014, **9**, e85744.

103 P. Radivojac, V. Vacic, C. Haynes, R. R. Cocklin, A. Mohan, J. W. Heyen, M. G. Goebl and L. M. Iakoucheva, Identification, analysis, and prediction of protein ubiquitination sites, *Proteins: Struct. Funct. Bioinform.*, 2010, **78**, 365–380.

104 Z. Li, S. Chen, J.-H. Jhong, Y. Pang, K.-Y. Huang, S. Li and T.-Y. Lee, UbiNet 2.0: a verified, classified, annotated and updated database of E3 ubiquitin ligase–substrate interactions, *Database*, 2021, **2021**, baab010.

105 H. Wang, Z. Wang, Z. Li and T.-Y. Lee, Incorporating deep learning with word embedding to identify plant ubiquitylation sites, *Front. Cell Dev. Biol.*, 2020, **8**, 572195.

106 A. Siraj, D. Y. Lim, H. Tayara and K. T. Chong, UbiComb: a hybrid deep learning model for predicting plant-specific protein ubiquitylation sites, *Genes*, 2021, **12**, 717.

107 W.-R. Qiu, X. Xiao, W.-Z. Lin and K.-C. Chou, iUbiq-Lys: prediction of lysine ubiquitination sites in proteins by extracting sequence evolution information via a gray system model, *J. Biomol. Struct. Dyn.*, 2015, **33**, 1731–1742.

108 W. Ning, H. Xu, P. Jiang, H. Cheng, W. Deng, Y. Guo and Y. Xue, HybridSucc: a hybrid-learning architecture for general and species-specific succinylation site prediction, *Genom. Proteom. Bioinform.*, 2020, **18**, 194–207.

109 M. Blanc, F. P. A. David and F. G. van der Goot, in *Protein Lipidation: Methods and Protocols*, ed. M. E. Linder, Springer, New York, NY, 2019, pp. 203–214.

110 J. Ren, L. Wen, X. Gao, C. Jin, Y. Xue and X. Yao, CSS-Palm 2.0: an updated software for palmitoylation sites prediction, *Protein Eng., Des. Sel.*, 2008, **21**, 639–644.

111 W. Ning, P. Jiang, Y. Guo, C. Wang, X. Tan, W. Zhang, D. Peng and Y. Xue, GPS-Palm: a deep learning-based graphic presentation system for the prediction of *S*-palmitoylation sites in proteins, *Brief. Bioinform*, 2021, **22**, 1836–1847.

112 Q. Zhao, Y. Xie, Y. Zheng, S. Jiang, W. Liu, W. Mu, Z. Liu, Y. Zhao, Y. Xue and J. Ren, GPS-SUMO: a tool for the prediction of sumoylation sites and SUMO-interaction motifs, *Nucleic Acids Res.*, 2014, **42**, W325–W330.

113 Z. Liu, Q. Ma, J. Cao, X. Gao, J. Ren and Y. Xue, GPS-PUP: computational prediction of pupylation sites in prokaryotic proteins, *Mol. BioSyst.*, 2011, **7**, 2737–2740.

114 Y. Yang, H. Wang, J. Ding and Y. Xu, iAcet-Sumo: identification of lysine acetylation and sumoylation sites in proteins by multi-class transformation methods, *Comput. Biol. Med.*, 2018, **100**, 144–151.

115 C.-C. Chang, C.-H. Tung, C.-W. Chen, C.-H. Tu and Y.-W. Chu, SUMOgo: prediction of sumoylation sites on lysines by motif screening models and the effects of various post-translational modifications, *Sci. Rep.*, 2018, **8**, 15512.

116 G. Chen, M. Cao, K. Luo, L. Wang, P. Wen and S. Shi, ProAcePred: prokaryote lysine acetylation sites prediction based on elastic net feature optimization, *Bioinformatics*, 2018, **34**, 3999–4006.

117 L. Kiemer, J. D. Bendtsen and N. Blom, NetAcet: prediction of N-terminal acetylation sites, *Bioinformatics*, 2005, **21**, 1269–1270.

118 G. Beauclair, A. Bridier-Nahmias, J.-F. Zagury, A. Saïb and A. Zamborlini, JASSA: a comprehensive tool for prediction of SUMOylation sites and SIMs, *Bioinformatics*, 2015, **31**, 3483–3491.

119 S. Maurer-Stroh, B. Eisenhaber and F. Eisenhaber, N-terminal *N*-myristoylation of proteins: prediction of substrate proteins from amino acid sequence, *J. Mol. Biol.*, 2002, **317**, 541–557.

120 W. Deng, Y. Wang, L. Ma, Y. Zhang, S. Ullah and Y. Xue, Computational prediction of methylation types of covalently modified lysine and arginine residues in proteins, *Brief. Bioinform.*, 2017, **18**, 647–658.

121 H. Chen, Y. Xue, N. Huang, X. Yao and Z. Sun, MeMo: a web tool for prediction of protein methylation modifications, *Nucleic Acids Res.*, 2006, **34**, W249–W253.

122 Y. Xue, Z. Liu, X. Gao, C. Jin, L. Wen, X. Yao and J. Ren, GPS-SNO: computational prediction of protein *S*-nitrosylation sites with a modified GPS algorithm, *PLoS One*, 2010, **5**, e11290.

123 Z. Liu, J. Cao, Q. Ma, X. Gao, J. Ren and Y. Xue, GPS-YNO2: computational prediction of tyrosine nitration sites in proteins, *Mol. BioSyst.*, 2011, **7**, 1197–1204.

124 Y.-J. Chen, C.-T. Lu, M.-G. Su, K.-Y. Huang, W.-C. Ching, H.-H. Yang, Y.-C. Liao, Y.-J. Chen and T.-Y. Lee, dbSNO 2.0: a resource for exploring structural environment, functional and disease association and regulatory network of protein *S*-nitrosylation, *Nucleic Acids Res.*, 2015, **43**, D503–D511.

125 Z. Liu, J. Cao, X. Gao, Q. Ma, J. Ren and Y. Xue, GPS-CCD: a novel computational program for the prediction of calpain cleavage sites, *PLoS One*, 2011, **6**, e19001.

126 Z. Pan, Z. Liu, H. Cheng, Y. Wang, T. Gao, S. Ullah, J. Ren and Y. Xue, Systematic analysis of the *in situ* crosstalk of tyrosine modifications reveals no additional natural selection on multiply modified residues, *Sci. Rep.*, 2014, **4**, 7331.

127 W. Deng, C. Wang, Y. Zhang, Y. Xu, S. Zhang, Z. Liu and Y. Xue, GPS-PAIL: prediction of lysine acetyltransferase-specific modification sites from protein sequences, *Sci. Rep.*, 2016, **6**, 39787.

128 N. Blom, J. Hansen, S. Brunak and D. Blaas, Cleavage site analysis in picornaviral polyproteins: discovering cellular targets by neural networks, *Protein Sci.*, 1996, **5**, 2203–2216.

129 P. Duckert, S. Brunak and N. Blom, Prediction of proprotein convertase cleavage sites, *Protein Eng., Des. Sel.*, 2004, **17**, 107–112.

130 Y. Xie, X. Luo, Y. Li, L. Chen, W. Ma, J. Huang, J. Cui, Y. Zhao, Y. Xue, Z. Zuo and J. Ren, DeepNitro: prediction of protein nitration and nitrosylation sites by deep learning, *Genom. Proteom. Bioinform.*, 2018, **16**, 294–306.

131 L. Wei, P. Xing, G. Shi, Z. Ji and Q. Zou, Fast prediction of protein methylation sites using a sequence-based feature selection technique, *IEEE/ACM Trans. Comput. Biol. Bioinform.*, 2019, **16**, 1264–1273.

132 Y. Xu, J. Ding, L.-Y. Wu and K.-C. Chou, iSNO-PseAAC: predict cysteine *S*-nitrosylation sites in proteins by incorporating position specific amino acid propensity into pseudo amino acid composition, *PLoS One*, 2013, **8**, e55844.

133 M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, Avogadro: an advanced semantic chemical editor, visualization, and analysis platform, *J. Cheminf.*, 2012, **4**, 1–17.

134 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, UCSF Chimera–a visualization system for exploratory research and analysis, *J. Comput. Chem.*, 2004, **25**, 1605–1612.

135 *Maestro*, Schrödinger, LLC, New York, NY, 2022.

136 *Molecular Operating Environment (MOE)*, Chemical Computing Group ULC, 1010 Sherbrooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2022.

137 W. DeLano, PYMOL, Schrödinger, LLC, New York, NY, 2020.

138 W. Humphrey, A. Dalke and K. Schulten, VMD: visual molecular dynamics, *J. Mol. Graphics*, 1996, **14**, 33–38.

139 D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. I. Cheatham, V. W. D. Cruziero, T. A. Darden, R. E. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A. W. Goetz, D. Greene, R. Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T.-S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. J. Mermelstein, K. M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, L. Xiao, D. M. York and P. A. Kollman, *AMBER*, 2022.

140 B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, CHARMM: the biomolecular simulation program, *J. Comput. Chem.*, 2009, **30**, 1545–1614.

141 S. B. Engelsen, S. Cros, W. Mackie and S. Pérez, A molecular builder for carbohydrates: application to polysaccharides and complex carbohydrates, *Biopolymers*, 1996, **39**, 417–433.

142 A. Bohne, E. Lang and C.-W. von der Lieth, W3-SWEET: carbohydrate modeling by internet, *J. Mol. Med.*, 1998, **4**, 33–43.

143 M. Frank, A. Bohne-Lang, T. Wetter and C.-W. V. D. Lieth, Rapid generation of a representative ensemble of N-glycan conformations, *In Silico Biol.*, 2002, **2**, 427–439.

144 J. Rosen, L. Miguet and S. Pérez, Shape: automatic conformation prediction of carbohydrates using a genetic algorithm, *J. Cheminform.*, 2009, **1**, 16.

145 S. B. Engelsen, P. I. Hansen and S. Pérez, POLYS 2.0: an open source software package for building three-dimensional structures of polysaccharides, *Biopolymers*, 2014, **101**, 733–743.

146 R. Woods, L. Foley, G. Chalmers, A. Khatamian, S. Makeneni, D. Rahbarinia, A. Singh, X. Xue, D. Thieker, J. Bellamy, R. Davis, L. Ding, H. Kim, Q. Li, M. Nagarajan, S. Voghooei, Y. Wang, A. Jackson, S. Lee, B. Miller, A. Santos, J. Tatem and D. Templeton, *GLYCAM web: Glycoprotein builder*, Woods Group, Complex Carbohydrate Research Center, University of Georgia, Athens, GA, 2005.

147 S. Jo, K. C. Song, H. Desaire, A. D. MacKerell Jr. and W. Im, Glycan reader: automated sugar identification and simulation preparation for carbohydrates and glycoproteins, *J. Comput. Chem.*, 2011, **32**, 3135–3141.

148 S.-J. Park, J. Lee, D. S. Patel, H. Ma, H. S. Lee, S. Jo and W. Im, Glycan Reader is improved to recognize most sugar types and chemical modifications in the Protein Data Bank, *Bioinformatics*, 2017, **33**, 3051–3057.

149 S.-J. Park, J. Lee, Y. Qi, N. R. Kern, H. S. Lee, S. Jo, I. Joung, K. Joo, J. Lee and W. Im, CHARMM-GUI glycan modeler for modeling and simulation of carbohydrates and glycoconjugates, *Glycobiology*, 2019, **29**, 320–331.

150 R. Danne, C. Poojari, H. Martinez-Seara, S. Rissanen, F. Lolicato, T. Róg and I. Vattulainen, doGlycans–Tools for preparing carbohydrate structures for atomistic simulations of glycoproteins, glycolipids, and carbohydrate polymers for GROMACS, *J. Chem. Inf. Model.*, 2017, **57**, 2401–2406.

151 S. Tsuchiya, N. P. Aoki, D. Shinmachi, M. Matsubara, I. Yamada, K. F. Aoki-Kinoshita and H. Narimatsu, Implementation of GlycanBuilder to draw a wide variety of ambiguous glycans, *Carbohydr. Res.*, 2017, **445**, 104–116.

152 M. Böhm, A. Bohne-Lang, M. Frank, A. Loss, M. A. Rojas-Macias and T. Lütteke, Glycosciences.DB: an annotated data collection linking glycomics and proteomics data (2018 update), *Nucleic Acids Res.*, 2019, **47**, D1195–D1201.

153 T. Lemmin and C. Soto, Glycosylator: a python framework for the rapid modeling of glycans, *BMC Bioinform.*, 2019, **20**, 513.

154 A. Warnecke, T. Sandalova, A. Achour and R. A. Harris, PyTMs: a useful PyMOL plugin for modeling common post-translational modifications, *BMC Bioinform.*, 2014, **15**, 370.

155 S. Jo, T. Kim and W. Im, Automated builder and database of protein/membrane complexes for molecular dynamics simulations, *PLoS One*, 2007, **2**, e880.

156 S. Jo, X. Cheng, S. M. Islam, L. Huang, H. Rui, A. Zhu, H. S. Lee, Y. Qi, W. Han, K. Vanommeslaeghe, A. D. MacKerell, B. Roux and W. Im, in *Advances in Protein*

Chemistry and Structural Biology, ed. T. Karabencheva-Christova, Academic Press, 2014, vol. 96, pp. 235–265.
157 J. Lee, X. Cheng, J. M. Swails, M. S. Yeom, P. K. Eastman, J. A. Lemkul, S. Wei, J. Buckner, J. C. Jeong, Y. Qi, S. Jo, V. S. Pande, D. A. Case, C. L. Brooks, A. D. MacKerell, J. B. Klauda and W. Im, CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field, *J. Chem. Theory Comput.*, 2016, **12**, 405–413.
158 H. Woo, S.-J. Park, Y. K. Choi, T. Park, M. Tanveer, Y. Cao, N. R. Kern, J. Lee, M. S. Yeom, T. I. Croll, C. Seok and W. Im, Developing a fully glycosylated full-length SARS-CoV-2 spike protein model in a viral membrane, *J. Phys. Chem. B*, 2020, **124**, 7128–7137.
159 O. C. Grant, D. Montgomery, K. Ito and R. J. Woods, Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition, *Sci. Rep.*, 2020, **10**, 14991.
160 L. Casalino, Z. Gaieb, J. A. Goldsmith, C. K. Hjorth, A. C. Dommer, A. M. Harbison, C. A. Fogarty, E. P. Barros, B. C. Taylor, J. S. McLellan, E. Fadda and R. E. Amaro, Beyond shielding: the roles of glycans in the SARS-CoV-2 spike protein, *ACS Cent. Sci.*, 2020, **6**, 1722–1734.
161 Y. K. Choi, Y. Cao, M. Frank, H. Woo, S.-J. Park, M. S. Yeom, T. I. Croll, C. Seok and W. Im, Structure, dynamics, receptor binding, and antibody binding of the fully glycosylated full-length SARS-CoV-2 spike protein in a viral membrane, *J. Chem. Theory Comput.*, 2021, **17**, 2479–2487.
162 Y. Cao, Y. K. Choi, M. Frank, H. Woo, S.-J. Park, M. S. Yeom, C. Seok and W. Im, Dynamic interactions of fully glycosylated SARS-CoV-2 spike protein with various antibodies, *J. Chem. Theory Comput.*, 2021, **17**, 6559–6569.
163 L. Casalino, A. C. Dommer, Z. Gaieb, E. P. Barros, T. Sztain, S.-H. Ahn, A. Trifan, A. Brace, A. T. Bogetti, A. Clyde, H. Ma, H. Lee, M. Turilli, S. Khalid, L. T. Chong, C. Simmerling, D. J. Hardy, J. D. Maia, J. C. Phillips, T. Kurth, A. C. Stern, L. Huang, J. D. McCalpin, M. Tatineni, T. Gibbs, J. E. Stone, S. Jha, A. Ramanathan and R. E. Amaro, AI-driven multi-scale simulations illuminate mechanisms of SARS-CoV-2 spike dynamics, *Int. J. High Perf. Comput. Appl.*, 2021, **35**, 432–451.
164 M. M. Lamers and B. L. Haagmans, SARS-CoV-2 pathogenesis, *Nat. Rev. Microbiol.*, 2022, **20**, 270–284.
165 K. Nagata, A. Randall and P. Baldi, Incorporating post-translational modifications and unnatural amino acids into high-throughput modeling of protein structures, *Bioinformatics*, 2014, **30**, 1681–1689.
166 D. Kony, W. Damm, S. Stoll and W. F. Van Gunsteren, An improved OPLS–AA force field for carbohydrates, *J. Comput. Chem.*, 2002, **23**, 1416–1429.
167 K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley and R. J. Woods, GLYCAM06: a generalizable biomolecular force field. Carbohydrates, *J. Comput. Chem.*, 2008, **29**, 622–655.
168 O. Guvench, S. S. Mallajosyula, E. P. Raman, E. Hatcher, K. Vanommeslaeghe, T. J. Foster, F. W. Jamison and A. D. MacKerell, CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate–protein modeling, *J. Chem. Theory Comput.*, 2011, **7**, 3162–3180.
169 L. Pol-Fachin, V. H. Rusu, H. Verli and R. D. Lins, GROMOS 53A6GLYC, an improved GROMOS force field for hexopyranose-based carbohydrates, *J. Chem. Theory Comput.*, 2012, **8**, 4681–4690.
170 J. Huang and A. D. MacKerell Jr, CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data, *J. Comput. Chem.*, 2013, **34**, 2135–2145.
171 G. A. Khoury, J. P. Thompson, J. Smadbeck, C. A. Kieslich and C. A. Floudas, Forcefield_PTM: ab initio charge and AMBER forcefield parameters for frequently occurring post-translational modifications, *J. Chem. Theory Comput.*, 2013, **9**, 5653–5674.
172 T. Yamashita, R. Okajima, K. Miyanabe and K. Tsumoto, Modified AMBER force-field (FUJI) parameters for sulfated and phosphorylated tyrosine residues: development and application to CCR5-derived peptide systems, *AIP Conf. Proc.*, 2019, **2186**, 030013.
173 P. G. Garay, E. E. Barrera and S. Pantano, Post-translational modifications at the coarse-grained level with the SIRAH force field, *J. Chem. Inf. Model.*, 2020, **60**, 964–973.
174 A. T. Shivgan, J. K. Marzinek, R. G. Huber, A. Krah, R. H. Henchman, P. Matsudaira, C. S. Verma and P. J. Bond, Extending the Martini coarse-grained force field to *N*-glycans, *J. Chem. Inf. Model.*, 2020, **60**, 3864–3883.
175 J. P. Stoppelman, T. T. Ng, P. S. Nerenberg and L.-P. Wang, Development and validation of AMBER-FB15-compatible force field parameters for phosphorylated amino acids, *J. Phys. Chem. B*, 2021, **125**, 11927–11942.
176 M. J. Robertson and G. Skiniotis, Development of OPLS-AA/M parameters for simulations of G protein-coupled receptors and other membrane proteins, *J. Chem. Theory Comput.*, 2022, **18**(7), 4482–4489.
177 M. Goldstein, Viscous liquids and the glass transition: a potential energy barrier picture, *J. Chem. Phys.*, 1969, **51**, 3728–3739.
178 H. Frauenfelder, S. G. Sligar and P. G. Wolynes, The energy landscapes and motions of proteins, *Science*, 1991, **254**, 1598–1603.
179 F. H. Stillinger, A topographic view of supercooled liquids and glass formation, *Science*, 1995, **267**, 1935–1939.
180 R. Du, V. S. Pande, A. Y. Grosberg, T. Tanaka and E. S. Shakhnovich, On the transition coordinate for protein folding, *J. Chem. Phys.*, 1998, **108**, 334–350.
181 J. Schlitter, M. Engels, P. Krüger, E. Jacoby and A. Wollmer, Targeted molecular dynamics simulation of conformational change—Application to the T ↔ R transition in insulin, *Mol. Simul.*, 1993, **10**, 291–308.
182 H. Grubmüller, B. Heymann and P. Tavan, Ligand Binding: Molecular mechanics calculation of the streptavidin–biotin rupture force, *Science*, 1996, **271**, 997–999.
183 D. Hamelberg, J. Mongan and J. A. McCammon, Accelerated molecular dynamics: a promising and efficient

simulation method for biomolecules, *J. Chem. Phys.*, 2004, **120**, 11919–11929.
184 G. Bussi and A. Laio, Using metadynamics to explore complex free-energy landscapes, *Nat. Rev. Phys.*, 2020, **2**, 200–212.
185 W. Wang, S. Cao, L. Zhu and X. Huang, Constructing Markov state models to elucidate the functional conformational changes of complex biomolecules, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2018, **8**, e1343.
186 B. E. Husic and V. S. Pande, Markov state models: from an art to a science, *J. Am. Chem. Soc.*, 2018, **140**, 2386–2396.
187 F. Noé and E. Rosta, Markov models of molecular kinetics, *J. Chem. Phys.*, 2019, **151**, 190401.
188 R. Elber, Milestoning: an efficient approach for atomically detailed simulations of kinetics in biophysics, *Annu. Rev. Biophys.*, 2020, **49**, 69–85.
189 J. Kästner, Umbrella sampling, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2011, **1**, 932–942.
190 W. E. W. Ren and E. Vanden-Eijnden, String method for the study of rare events, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2002, **66**, 052301.
191 Y. Sugita and Y. Okamoto, Replica-exchange molecular dynamics method for protein folding, *Chem. Phys. Lett.*, 1999, **314**, 141–151.
192 M. L. Schneider and C. B. Post, Solution structure of a band 3 peptide inhibitor bound to aldolase: a proposed mechanism for regulating binding by tyrosine phosphorylation, *Biochemistry*, 1995, **34**, 16574–16584.
193 M.-H. Feng, M. Philippopoulos, A. D. MacKerell and C. Lim, Structural characterization of the phosphotyrosine binding region of a high-affinity SH2 domain−phosphopeptide complex by molecular dynamics simulation and chemical shift calculations, *J. Am. Chem. Soc.*, 1996, **118**, 11265–11277.
194 J. Mendieta and F. Gago, In silico activation of Src tyrosine kinase reveals the molecular basis for intramolecular autophosphorylation, *J. Mol. Graph. Model.*, 2004, **23**, 189–198.
195 I. Bártová, J. Koča and M. Otyepka, Regulatory phosphorylation of cyclin-dependent kinase 2: insights from molecular dynamics simulations, *J. Mol. Model.*, 2008, **14**, 761.
196 I. Bártová, M. Otyepka, Z. Kříž and J. Koča, Activation and inhibition of cyclin-dependent kinase-2 by phosphorylation; a molecular dynamics study reveals the functional importance of the glycine-rich loop, *Protein Sci.*, 2004, **13**, 1449–1457.
197 I. Bártová, M. Otyepka, Z. KřÍž and J. Koča, The mechanism of inhibition of the cyclin-dependent kinase-2 as revealed by the molecular dynamics study on the complex CDK2 with the peptide substrate HHASPRK, *Protein Sci.*, 2005, **14**, 445–451.
198 P. Roche, L. Mouawad, D. Perahia, J.-P. Samama and D. Kahn, Molecular dynamics of the FixJ receiver domain: movement of the β4–α4 loop correlates with the in and out flip of Phe101, *Protein Sci.*, 2002, **11**, 2622–2630.
199 K. Tomoo, X. Shen, K. Okabe, Y. Nozoe, S. Fukuhara, S. Morino, M. Sasaki, T. Taniguchi, H. Miyagawa, K. Kitamura, K. Miura and T. Ishida, Structural features of human initiation factor 4E, studied by X-ray crystal analyses and molecular dynamics simulations, *J. Mol. Biol.*, 2003, **328**, 365–383.
200 G. H. Peters, T. M. Frimurer, J. N. Andersen and O. H. Olsen, Molecular dynamics simulations of Protein–Tyrosine Phosphatase 1B. II. Substrate-enzyme interactions and dynamics, *Biophys. J.*, 2000, **78**, 2191–2200.
201 C. M. Stultz, A. D. Levin and E. R. Edelman, Phosphorylation-induced conformational changes in a mitogen-activated protein kinase substrate: implications for tyrosine hydroxylase activation, *J. Biol. Chem.*, 2002, **277**, 47653–47661.
202 A. Phan-Chan-Du, C. Hemmerlin, D. Krikorian, M. Sakarellos-Daitsiotis, V. Tsikaris, C. Sakarellos, M. Marinou, A. Thureau, M. T. Cung and S. J. Tzartos, Solution conformation of the antibody-bound tyrosine phosphorylation site of the nicotinic acetylcholine receptor β-subunit in its phosphorylated and nonphosphorylated states, *Biochemistry*, 2003, **42**, 7371–7380.
203 P. Kumar, M. S. Chimenti, H. Pemble, A. Schönichen, O. Thompson, M. P. Jacobson and T. Wittmann, Multisite phosphorylation disrupts arginine-glutamate salt bridge networks required for binding of Cytoplasmic Linker-associated Protein 2 (CLASP2) to End-binding Protein 1 (EB1), *J. Biol. Chem.*, 2012, **287**, 17050–17064.
204 J. L. Smart and J. A. McCammon, Phosphorylation stabilizes the N-termini of α-helices, *Biopolymers*, 1999, **49**, 225–233.
205 T. Shen, C. F. Wong and J. A. McCammon, Atomistic brownian dynamics simulation of peptide phosphorylation, *J. Am. Chem. Soc.*, 2001, **123**, 9107–9111.
206 J. N. Mavinahalli, A. Madhumalar, R. W. Beuerman, D. P. Lane and C. Verma, Differences in the transactivation domains of p53 family members: a computational study, *BMC Genom.*, 2010, **11**, S5.
207 M. Germana Paterlini and D. D. Thomas, The α-helical propensity of the cytoplasmic domain of phospholamban: a molecular dynamics simulation of the effect of phosphorylation and mutation, *Biophys. J.*, 2005, **88**, 3243–3251.
208 H. Yalinca, C. J. C. Gehin, V. Oleinikovas, H. A. Lashuel, F. L. Gervasio and A. Pastore, The role of post-translational modifications on the energy landscape of Huntingtin N-terminus, *Front. Mol. Biosci.*, 2019, **6**, 95.
209 D. J. Mandell, I. Chorny, E. S. Groban, S. E. Wong, E. Levine, C. S. Rapp and M. P. Jacobson, Strengths of hydrogen bonds involving phosphorylated amino acid side chains, *J. Am. Chem. Soc.*, 2007, **129**, 820–827.
210 M. M. Chen, A. I. Bartlett, P. S. Nerenberg, C. T. Friel, C. P. R. Hackenberger, C. M. Stultz, S. E. Radford and B. Imperiali, Perturbing the folding energy landscape of the bacterial immunity protein Im7 by site-specific N-linked glycosylation, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 22528–22533.
211 T. Hagai, A. Azia, Á. Tóth-Petróczy and Y. Levy, Intrinsic disorder in ubiquitination substrates, *J. Mol. Biol.*, 2011, **412**, 319–324.

212 D. Shental-Bechor and Y. Levy, Effect of glycosylation on protein folding: a close look at thermodynamic stabilization, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 8256–8261.

213 D. Shental-Bechor and Y. Levy, Folding of glycoproteins: toward understanding the biophysics of the glycosylation code, *Curr. Opin. Struct. Biol.*, 2009, **19**, 524–533.

214 H. S. Lee, Y. Qi and W. Im, Effects of N-glycosylation on protein conformation and dynamics: protein Data Bank analysis and molecular dynamics simulation study, *Sci. Rep.*, 2015, **5**, 8926.

215 J. L. Price, D. Shental-Bechor, A. Dhar, M. J. Turner, E. T. Powers, M. Gruebele, Y. Levy and J. W. Kelly, Context-dependent effects of asparagine glycosylation on Pin WW folding kinetics and thermodynamics, *J. Am. Chem. Soc.*, 2010, **132**, 15359–15367.

216 P. B. Lawrence, Y. Gavrilov, S. S. Matthews, M. I. Langlois, D. Shental-Bechor, H. M. Greenblatt, B. K. Pandey, M. S. Smith, R. Paxman, C. D. Torgerson, J. P. Merrell, C. C. Ritz, M. B. Prigozhin, Y. Levy and J. L. Price, Criteria for selecting PEGylation sites on proteins for higher thermodynamic and proteolytic stability, *J. Am. Chem. Soc.*, 2014, **136**, 17547–17560.

217 K. A. Ball, J. R. Johnson, M. K. Lewinski, J. Guatelli, E. Verschueren, N. J. Krogan and M. P. Jacobson, Non-degradative ubiquitination of protein kinases, *PLoS Comput. Biol.*, 2016, **12**, e1004898.

218 Ž. Sovová, J. Štikarová, J. Kaufmanová, P. Májek, J. Suttnar, P. Šácha, M. Malý and J. E. Dyr, Impact of posttranslational modifications on atomistic structure of fibrinogen, *PLoS One*, 2020, **15**, e0227543.

219 Y. Gavrilov, D. Shental-Bechor, H. M. Greenblatt and Y. Levy, Glycosylation may reduce protein thermodynamic stability by inducing a conformational distortion, *J. Phys. Chem. Lett.*, 2015, **6**, 3572–3577.

220 T. Shimazu, M. D. Hirschey, L. Hua, K. E. Dittenhafer-Reed, B. Schwer, D. B. Lombard, Y. Li, J. Bunkenborg, F. W. Alt, J. M. Denu, M. P. Jacobson and E. Verdin, SIRT3 deacetylates mitochondrial 3-hydroxy-3-methylglutaryl CoA synthase 2 and regulates ketone body production, *Cell Metab.*, 2010, **12**, 654–661.

221 M. A. Young, S. Gonfloni, G. Superti-Furga, B. Roux and J. Kuriyan, Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation, *Cell*, 2001, **105**, 115–126.

222 D. Winogradoff, I. Echeverria, D. A. Potoyan and G. A. Papoian, The acetylation landscape of the H4 histone tail: disentangling the interplay between the specific and cumulative effects, *J. Am. Chem. Soc.*, 2015, **137**, 6245–6253.

223 W. Lyu (Lv), F. Arnesano, P. Carloni, G. Natile and G. Rossetti, Effect of in vivo post-translational modifications of the HMGB1 protein upon binding to platinated DNA: a molecular simulation study, *Nucleic Acids Res.*, 2018, **46**, 11687–11697.

224 N. Šoštarić and V. van Noort, Molecular dynamics shows complex interplay and long-range effects of post-translational modifications in yeast protein interactions, *PLoS Comput. Biol.*, 2021, **17**, e1008988.

225 N. R. Latorraca, M. Masureel, S. A. Hollingsworth, F. M. Heydenreich, C.-M. Suomivuori, C. Brinton, R. J. L. Townshend, M. Bouvier, B. K. Kobilka and R. O. Dror, How GPCR phosphorylation patterns orchestrate arrestin-mediated signaling, *Cell*, 2020, **183**, 1813–1825.e18.

226 M. C. Chan and D. Shukla, The effects of N-linked glycosylation on SLC6 transporters, *bioRxiv*, 2022, prepint, DOI: 10.1101/2022.07.12.499387.

227 D. Hamelberg, T. Shen and J. A. McCammon, Phosphorylation effects on cis/trans isomerization and the backbone conformation of serine–proline motifs: accelerated molecular dynamics analysis, *J. Am. Chem. Soc.*, 2005, **127**, 1969–1974.

228 H. A. Velazquez and D. Hamelberg, Conformational selection in the recognition of phosphorylated substrates by the catalytic domain of human Pin1, *Biochemistry*, 2011, **50**, 9605–9615.

229 E. Papaleo, V. Ranzani, F. Tripodi, A. Vitriolo, C. Cirulli, P. Fantucci, L. Alberghina, M. Vanoni, L. D. Gioia and P. Coccetti, An acidic loop and cognate phosphorylation sites define a molecular switch that modulates ubiquitin charging activity in Cdc34-like enzymes, *PLoS Comput. Biol.*, 2011, **7**, e1002056.

230 N. A. Jonniya, M. F. Sk and P. Kar, Investigating phosphorylation-induced conformational changes in WNK1 kinase by molecular dynamics simulations, *ACS Omega*, 2019, **4**, 17404–17416.

231 H.-F. Chen, Molecular dynamics simulation of phosphorylated KID post-translational modification, *PLoS One*, 2009, **4**, e6516.

232 A. Kuzmanic, L. Sutto, G. Saladino, A. R. Nebreda, F. L. Gervasio and M. Orozco, Changes in the free-energy landscape of p38α MAP kinase through its canonical activation and binding events as studied by enhanced molecular dynamics simulations, *eLife*, 2017, **6**, e22175.

233 A. S. Moffett and D. Shukla, Structural consequences of multisite phosphorylation in the BAK1 kinase domain, *Biophys. J.*, 2020, **118**, 698–707.

234 J. Lätzer, T. Shen and P. G. Wolynes, Conformational switching upon phosphorylation: a predictive framework based on energy landscape principles, *Biochemistry*, 2008, **47**, 2110–2122.

235 D. Petrov and B. Zagrovic, Microscopic analysis of protein oxidative damage: effect of carbonylation on structure, dynamics, and aggregability of villin headpiece, *J. Am. Chem. Soc.*, 2011, **133**, 7016–7024.

236 G. R. Bowman, D. L. Ensign and V. S. Pande, Enhanced modeling via network theory: adaptive sampling of Markov state models, *J. Chem. Theory Comput.*, 2010, **6**, 787–794.

237 S. Doerr and G. De, Fabritiis, On-the-fly learning and sampling of ligand binding by high-throughput molecular simulations, *J. Chem. Theory Comput.*, 2014, **10**, 2064–2069.

238 E. Hruska, J. R. Abella, F. Nüske, L. E. Kavraki and C. Clementi, Quantitative comparison of adaptive

sampling methods for protein dynamics, *J. Chem. Phys.*, 2018, **149**, 244119.

239 R. M. Betz and R. O. Dror, How effectively can adaptive sampling methods capture spontaneous ligand binding?, *J. Chem. Theory Comput.*, 2019, **15**, 2053–2063.

240 Z. Shamsi, A. S. Moffett and D. Shukla, Enhanced unbiased sampling of protein dynamics using evolutionary coupling information, *Sci. Rep.*, 2017, **7**, 12700.

241 Z. Shamsi, K. J. Cheng and D. Shukla, Reinforcement learning based adaptive sampling: REAPing rewards by exploring protein conformational landscapes, *J. Phys. Chem. B*, 2018, **122**, 8386–8395.

242 K. Shin, D. P. Tran, K. Takemura, A. Kitao, K. Terayama and K. Tsuda, Enhancing biomolecular sampling with reinforcement learning: a tree search molecular dynamics simulation method, *ACS Omega*, 2019, **4**, 13853–13862.

243 A. Pérez, P. Herrera-Nieto, S. Doerr and G. De, Fabritiis, AdaptiveBandit: a multi-armed bandit framework for adaptive sampling in molecular simulations, *J. Chem. Theory Comput.*, 2020, **16**, 4685–4693.

244 D. E. Kleiman and D. Shukla, Multiagent reinforcement learning-based adaptive sampling for conformational dynamics of proteins, *J. Chem. Theory Comput.*, 2022, **18**, 5422–5434.

245 V. S. Pande, K. Beauchamp and G. R. Bowman, Everything you wanted to know about Markov State Models but were afraid to ask, *Methods*, 2010, **52**, 99–105.

246 J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte and F. Noé, Markov models of molecular kinetics: generation and validation, *J. Chem. Phys.*, 2011, **134**, 174105.

247 J. D. Chodera and F. Noé, Markov state models of biomolecular conformational dynamics, *Curr. Opin. Struct. Biol.*, 2014, **25**, 135–144.

248 G. R. Bowman, V. S. Pande and F. Noé, *An introduction to Markov state models and their application to long timescale molecular simulation*, Springer, Heidelberg, Germany, 2014.

249 D. Shukla, C. X. Hernández, J. K. Weber and V. S. Pande, Markov state models provide insights into dynamic modulation of protein function, *Acc. Chem. Res.*, 2015, **48**, 414–422.

250 X. Wang, I. C. Unarta, P. P.-H. Cheung and X. Huang, Elucidating molecular mechanisms of functional conformational changes of proteins via Markov state models, *Curr. Opin. Struct. Biol.*, 2021, **67**, 69–77.

251 Y. Ge and V. A. Voelz, in *Protein-Ligand Interactions and Drug Design*, ed. F. Ballante, Springer US, New York, NY, 2021, pp. 239–259.

252 A. S. Moffett, K. W. Bender, S. C. Huber and D. Shukla, Allosteric control of a plant receptor kinase through S-glutathionylation, *Biophys. J.*, 2017, **113**, 2354–2363.

253 C. L. McClendon, L. Hua, G. Barreiro and M. P. Jacobson, Comparing conformational ensembles using the Kullback–Leibler divergence expansion, *J. Chem. Theory Comput.*, 2012, **8**, 2115–2126.

254 S. Shukla, C. Zhao and D. Shukla, Dewetting controls plant hormone perception and initiation of drought resistance signaling, *Structure*, 2019, **27**, 692–702.e3.

255 B. Selvam, J. Feng and D. Shukla, Atomistic insights into the mechanism of dual affinity switching in plant nitrate transporter NRT1.1, *bioRxiv*, 2022, preprint, DOI: 10.1101/2022.10.17.512638.

256 E. Vanden-Eijnden, in *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology*, Volume 1, ed. M. Ferrario, G. Ciccotti and K. Binder, Springer, Berlin, Heidelberg, 2006, pp. 453–493.

257 Y.-Y. Wang, Y.-H. Cheng, K.-E. Chen and Y.-F. Tsay, Nitrate transport, signaling, and use efficiency, *Annu. Rev. Plant Biol.*, 2018, **69**, 85–122.

258 J. Sun, J. R. Bankston, J. Payandeh, T. R. Hinds, W. N. Zagotta and N. Zheng, Crystal structure of the plant dual-affinity nitrate transporter NRT1.1, *Nature*, 2014, **507**, 73–77.

259 J. L. Parker and S. Newstead, Molecular basis of nitrate uptake by the plant nitrate transporter NRT1.1, *Nature*, 2014, **507**, 68–72.

260 M. C. Chan, E. Procko and D. Shukla, Structural rearrangement of the serotonin transporter intracellular gate induced by Thr276 phosphorylation, *ACS Chem. Neurosci.*, 2022, **13**, 933–945.

261 H. Zhang, J. He, G. Hu, F. Zhu, H. Jiang, J. Gao, H. Zhou, H. Lin, Y. Wang, K. Chen, F. Meng, M. Hao, K. Zhao, C. Luo and Z. Liang, Dynamics of post-translational modification inspires drug design in the kinase family, *J. Med. Chem.*, 2021, **64**, 15111–15125.

262 E. L. Wu, X. Cheng, S. Jo, H. Rui, K. C. Song, E. M. Dávila-Contreras, Y. Qi, J. Lee, V. Monje-Galvan, R. M. Venable, J. B. Klauda and W. Im, CHARMM-GUI Membrane Builder toward realistic biological membrane simulations, *J. Comput. Chem.*, 2014, **35**, 1997–2004.

263 D. J. Smith, J. B. Klauda and A. J. Sodt, Simulation best practices for lipid membranes [Article v1.0], *Living J. Comp. Mol. Sci.*, 2019, **1**, 5966.

264 M. Audagnotto and M. Dal, Peraro, Protein post-translational modifications: in silico prediction tools and molecular modeling, *Comput. Struct. Biotechnology J.*, 2017, **15**, 307–319.

265 O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko and K.-R. Müller, Machine learning force fields, *Chem. Rev.*, 2021, **121**, 10142–10186.

266 D. L. Mobley, C. C. Bannan, A. Rizzi, C. I. Bayly, J. D. Chodera, V. T. Lim, N. M. Lim, K. A. Beauchamp, D. R. Slochower, M. R. Shirts, M. K. Gilson and P. K. Eastman, Escaping atom types in force fields using direct chemical perception, *J. Chem. Theory Comput.*, 2018, **14**, 6076–6092.

267 J. S. Smith, O. Isayev and A. E. Roitberg, ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost, *Chem. Sci.*, 2017, **8**, 3192–3203.

268 A. Schönichen, B. A. Webb, M. P. Jacobson and D. L. Barber, Considering protonation as a posttranslational modification

regulating protein structure and function, *Annu. Rev. Biophys.*, 2013, **42**, 289–314.

269 R. Kawade, D. Kuroda and K. Tsumoto, How the protonation state of a phosphorylated amino acid governs molecular recognition: insights from classical molecular dynamics simulations, *FEBS Lett.*, 2020, **594**, 903–912.

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

271 R. F. Service, 'The game has changed.' AI triumphs at protein folding, *Science*, 2020, **370**, 1144–1145.

272 J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli and D. Hassabis, Highly accurate protein structure prediction with AlphaFold, *Nature*, 2021, **596**, 583–589.

273 M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G. R. Lee, J. Wang, Q. Cong, L. N. Kinch, R. D. Schaeffer, C. Millán, H. Park, C. Adams, C. R. Glassman, A. DeGiovanni, J. H. Pereira, A. V. Rodrigues, A. A. van Dijk, A. C. Ebrecht, D. J. Opperman, T. Sagmeister, C. Buhlheller, T. Pavkov-Keller, M. K. Rathinaswamy, U. Dalwadi, C. K. Yip, J. E. Burke, K. C. Garcia, N. V. Grishin, P. D. Adams, R. J. Read and D. Baker, Accurate prediction of protein structures and interactions using a three-track neural network, *Science*, 2021, **373**, 871–876.

274 E. S. Groban, A. Narayanan and M. P. Jacobson, Conformational changes in protein loops and helices induced by post-translational phosphorylation, *PLoS Comput. Biol.*, 2006, **2**, e32.

275 H. Bagdonas, C. A. Fogarty, E. Fadda and J. Agirre, The case for post-predictional modifications in the AlphaFold protein structure database, *Nat. Struct. Mol. Biol.*, 2021, **28**, 869–870.

276 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, The Protein Data Bank, *Nucleic Acids Res.*, 2000, **28**, 235–242.

277 I. Bludau, S. Willems, W.-F. Zeng, M. T. Strauss, F. M. Hansen, M. C. Tanzer, O. Karayel, B. A. Schulman and M. Mann, The structural context of posttranslational modifications at a proteome-wide scale, *PLoS Biol.*, 2022, **20**, e3001636.

278 V. Csizmok and J. D. Forman-Kay, Complex regulatory mechanisms mediated by the interplay of multiple post-translational modifications, *Curr. Opin. Struct. Biol.*, 2018, **48**, 58–67.

279 A. Sethi, J. Eargle, A. A. Black and Z. Luthey-Schulten, Dynamical networks in tRNA:protein complexes, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 6620–6625.

280 Z. Liang, G. M. Verkhivker and G. Hu, Integration of network models and evolutionary analysis into high-throughput modeling of protein dynamics and allosteric regulation: theory, tools and applications, *Brief. Bioinform.*, 2020, **21**, 815–835.

281 M. C. R. Melo, R. C. Bernardi, C. de la Fuente-Nunez and Z. Luthey-Schulten, Generalized correlation-based dynamical network analysis: a new high-performance approach for identifying allosteric communications in molecular dynamics trajectories, *J. Chem. Phys.*, 2020, **153**, 134104.

282 Z. R. Thornburg, D. M. Bianchi, T. A. Brier, B. R. Gilbert, T. M. Earnest, M. C. R. Melo, N. Safronova, J. P. Sáenz, A. T. Cook, K. S. Wise, C. A. Hutchison, H. O. Smith, J. I. Glass and Z. Luthey-Schulten, Fundamental behaviors emerge from simulations of a living minimal cell, *Cell*, 2022, **185**, 345–360.e28.

283 Z. Luthey-Schulten, Integrating experiments, theory and simulations into whole-cell models, *Nat. Methods*, 2021, **18**, 446–447.

284 Z. Luthey-Schulten, Z. R. Thornburg and B. R. Gilbert, Integrating cellular and molecular structures and dynamics into whole-cell models, *Curr. Opin. Struct. Biol.*, 2022, **75**, 102392.

285 G. Kustatscher, T. Collins, A.-C. Gingras, T. Guo, H. Hermjakob, T. Ideker, K. S. Lilley, E. Lundberg, E. M. Marcotte, M. Ralser and J. Rappsilber, Understudied proteins: opportunities and challenges for functional proteomics, *Nat. Methods*, 2022, 1–6.

286 A. T. Weigle, M. Carr and D. Shukla, Impact of increased membrane realism on conformational sampling of proteins, *J. Chem. Theory Comput.*, 2021, **17**, 5342–5357.