# Markov State Models of Molecular Simulations to Study Protein Folding and Dynamics

**Vivek Junghare, Sourya Bhattacharya, Khalid Ansari, and Saugata Hazra**

**Abstract** Proteins are essential units of life that govern several functions. Understanding their behavior is closely related to their conformations, native folds, and change in conformations. Thus, the dynamic information of protein becomes essential to understand its properties at the molecular level. The molecular dynamics (MD) simulation approach provides atomistic-level dynamic information about proteins. However, more extended or complex MD simulations of protein are challenging to analyze and to gather meaningful confirmation from several snapshots of the dynamic system. To achieve it, i.e., analyzing MD simulation data, Markov State Model (MSM) is a powerful tool that has a statistical background. It represents the MD simulation system as a combination of finite memoryless states, i.e., states that are not dependent on prior states and transition probability among such states. MSM applications have grown from peptides to membrane protein simulations. The present book chapter sheds light on MD simulation's role in protein dynamics and why MSM is required. The brief theoretical aspects of MSM techniques are demonstrated. Lastly, the chapter discusses the application of MSM in different protein folding and dynamics.

**Keywords** Molecular dynamics (MD) simulation · Markov state model · Dynamics · Sampling · Statistical approach · Transition count matrix

V. Junghare · S. Bhattacharya
Department of Biosciences and Bioengineering, Indian Institute of Technology Roorkee, Roorkee, India

K. Ansari
Department of Physics, Indian Institute of Technology Roorkee, Roorkee, India

S. Hazra (✉)
Department of Biosciences and Bioengineering, Indian Institute of Technology Roorkee, Roorkee, India

Center for Nanotechnology, Indian Institute of Technology Roorkee, Roorkee, India
e-mail: saugata.hazra@bt.iitr.ac.in

# 1   Introduction

Protein dynamics and folding have been challenging phenomena essential for the molecular-level understanding of protein function. Molecular dynamics (MD) simulation is a valuable tool that comprehends macromolecular structural and functional insights. Data assembled after the MD simulation study can confer good knowledge about the macromolecular structure and provide detailed informational insights [1].

## 1.1   *Importance of Molecular Dynamics*

Proteins and nucleic acids are dynamic entities, and their dynamics play a significant role in their functions. Crystal structures stored at the PDB provide a halfway and limited perspective on three-dimensional (3D) construction. Especially protein molecules undergo crucial conformational changes during a particular function [2, 3]. One such change is the structural rearrangement in the protein molecule upon binding a substrate or inhibitor [4, 5]. This can be effectively verified by comparing apo and ligand-bound 3D protein structures. The conformational changes are usual parameters of enzymes' catalytic mechanisms [6]. One of the common instances is loop movement or domain rearrangements that change the local composition of the active site's chemical environment to perform a function. Sometimes, these alterations activate the catalytic process by bringing protein subunits together. Moreover, one can correlate protein function only when dynamic properties are considered [7–9].

There are several ways to deal with the conformation correlated with the relevant macromolecular function. One of the conventional ways is to gather experimentally determined structures covering the conformational space using X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, or cryo-electron microscopy (cryo-EM) methods. These methods can be used to study structures of macromolecules in different environments or bound with other substrates or ligands. However, these experimental studies are time taking and need specific high-end instruments.

On the other hand, theoretical strategies are the most helpful method for getting an image of the macromolecular dynamic properties of a protein. Protein folding occurs in a timescale of a few microseconds, allosteric transitions in microseconds to milliseconds, relative motions of protein domains in nanoseconds to seconds, and dynamics of side chains in picoseconds to nanoseconds (Fig. 1) [10]. Additionally, it is observed that longer timescale motions can influence shorter timescale dynamics and vice versa. Hence, long timescale simulations have always been a well-chosen option [11, 12]. Long-time simulations provide an opportunity to understand the flexibility of proteins and their related ensemble of alternative structural states, which are crucial for understanding the folding and dynamics of proteins [13, 14].
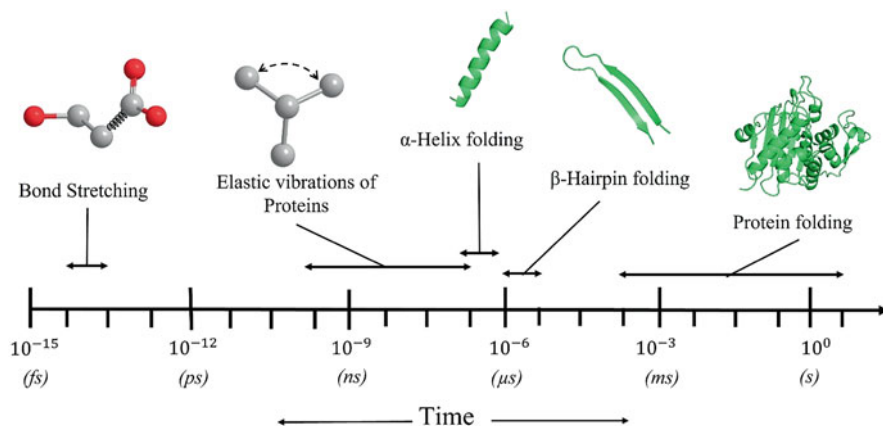
**Fig. 1** The figure represents the protein motion concerning the time axis. The MD simulations must be performed at femtosecond time steps to capture the bond stretching motion and similarly for other represented motions that are more time-scaled atomistic. Hence, more computational power is required

Protein-conformational changes play a vital role in its functioning [15, 16]. Hence it is not enough to study just one PDB conformer. Modern-day advances in simulation algorithms and calculations have promoted the idea of "conformational ensembles" as an option in contrast to examining a single structure from PDB. These ensembles or conformers can be examined to determine thermodynamic properties, entropy, free energy, conformational changes, or protein folding phenomenon [16–18]. There are two significant difficulties in analyzing MD simulations of biomolecules: adequate conformational sampling and exact physical force fields. Despite remarkable improvements in modern computing capacity, conventional MD (cMD) simulations are still essentially constrained to shorter timescales than those demonstrated by various biomolecular movements and functions [19–22]. Hence, to gather multiple conformations, a specified tool is required.

Furthermore, protein folding remains one of biology's fundamental and least understood phenomena. This fascinating phenomenon of conversion of the primary sequence of a protein to the native 3D structure remains less understood. Small molecular weight proteins with ~10–100 amino acid residues fold in the microsecond to sub-millisecond timescales, known as "fast-folding" proteins. They are magnificent model systems to study and analyze protein folding through long timescale cMD simulations in explicit water [23]. Protein folding needs a broad measure of conformational examination and computational ability to describe the free energy landscape appropriately. Advancement in computation with more extended simulations is insufficient to expand the conformational sampling in the molecular framework. The complicated state of the free energy landscape makes the majority of the simulations investigate only a small region around the energy least near to the initial conformation. With the accessibility of the current advanced HPC systems, a conspicuous methodology is to play out a series of parallel simulations

with several initial energy-minimized conformations. Although this could be proficient, it requires detailed information on the framework to simulate and cannot be applied as an overall strategy.

Nevertheless, protein folding has been analyzed using cMD and utilizing productive examining methods such as replica-exchange MD [24], Markov State Models (MSM), biasing MD simulations such as bias-exchange metadynamics [25], and transition path sampling [26]. This chapter sheds light on how MSM helps tackle protein dynamics and folding problems.

## 1.2 Motivation Behind Using MSM Technique

At times, protein folding and dynamics require long timescale simulations, or the system becomes highly complex or enormous (such as in the case of membrane protein simulation). The first microsecond-length all-atom MD simulation of a small protein was carried out by Duan and Kollman [27]. Further advancements in computer power open up possibilities of MD simulations of thousands of protein atoms, long time-scaled simulation of proteins, etc. Biomacromolecules frequently perform their functions through dynamic transitions between conformational states. For instance, the AdeB efflux pump undergoes carbapenem resistance through conformational modifications [28]. By performing long timescale dynamics based on several short MD simulations, MSM has emerged as a prominent method for bridging this timescale gap [2, 29].

Representing physical, chemical, or biological systems using stochastic processes is standard practice. The objective is to analyze the stochastic model and roughly compute the exciting properties of the system. Direct sampling and building a coarse-grained model of the system are two methods for carrying out such analysis. In a direct sampling strategy, one attempt to produce a statistically significant number of occurrences representing the system property in question. Here, making sufficient statistics for accurate estimates requires much computation. Estimation through direct numerical simulation is impossible, especially if the state space is continuous and has a high dimension [30]. In the coarse-grained model, discretization of the systems state space is used. This is achievable using MSM. The advantage is that it uses discrete finite space. Due to this, the vast systems became finite discrete models that can be solved numerically to find their properties. It uses transition path theory (TPT) to analyze systems' discrete states. In summary, the analysis of the ensemble of reactive trajectories, or trajectories that originate from a specific set of states A and go to B. Hence using such a technique provides a more comprehensive analysis of biological protein simulations.
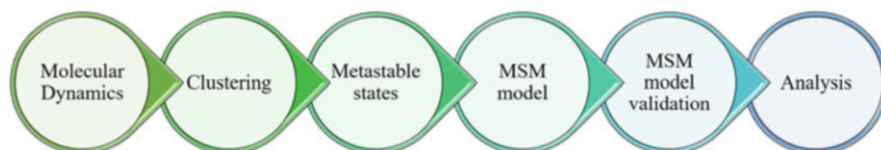
## 2    Markov State Model

A theoretical model, often known as the Markov State Model (MSM), is frequently used to study the dynamic nature of biological systems. The basic idea of MSM is making the square matrix known as the transition probability matrix (TPM). In the case of protein dynamics, MSM can be used after obtaining initial data from MD simulation trajectories.

### 2.1    Building of MSM

To develop MSM, an adaptive sampling algorithm is frequently used. Adaptive sampling is a statistical approach for solving protein dynamics on large timescales (100 μs to the ms) to sample conformational transitions. The adaptive sampling algorithm is based on iterations, which are used until the desired sampling criteria are reached [19]. The adaptive sampling process is divided into three steps: (i) to run an MD simulation and get many short trajectories, (ii) build an MSM using trajectories, and (iii) run a simulation trajectory based on obtained results from the MSM. MSM uses a matrix, so it needs microstates that can be prepared in two ways: one is based on geometric distributions (distance metric), and the other is based on a free energy map (kinetic-based metric). The preferred one is to choose free energy minima, i.e., kinetic distribution, instead of the geometric distribution. The pathway of MSM is illustrated in Fig. 2.

### 2.2    Microstates and Macrostates Generation

Microstates are required to construct MSM. They are the nonoverlapping discrete configurational space. Every transition among these microstates is not dependent on the previous state. This phenomenon is known as memoryless transition. In this regard, one needs microstates where shifts can happen smoothly and rapidly. For this, there is a requirement to group configurations, often known as clustering. Since many clustering techniques are available, one must choose them wisely. One of the clustering techniques is choosing a distance metric. The k-centers, k-medoids, and



**Fig. 2**  The schematic pathway of the Markov State Model (MSM)

hybrid k-centers/k-medoids clustering are some of the essential clustering algorithms. To determine states, one needs to go through the MD simulations first and then find the suitable conformations based on either the root mean square deviation (RMSD) chosen appropriately 2 to 3 Å or based on the energy barriers. Most of the time, it is assumed that as the degree of structural similarity is higher, the corresponding kinetic similarity is also higher. It is known as the kinetic clustering of microstates into larger macrostates [31].

In Markovian microstate formation, there is a timeframe difference at which the states occur, often known as lag time or Markovian lag ($\tau$). Hence, after lag time $\tau$, the state will not be dependent on the previous state. MSM building requires a transition probability among these microstates, which depends on the number of microstates and lag time. Markovian lag should be large enough but not too large so that it does not alter significantly from other trajectories, which are often considered microstates. Markovian lag is just a method of selecting steps for trajectories that must be chosen carefully.

Additionally, in the case of tens of thousands of microstates or huge system sizes (such as membrane protein simulation), kinetic-based clustering can be performed that are supersets of microstates and are named macrostates. These macrostates are obtained using coarse-graining the model. This method collects microstates that are quickly clumping together and are collected to form macrostates. Available lumping procedures from microstates to macrostates are perron cluster cluster analysis (PCCA), their improved version (PCCAC), Bayesian agglomerative clustering engine (BACE), and super level set hierarchical clustering (SHC).

## 2.3  MSM Model and Validation

After obtaining the microstates, the next step is constructing the transition count matrix (TCM). It is a matrix that describes the transition from one state to another. The transition count matrix in general form is shown below:

$$M = \begin{bmatrix} a_{11} & a_{12}\ldots\ldots & a_{1n} \\ a_{21} & a_{22}\ldots\ldots & a_{2n} \\ \vdots & \ddots \quad \cdots \quad \cdots & \\ a_{n1} & a_{n2} \quad \ddots \quad \cdots & a_{nn} \end{bmatrix}$$

where $a_{ij}$ denotes the transition from $i$th state to $j$th state. For example, if the states chosen from trajectories named $A$, $B$, and the trajectory are given as:

Trajectory : *AABBBABABAABB*.

Also, if the trajectory is chosen one step, then the number of transitions from $A$ to $A$ is 2 ($N_{AA} = 2$), from $A$ to $B$ is 4 ($N_{AB} = 4$), from $B$ to $A$ is 3 ($N_{BA} = 3$), and from $B$ to $B$ is 3 ($N_{BB} = 3$). Then the TCM can be written as mentioned in Table 1.

**Table 1** Transition count matrix representing the transition between states *A* and *B*

| From\To | A | B |
|---------|---|---|
| A | 2 | 4 |
| B | 3 | 3 |

**Table 2** Transpose of the transition count matrix

| From\To | A | B |
|---------|---|---|
| A | 2 | 3 |
| B | 4 | 3 |

The transition count matrix is usually not symmetric, so it is necessary to make a symmetric matrix and any symmetric matrix. One must follow the symmetry property of the matrix, which is defined as any (square) matrix. It is written as the sum of a symmetric matrix and an antisymmetric matrix [32].

$$M = \frac{[M + M^T]}{2} + \frac{[M - M^T]}{2}$$

where $M^T$ is the transpose of $M$, $[M + M^T]$ is symmetric, and $[M - M^T]$ is antisymmetric.

This matrix should be symmetric because the transition between states depends not only on the forward direction but also on the reverse direction and is transposable. The transpose matrix describes moving from one state to another in either a forward or reverse direction. The transpose of TCM is shown below:

$$M^T = \begin{bmatrix} a_{11} & a_{21}......... & a_{n1} \\ a_{12} & a_{22}......... & a_{n2} \\ \vdots & \ddots \quad ...... \quad ... & \\ a_{1n} & a_{2n} \quad \ddots \quad ... & a_{nn} \end{bmatrix}$$

For the transpose matrix, the row (horizontal elements) is changed into a column (vertical components) and vice versa, as shown in Table 2.

Averaging the transition matrix counts by adding a transition matrix, and their transpose matrix gives symmetry.

$$M^{symm} = \frac{M + M^T}{2}$$

The symmetry matrix is shown below:

$$M_{ij}^{symm} = \frac{1}{2} \begin{bmatrix} a_{11}+a_{11} & a_{12}+a_{21}......... & a_{1n}+a_{n1} \\ a_{21}+a_{12} & a_{22}+a_{22}......... & a_{2n}+a_{n2} \\ \vdots & \ddots \quad ...... \quad ... & \\ a_{n1}+a_{1n} & a_{n2}+a_{2n} \quad \ddots \quad ... & a_{nn}+a_{nn} \end{bmatrix}$$

For the present example, the symmetric matrix is shown in Table 3.

After this, reversible TPM will be calculated for each element of the matrix. There are two requirements for the TPM that must be rigorously followed. First, the total

**Table 3** Symmetry matrix for present trajectory

| From\To | A | B |
|---------|---|---|
| A | 2 | 3.5 |
| B | 3.5 | 3 |

**Table 4** The transition probability matrix for a given trajectory

| From\To | A | B |
|---------|---|---|
| A | 0.222 | 0.778 |
| B | 0.636 | 0.364 |

probability in each row is equal to unity, and second, elements should be nonnegative. There is no negative value meaning because probability only contains values between zero and one. Another essential point about transition probability is that it depends only on the time difference, i.e., the transition should be homogeneous [33].

$$P_{ij} = \frac{M_{ij}^{symm}}{\sum_i^j \left( M_{ij}^{symm} \right)}$$

The transition probability matrix is shown below:

$$M_{prob} = \begin{bmatrix} P_{11} & P_{12}........ & P_{1n} \\ P_{21} & P_{22}........ & P_{2n} \\ \vdots & \ddots \quad \cdots \quad \cdots & \\ P_{n1} & P_{n2} \quad \ddots \quad \cdots & P_{nn} \end{bmatrix}$$

For the present example, the transition probability matrix will be shown in Table 4.

$$\text{Auxiliary equation} : |M - \lambda I| = 0;$$

where $I$ is an identity matrix, and $\lambda$ is for eigenvalues.

After solving the auxiliary equation for the TPM, one can get the eigenvectors and corresponding eigenvalues. The total sum of eigenvalues is to be zero. From eigenvalues data, one can analyze that the most positive value gives the most fluctuation from the equilibrium states, and the least negative value is in the most equilibrium states. There are several methods and tests to validate the models, such as Chapman–Kolmogorov equation model-based test, correlation function test, Bayesian Model selection, Swope–Pitera eigenvalue test, etc.

## 3 MSM to Understand Protein Folding and Dynamics

The initial studies of using MSM were started by studying peptide folding [34–36] and other small systems [37]. Further, it was applied in protein folding, protein–ligand binding, nucleic acids, and other biological problems (Fig. 3). It is used to

analyze small-timescale and large-timescale simulations to gather relevant information. We now discuss how MSM is used to understand protein folding and dynamics, focusing on ensemble sampling and conformational fluctuations.

## 3.1 Peptide Modeling

Researchers have tried to address the issues related to understanding the mechanism of protein folding and finding the nature of folds. MD simulations have been regularly used along with experimental studies. In 2004, Swope et al. developed an algorithm to study the kinetics of protein folding. They applied it to a small peptide, a C-terminal alpha-hairpin motif from protein G. They used a Boltzmann-weighted ensemble to formulate the transition function from MD simulation [35]. They found the pattern and number of hydrogen bonds in a peptide. The Markov model depends on finding the finite number of metastable states; thus, identifying them is a critical and essential step. Hence, the clustering algorithm
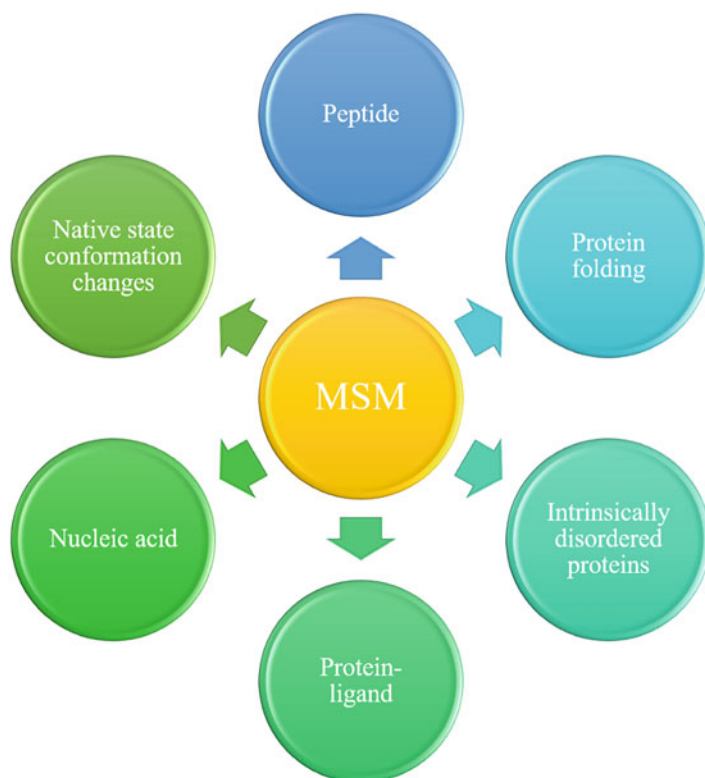


**Fig. 3** Applications of MSM in protein folding and protein dynamics

was applied to get kinetics-based states that were long-lived in dynamic systems. This kinetics-based clustering was used by Noe et al., who tested $ALA_8$ and $ALA_{12}$ peptides [36]. This study, by Noé et al., brought a new direction to form metastable states, which consider dynamic behavior and not geometric proximity. Following this method, the automated algorithm was proposed, which detects the kinetically metastable states and was tested on three peptides [38]. After this, the master equation was developed by Buchete & Hummer for studying MD simulation of peptide folding at an atomistic level [39]. $ALA_5$ peptide was used for the study, which was intended to form a small helix. In recent studies, this technique has been used to study peptides like amyloid-β peptide (Aβ), which is responsible for Alzheimer's disease [40].

## 3.2   Protein Folding

Protein folding prediction through an in silico approach has been a mystery since the inception of protein simulation. Protein folds have numerous possibilities, as stated by the Leventhal paradox [41]. However, protein folds within a few microseconds in natural states and retains its native fold to function [42]. At the same time, predicting protein folding, understanding different folding conformations, and the folding rates also matter [43]. Several mechanisms have been proposed to explain the protein folding process, from a simple two-state model [44] to more complex models [45]. Also, it has been observed that some proteins do not fold and exist in an intrinsically disordered state [46].

Additionally, the misfolding of protein also occurs and has been observed in neurodegenerative disorders [47]. Thus, gathering the information on the folded and unfolded states is not enough, but the intermediate, misfolded, and disordered should also be analyzed. MSM uses the MD simulation data to find transition probability between different finite states. Initially, the model is constructed using geometric conformation similarity [48, 49]. The obvious choice is to use RMSD between the conformations by limiting it to a smaller cut-off value [50]. However, the RMSD is based on a protein backbone and is used to generate distance metrics. Hence, side chain and dihedral angle flip may hinder the results. The assumption is that the conformations with smaller deflections may have similar kinetic stability. However, finding more kinetically relevant metastable states should be carried out. Different clustering algorithms have been used [51], such as k-centers clustering, k-medoids clustering, and a hybrid of both k-clustering methods. The k-center clustering algorithm aims to find clusters with approximately the same radius and map different conformations to the nearest center of the cluster so that the distance from a distance is minimum. Li et al. & Voelz et al. used this clustering algorithm to improve the microstate generation efficiently [29, 52]. In the case of k-medoid clustering, the optimization is performed for the average distance between the center and other cluster points. In protein folding, this algorithm creates many clusters in the folded

scenario and very few in the unfolded system [53]. The hybrid approach of both the k-clustering techniques was used to build MSMBuilder2 [54].

## 3.3  Protein–Ligand Binding

Analyzing the interaction of a protein with its substrate/inhibitor can provide critical information about the protein's function [5, 55]. The binding of small molecules to proteins or detecting new binding sites could be performed using MSM methods. Earlier, binding kinetics has been studied by constructing MSM to find long-lived intermediates of trypsin inhibitors [56]. The induced fit model (conformation changes due to ligand binding) and conformation selection model (ligand bind to protein without changing in protein's conformation) are used to detect protein–ligand recognition [57–59]. But later, it was observed that both are found in real-life scenarios [60–62]. In an earlier study to find the contribution of both methods, an analytical model based on a three-pronged approach of MD simulation, flux, and MSM was developed [63]. The choline-binding protein (ChoX) was used as a case study, and MD and MSM methods were used to find parameters for flux analysis [61].

## 3.4  Analyzing Intrinsically Disordered Proteins

Intrinsically Disordered Proteins (IDPs) are proteins that do not have a stable 3D structure. They bind to nucleic acids or other proteins for their functions. IDPs are dynamic ensembles that continuously change their internal conformation with high structural heterogeneity [64, 65]. However, IDPs are responsible for several cellular functions and are involved in many diseases like diabetes, cancer, neurodegenerative diseases, and cardiovascular diseases [66–69]. While interacting with partners, IDPs are coupled binding and folding reactions, which is essential for their function. Similar to ligand binding, induced fit, conformational selection, and a combination of both models are used to study IDPs. However, the kinetics of the binding-folding reaction, specifically binding to a partner or conformation without a partner, requires detailed investigation [70, 71]. Here, MD simulation can provide a contemporary way to analyze IDP folding at the atomistic level. To achieve this, MD simulations of IDPs should be performed so that the whole binding-folding pathways can be analyzed. Such simulation trajectories are complex to study; however, MSM techniques can help to identify metastable states in the pathway and the transition probability [72, 73].

## 3.5   Native State Conformation Changes

Generally, the rational structure-based drug design does not take into account protein-conformational changes. Approximately 15% of proteins have deep active sites related to their activity [74]. Hence, conformational heterogeneity is essential to understand protein behavior. This could provide information on the novel active sites or transient catalytic sites, which are allosteric or can block protein–protein interaction [75–78]. Since MD simulation can provide the system's dynamical behavior, if coupled with MSM, it can provide a set of ensembles where the metastable state is in an equilibrium state. Also, the advancement in MSM to capture kinetic and thermodynamic properties makes it a more viable option to identify the transient active site. There are several examples where similar approaches have been used to find cryptic pockets and allosteric sites. Among such studies, the TEM-1 beta-lactamase was used and observed that several such allosteric sites were present [79]. Such studies could also be performed with novel proteins to find active or allosteric sites.

## 4   Summary

Advancements in computational power, such as parallel programming and GPUs, have made the MD simulation more achievable. However, analyzing the simulation data is challenging. MSM is based on finite ensembles and uses clustering methods to create ensembles. Before MSM, geometric clustering was used, but MSM provides enhanced metastable states, which means it is the kinetic energy-based state. It is a coarse-graining of a system's dynamics, which depicts the underlying free energy landscape that governs the system's structure and dynamics. Identifying states in a kinetically relevant scheme and effectively using state decomposition to construct a transition matrix are the two main issues for creating an MSM. To build the MSM model, the traditional geometric clustering method is used to develop microstates. These microstates are further used to build a transition matrix. This step takes care of finding kinetically related microstates. This information is used to build MSM. However, adaptive sampling is used to improve the MSM model. Further, validations can be done by Bayesian Model selection, Swope–Pitera eigenvalue test, and other such tests (Fig. 4).

Protein folding and the dynamics of the native 3D structure are critical biological phenomena [80]. MD simulation can provide a way to understand these processes in millisecond simulations [81–83]; however, analyzing such data requires sophisticated protocols and methods [84, 85]. MSM provides a convenient and interpretable solution [86]. With the current advancement in computational power and algorithm, the use of MSM has increased and will continue to grow. This technique can also analyze and comprehend complicated systems such as membrane proteins, peptide
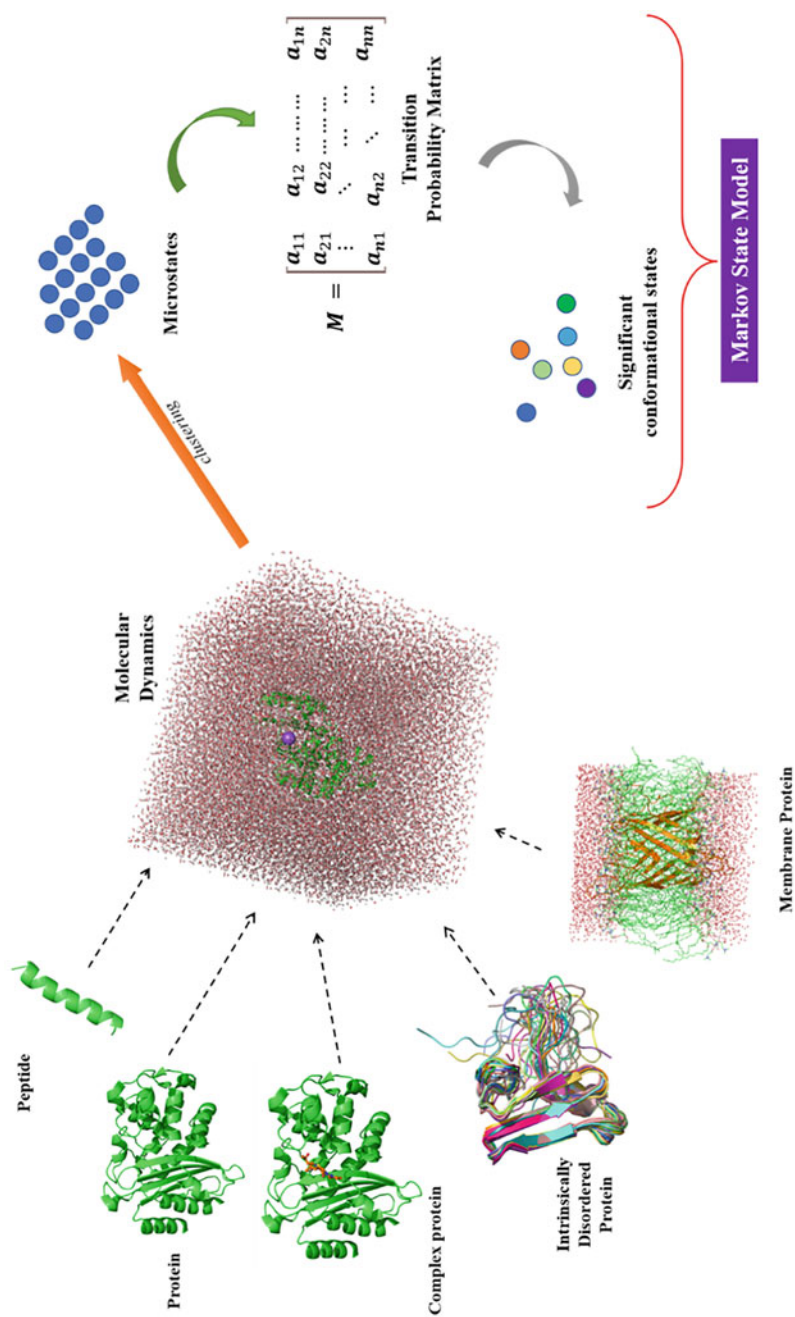
**Fig. 4** Various applications which could be studied using Markov State Model

folding, IDPs, and other biological systems; hence, it is emerging as a critical in silico approach.

# References

1. J. Gelpi, A. Hospital, R. Goñi, M. Orozco, Molecular dynamics simulations: advances and applications. Adv. Appl. Bioinforma. Chem. **8**, 37 (2015). https://doi.org/10.2147/AABC. S70333

2. J. Chodera, F. Noé, Markov state models of biomolecular conformational dynamics. Curr. Opin. Struct. Biol. **25**, 135–144 (2014)

3. S. Hazra, A. Szewczak, S. Ort, M. Konrad, A. Lavie, Post-translational phosphorylation of serine 74 of human deoxycytidine kinase favors the enzyme adopting the open conformation making it competent for nucleoside binding and release. Biochemistry **50**, 2870–2880 (2011). https://doi.org/10.1021/bi2001032

4. M.F. Chek, S.-Y. Kim, T. Mori, H.T. Tan, K. Sudesh, T. Hakoshima, Asymmetric open-closed dimer mechanism of polyhydroxyalkanoate synthase PhaC. IScience **23**, 101084 (2020). https://doi.org/10.1016/j.isci.2020.101084

5. S. Hazra, H. Xu, J.S. Blanchard, Tebipenem, a new Carbapenem antibiotic, is a slow substrate that inhibits the β-lactamase from *mycobacterium tuberculosis*. Biochemistry **53**, 3671–3678 (2014). https://doi.org/10.1021/bi500339j

6. M. Kokkinidis, N.M. Glykos, V.E. Fadouloglou, Protein flexibility and enzymatic catalysis. Adv. Protein. Chem. Struct. Biol. **87**, 181–218 (2012). https://doi.org/10.1016/B978-0-12-398312-1.00007-X

7. M. Karplus, Role of conformation transitions in adenylate kinase. Proc. Natl. Acad. Sci. U S A **107**, E71 (2010). https://doi.org/10.1073/pnas.1002180107

8. R.C. Stevens, W.N. Lipscomb, Allosteric control of quaternary states in E. coli aspartate transcarbamylase. Biochem. Biophys. Res. Commun. **171**, 1312–1318 (1990). https://doi.org/10.1016/0006-291X(90)90829-C

9. S. Bhattacharya, A.K. Padhi, V. Junghare, N. Das, D. Ghosh, P. Roy, K.Y.J. Zhang, S. Hazra, Understanding the molecular interactions of inhibitors against Bla1 beta-lactamase towards unraveling the mechanism of antimicrobial resistance. Int. J. Biol. Macromol. **177**, 337–350 (2021). https://doi.org/10.1016/j.ijbiomac.2021.02.069

10. S.A. Adcock, J.A. McCammon, Molecular dynamics: survey of methods for simulating the activity of proteins. Chem. Rev. **106**, 1589–1615 (2006). https://doi.org/10.1021/cr040426m

11. K.A. Henzler-Wildman, M. Lei, V. Thai, S.J. Kerns, M. Karplus, D. Kern, A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. Nature **450**, 913–916 (2007). https://doi.org/10.1038/nature06407

12. S. Hammes-Schiffer, S.J. Benkovic, Relating protein motion to catalysis. Annu. Rev. Biochem. **75**, 519–541 (2006). https://doi.org/10.1146/ANNUREV.BIOCHEM.75.103004.142800

13. G.M. Lee, C.S. Craik, Trapping moving targets with small molecules. Science **324**(2009), 213–215 (1979). https://doi.org/10.1126/SCIENCE.1169378

14. S.J. Teague, Implications of protein flexibility for drug discovery. Nat. Rev. Drug Discov. **2**, 527–541 (2003). https://doi.org/10.1038/nrd1129

15. M. Pal, S. Bhattacharya, G. Kalyan, S. Hazra, Cadherin profiling for therapeutic interventions in epithelial Mesenchymal transition (EMT) and tumorigenesis. Exp. Cell Res. **368**, 137–146 (2018). https://doi.org/10.1016/j.yexcr.2018.04.014

16. G. Kalyan, V. Junghare, S. Bhattacharya, S. Hazra, Understanding structure-based dynamic interactions of antihypertensive peptides extracted from food sources. J. Biomol. Struct. Dyn. **39**, 635–649 (2021). https://doi.org/10.1080/07391102.2020.1715836

17. M.C. Baxa, E.J. Haddadian, J.M. Jumper, K.F. Freed, T.R. Sosnick, Loss of conformational entropy in protein folding calculated using realistic ensembles and its implications for NMR-based calculations. Proc. Natl. Acad. Sci. **111**, 15396–15401 (2014). https://doi.org/10.1073/pnas.1407768111

18. A.N. Naganathan, M. Orozco, The native ensemble and folding of a protein molten-globule: Functional consequence of downhill folding. J. Am. Chem. Soc. **133**, 12154–12161 (2011). https://doi.org/10.1021/ja204053n

19. T.J. Lane, D. Shukla, K.A. Beauchamp, V.S. Pande, To milliseconds and beyond: challenges in the simulation of protein folding. Curr. Opin. Struct. Biol. **23**, 58–65 (2013). https://doi.org/10.1016/j.sbi.2012.11.002

20. S. Piana, J.L. Klepeis, D.E. Shaw, Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. Curr. Opin. Struct. Biol. **24**, 98–105 (2014). https://doi.org/10.1016/j.sbi.2013.12.006

21. S. Bhattacharya, V. Junghare, N.K. Pandey, D. Ghosh, H. Patra, S. Hazra, An insight into the complete biophysical and biochemical characterization of novel class a beta-lactamase (Bla1) from bacillus anthracis. Int. J. Biol. Macromol. **145**, 510–526 (2020). https://doi.org/10.1016/j.ijbiomac.2019.12.136

22. S. Bhattacharya, V. Junghare, N.K. Pandey, S. Baidya, H. Agarwal, N. Das, A. Banerjee, D. Ghosh, P. Roy, H.K. Patra, S. Hazra, Variations in the SDN loop of class a beta-lactamases: a study of the molecular mechanism of BlaC (mycobacterium tuberculosis) to Alter the stability and catalytic activity towards antibiotic resistance of MBIs. Front. Microbiol. **12**, 710291 (2021). https://doi.org/10.3389/fmicb.2021.710291

23. K. Lindorff-Larsen, S. Piana, R.O. Dror, D.E. Shaw, How fast-folding proteins fold. Science **334**(2011), 517–520 (1979). https://doi.org/10.1126/science.1208351

24. Y. Sugita, Y. Okamoto, Replica-exchange molecular dynamics method for protein folding. Chem. Phys. Lett. **314**, 141–151 (1999). https://doi.org/10.1016/S0009-2614(99)01123-9

25. F. Marinelli, F. Pietrucci, A. Laio, S. Piana, A kinetic model of trp-cage folding from multiple biased molecular dynamics simulations. PLoS Comput. Biol. **5**, e1000452 (2009). https://doi.org/10.1371/journal.pcbi.1000452

26. J. Juraszek, P.G. Bolhuis, Sampling the multiple folding mechanisms of Trp-cage in explicit solvent. Proc. Natl. Acad. Sci. U S A **103**, 15859–15864 (2006). https://doi.org/10.1073/pnas.0606692103

27. Y. Duan, P.A. Kollman, Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. Science **282**(1998), 740–744 (1979). https://doi.org/10.1126/SCIENCE.282.5389.740

28. S. Roy, V. Junghare, S. Dutta, S. Hazra, S. Basu, Differential binding of carbapenems with the AdeABC efflux pump and modulation of the expression of AdeB linked to novel mutations within two-component system AdeRS in carbapenem-resistant acinetobacter baumannii. mSystems **7**, e0021722 (2022). https://doi.org/10.1128/msystems.00217-22

29. V.A. Voelz, G.R. Bowman, K. Beauchamp, V.S. Pande, Molecular simulation of *ab initio* protein folding for a millisecond folder NTL9(1−39). J. Am. Chem. Soc. **132**, 1526–1528 (2010). https://doi.org/10.1021/ja9090353

30. C. Hartmann, R. Banisch, M. Sarich, T. Badowski, C. Schütte, Characterization of rare events in molecular dynamics. Entropy **16**, 350–376 (2013). https://doi.org/10.3390/e16010350

31. V.S. Pande, K. Beauchamp, G.R. Bowman, Everything you wanted to know about Markov state models but were afraid to ask. Methods **52**, 99–105 (2010). https://doi.org/10.1016/j.ymeth.2010.06.002

32. G.B. Arfken, H.J. Weber, F.E. Harris, *Mathematical Methods for Physicists* (Elsevier, Amsterdam, 2013). https://doi.org/10.1016/C2009-0-30629-7

33. N.G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam, 2007). https://doi.org/10.1016/B978-0-444-52965-7.X5000-4

34. N. Singhal, C.D. Snow, V.S. Pande, Using path sampling to build better Markovian state models: predicting the folding rate and mechanism of a tryptophan zipper beta hairpin. J. Chem. Phys. **121**, 415 (2004). https://doi.org/10.1063/1.1738647

35. W.C. Swope, J.W. Pitera, F. Suits, M. Pitman, M. Eleftheriou, B.G. Fitch, R.S. Germain, A. Rayshubski, T.J.C. Ward, Y. Zhestkov, R. Zhou, Describing protein folding kinetics by molecular dynamics simulations. 2. Example applications to alanine dipeptide and a β-hairpin peptide. J. Phys. Chem. B **108**, 6582–6594 (2004). https://doi.org/10.1021/jp037422q

36. F. Noé, I. Horenko, C. Schütte, J.C. Smith, Hierarchical analysis of conformational dynamics in biomolecules: transition networks of metastable states. J. Chem. Phys. **126**, 155102 (2007). https://doi.org/10.1063/1.2714539

37. F. Noé, S. Doose, I. Daidone, M. Löllmann, M. Sauer, J.D. Chodera, J.C. Smith, Dynamical fingerprints for probing individual relaxation processes in biomolecular dynamics with simulations and kinetic experiments. Proc. Natl. Acad. Sci. U S A **108**, 4822–4827 (2011). https://doi.org/10.1073/pnas.1004646108

38. J.D. Chodera, N. Singhal, V.S. Pande, K.A. Dill, W.C. Swope, Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. J. Chem. Phys. **126**, 155101 (2007). https://doi.org/10.1063/1.2714538

39. N.-V. Buchete, G. Hummer, Coarse master equations for peptide folding dynamics. J. Phys. Chem. B **112**, 6057–6069 (2008). https://doi.org/10.1021/jp0761665

40. A. Paul, S. Samantray, M. Anteghini, M. Khaled, B. Strodel, Thermodynamics and kinetics of the amyloid-β peptide revealed by Markov state models based on MD data in agreement with experiment. Chem. Sci. **12**, 6652–6669 (2021). https://doi.org/10.1039/D0SC04657D

41. C. Levinthal, How to fold graciously, in *Mossbauer Spectroscopy in Biological Systems*, ed. by P. DeBrunner, J. Tsibris, E. Munck, (University of Illinois Press, Urbana, 1969), pp. 22–26

42. H. White Frederick, J. Bello, D. Harker, E. de Jarnette, Regeneration of native secondary and tertiary structures by air oxidation of reduced Ribonuclease. J. Biol. Chem. **236**, 1353–1360 (1961). https://doi.org/10.1016/S0021-9258(18)64176-6

43. V.A. Voelz, V.R. Singh, W.J. Wedemeyer, L.J. Lapidus, V.S. Pande, Unfolded-state dynamics and structure of protein L characterized by simulation and experiment. J. Am. Chem. Soc. **132**, 4702–4709 (2010). https://doi.org/10.1021/ja908369h

44. J. Kubelka, T.K. Chiu, D.R. Davies, W.A. Eaton, J. Hofrichter, Sub-microsecond protein folding. J. Mol. Biol. **359**, 546–553 (2006). https://doi.org/10.1016/j.jmb.2006.03.034

45. P.S. Kim, R.L. Baldwin, Intermediates in the folding reactions of small proteins. Annu. Rev. Biochem. **59**, 631–660 (1990). https://doi.org/10.1146/annurev.bi.59.070190.003215

46. H.J. Dyson, P.E. Wright, Coupling of folding and binding for unstructured proteins. Curr. Opin. Struct. Biol. **12**, 54–60 (2002). https://doi.org/10.1016/S0959-440X(02)00289-0

47. J. Hardy, D.J. Selkoe, The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. Science **297**(2002), 353–356 (1979). https://doi.org/10.1126/science.1072994

48. G.R. Bowman, X. Huang, V.S. Pande, Using generalized ensemble simulations and Markov state models to identify conformational states. Methods **49**, 197–201 (2009). https://doi.org/10.1016/j.ymeth.2009.04.013

49. M. Senne, B. Trendelkamp-Schroer, A.S.J.S. Mey, C. Schütte, F. Noé, EMMA: a software package for Markov model building and analysis. J. Chem. Theory Comput. **8**, 2223–2238 (2012). https://doi.org/10.1021/ct300274u

50. L.-T. Da, F.K. Sheong, D.-A. Silva, X. Huang, Application of Markov state models to simulate long timescale dynamics of biological macromolecules. Adv. Exp. Med. Biol. **805**, 29–66 (2014). https://doi.org/10.1007/978-3-319-02970-2_2

51. J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J.D. Chodera, C. Schütte, F. Noé, Markov models of molecular kinetics: generation and validation. J. Chem. Phys. **134**, 174105 (2011). https://doi.org/10.1063/1.3565032

52. Y. Li, Z. Dong, Effect of clustering algorithm on establishing Markov state model for molecular dynamics simulations. J. Chem. Inf. Model. **56**, 1205–1215 (2016). https://doi.org/10.1021/acs.jcim.6b00181

53. G.R. Bowman, An overview and practical guide to building Markov state models. Adv. Exp. Med. Biol. **797**, 7–22 (2014). https://doi.org/10.1007/978-94-007-7606-7_2

54. K.A. Beauchamp, G.R. Bowman, T.J. Lane, L. Maibaum, I.S. Haque, V.S. Pande, MSMBuilder2: modeling conformational dynamics on the picosecond to millisecond scale. J. Chem. Theory Comput. **7**, 3412–3419 (2011). https://doi.org/10.1021/ct200463m

55. G. Kalyan, V. Junghare, M.F. Khan, S. Pal, S. Bhattacharya, S. Guha, K. Majumder, S. Chakrabarty, S. Hazra, Anti-hypertensive peptide predictor: a machine learning-empowered web server for prediction of food-derived peptides with potential angiotensin-converting enzyme-I inhibitory activity. J. Agric. Food Chem. **69**, 14995–15004 (2021). https://doi.org/10.1021/acs.jafc.1c04555

56. U. Kahler, A.S. Kamenik, F. Waibl, J. Kraml, K.R. Liedl, Protein-protein binding as a two-step mechanism: preselection of encounter poses during the binding of BPTI and trypsin. Biophys. J. **119**, 652–666 (2020). https://doi.org/10.1016/j.bpj.2020.06.032

57. B. Ma, M. Shatsky, H.J. Wolfson, R. Nussinov, Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. Protein Sci. **11**, 184–197 (2002). https://doi.org/10.1110/ps.21302

58. D.E. Koshland, Application of a theory of enzyme specificity to protein synthesis. Proc. Natl. Acad. Sci. U S A **44**, 98–104 (1958). https://doi.org/10.1073/pnas.44.2.98

59. V. Junghare, R. Alex, A. Baidya, M. Paul, R.R. Alyethodi, G.S. Sengar, S. Kumar, U. Singh, R. Deb, S. Hazra, *In silico* modeling revealed new insights into the mechanism of action of enzyme 2′-5′-oligoadenylate synthetase in cattle. J. Biomol. Struct. Dyn. **40**, 14013–14026 (2021). https://doi.org/10.1080/07391102.2021.2001373

60. H.-X. Zhou, From induced fit to conformational selection: a continuum of binding mechanism controlled by the timescale of conformational transitions. Biophys. J. **98**, L15–L17 (2010). https://doi.org/10.1016/j.bpj.2009.11.029

61. S. Gu, D.-A. Silva, L. Meng, A. Yue, X. Huang, Quantitatively characterizing the ligand binding mechanisms of choline binding protein using Markov state model analysis. PLoS Comput. Biol. **10**, e1003767 (2014). https://doi.org/10.1371/journal.pcbi.1003767

62. M.S. Formaneck, L. Ma, Q. Cui, Reconciling the "old" and "new" views of protein allostery: a molecular simulation study of chemotaxis Y protein (CheY). Proteins **63**, 846–867 (2006). https://doi.org/10.1002/prot.20893

63. G.G. Hammes, Y.-C. Chang, T.G. Oas, Conformational selection or induced fit: a flux description of reaction mechanism. Proc. Natl. Acad. Sci. **106**, 13737–13741 (2009). https://doi.org/10.1073/pnas.0907195106

64. P. Tompa, Intrinsically disordered proteins: a 10-year recap. Trends Biochem. Sci. **37**, 509–516 (2012). https://doi.org/10.1016/j.tibs.2012.08.004

65. V.N. Uversky, Dancing protein clouds: the strange biology and chaotic physics of intrinsically disordered proteins. J. Biol. Chem. **291**, 6681–6688 (2016). https://doi.org/10.1074/jbc.R115.685859

66. H. Xie, S. Vucetic, L.M. Iakoucheva, C.J. Oldfield, A.K. Dunker, Z. Obradovic, V.N. Uversky, Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. J. Proteome Res. **6**, 1917–1932 (2007). https://doi.org/10.1021/pr060394e

67. V.N. Uversky, C.J. Oldfield, A.K. Dunker, Intrinsically disordered proteins in human diseases: introducing the $D^2$ concept. Annu. Rev. Biophys. **37**, 215–246 (2008). https://doi.org/10.1146/annurev.biophys.37.032807.125924

68. V.N. Uversky, Intrinsic disorder-based protein interactions and their modulators. Curr. Pharm. Des. **19**, 4191–4213 (2013). https://doi.org/10.2174/1381612811319230005

69. S.J. Metallo, Intrinsically disordered proteins are potential drug targets. Curr. Opin. Chem. Biol. **14**, 481–488 (2010). https://doi.org/10.1016/j.cbpa.2010.06.169

70. L. Mollica, L.M. Bessa, X. Hanoulle, M.R. Jensen, M. Blackledge, R. Schneider, Binding mechanisms of intrinsically disordered proteins: theory, simulation, and experiment. Front. Mol. Biosci. **3**, 52 (2016). https://doi.org/10.3389/fmolb.2016.00052

71. T. Kiefhaber, A. Bachmann, K.S. Jensen, Dynamics and mechanisms of coupled protein folding and binding reactions. Curr. Opin. Struct. Biol. **22**, 21–29 (2012). https://doi.org/10.1016/j.sbi.2011.09.010

72. J.C. Ezerski, P. Zhang, N.C. Jennings, M.N. Waxham, M.S. Cheung, Molecular dynamics ensemble refinement of intrinsically disordered peptides according to deconvoluted spectra from circular dichroism. Biophys. J. **118**, 1665–1678 (2020). https://doi.org/10.1016/j.bpj.2020.02.015

73. G. Pérez-Hernández, F. Paul, T. Giorgino, G. de Fabritiis, F. Noé, Identification of slow molecular order parameters for Markov model construction. J. Chem. Phys. **139**, 015102 (2013). https://doi.org/10.1063/1.4811489

74. A.L. Hopkins, C.R. Groom, The druggable genome. Nat. Rev. Drug Discov. **1**, 727–730 (2002). https://doi.org/10.1038/nrd892

75. M.R. Arkin, M. Randal, W.L. DeLano, J. Hyde, T.N. Luong, J.D. Oslob, D.R. Raphael, L. Taylor, J. Wang, R.S. McDowell, J.A. Wells, A.C. Braisted, Binding of small molecules to an adaptive protein-protein interface. Proc. Natl. Acad. Sci. U S A **100**, 1603–1608 (2003). https://doi.org/10.1073/PNAS.252756299

76. D.F. Ceccarelli, X. Tang, B. Pelletier, S. Orlicky, W. Xie, V. Plantevin, D. Neculai, Y.C. Chou, A. Ogunjimi, A. Al-Hakim, X. Varelas, J. Koszela, G.A. Wasney, M. Vedadi, S. Dhe-Paganon, S. Cox, S. Xu, A. Lopez-Girona, F. Mercurio, J. Wrana, D. Durocher, S. Meloche, D.R. Webb, M. Tyers, F. Sicheri, An allosteric inhibitor of the human Cdc34 ubiquitin-conjugating enzyme. Cell **145**, 1075–1087 (2011). https://doi.org/10.1016/J.CELL.2011.05.039

77. J.A. Hardy, J.A. Wells, Searching for new allosteric sites in enzymes. Curr. Opin. Struct. Biol. **14**, 706–715 (2004). https://doi.org/10.1016/J.SBI.2004.10.009

78. J.R. Horn, B.K. Shoichet, Allosteric inhibition through core disruption. J. Mol. Biol. **336**, 1283–1291 (2004). https://doi.org/10.1016/J.JMB.2003.12.068

79. G.R. Bowman, P.L. Geissler, Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. Proc. Natl. Acad. Sci. U S A **109**, 11681–11686 (2012). https://doi.org/10.1073/PNAS.1209309109/SUPPL_FILE/PNAS.1209309109_SI.PDF

80. D.B. Singh, T. Tripathi, *Frontiers in Protein Structure, Function, and Dynamics* (Springer Nature, Singapore, 2020)

81. R. Shukla, T. Tripathi, Molecular dynamics simulation of protein and protein-ligand complexes, in *Computer-Aided Drug Design*, ed. by D.B. Singh, (Springer Nature, Singapore, 2020), pp. 133–161

82. R. Shukla, T. Tripathi, Molecular dynamics simulation in drug discovery: opportunities and challenges, in *Innovations and Implementations of Drug Discovery Strategies in Rational Drug Design*, ed. by S.K. Singh, (Springer Nature, Singapore, 2021), pp. 295–316

83. K. Prince, S. Sasidharan, N. Nag, T. Tripathi, P. Saudagar, Integration of spectroscopic and computational data to analyze protein structure, function, folding, and dynamics, in *Advanced Spectroscopic Methods to Study Biomolecular Structure and Dynamics*, ed. by P. Saudagar, T. Tripathi, (Academic Press, San Diego, 2023), pp. 483–502

84. T. Tripathi, V.K. Dubey, *Advances in Protein Molecular and Structural Biology Methods*, 1st edn. (Academic Press, Cambridge, MA, 2022)

85. P. Saudagar, T. Tripathi, *Advanced Spectroscopic Methods to Study Biomolecular Structure and Dynamics*, 1st edn. (Academic Press, San Diego, 2023)

86. D. Ensign, P. Kasson, V.S. Pande, Heterogeneity even at the speed limit of folding: large-scale molecular dynamics study of a fast-folding variant of the villin headpiece. J. Mol. Biol. **374**(3), 806–816 (2007)