Current Research in Structural Biology 3 (2021) 229–238

Contents lists available at ScienceDirect

Current Research in Structural Biology

journal homepage: www.journals.elsevier.com/current-research-in-structural-biology

Hypervariability of accessible and inaccessible conformational space of proteins☆

Ashraya Ravikumar , Narayanaswamy Srinivasan *

Molecular Biophysics Unit, Indian Institute of Science, Bengaluru, 560012, India

ARTICLE INFO

Keywords:
Protein dynamics
Molecular dynamics simulations
Protein stereochemistry
Bond length and angles
Ramachandran map
(φ,ψ) angle transitions

ABSTRACT

Proteins undergo motions in a range of amplitudes, from domain motions to backbone rotations, leading to changes in (φ,ψ) torsion angles and small-scale bond vibrations and angle bending. Here, we study the extent of variations in (φ,ψ) values in proteins and the effects of bond geometry variations due to vibrational motions in a protein on the accessible, (steric clash-free) (φ,ψ) space. We perform 1-fs timestep unconstrained molecular dynamics simulations on super-high-resolution protein structures. Extent of variations in bond geometry during the simulation is within acceptable ranges of bond lengths and angles. However, the steric clash-free (φ,ψ) space continuously changes as seen in bond geometry-specific (φ,ψ) steric maps at the residue level during simulations. (φ,ψ) regions that have steric clash at one timepoint can become steric clash-free at a different timepoint through minor adjustments to backbone bond lengths and angles. Also instances of (φ,ψ) transitions from the left to right half of the (φ,ψ) map in consecutive snapshots of the trajectory are seen. Although the two quadrants are separated by a steric clash-prone region, corresponding to a high-energy barrier, height of this barrier is lowered by adjusting the bond geometry such that a bridging region of steric clash-free, low-energy (φ,ψ) values is formed. We demonstrate the idea of dynamically varying nature of acceptable and accessible (φ,ψ) steric space in proteins, which has implications for protein folding; proteins could sample (φ,ψ) space which is originally considered to be inaccessible, during folding, through minor adjustments to their backbone bond geometry.

1. Introduction

It is well-known that proteins are dynamic biomolecules and their dynamical features are essential for their function (Frauenfelder et al., 1991; Henzler-Wildman and Kern, 2007; Teague, 2003). They undergo a range of motions, from large-scale, slow dynamics such as domain or subunit movements, torsion angle rotations and sidechain rotamer changes, to small-scale, fast motions such as bond vibrations (Haran and Mazal, 2020; Henzler-Wildman and Kern, 2007). The torsion angles φ and ψ, also called the Ramachandran angles, are defined about the backbone N-Cα and Cα-C bonds. While these torsion angles are often depicted as having fixed values, as seen in static protein structures solved using X-ray crystallography or Cryo Electron Microscopy, they also show variations between different conformations of a protein (Levinson et al., 2006; Oruganty et al., 2013; Ravikumar et al., 2021).

The two-dimensional (φ,ψ) space is demarcated by allowed and disallowed regions, either based on steric considerations (Ramachandran et al., 1963) or observations from large number of protein structures (Chen et al., 2010; Laskowski et al., 1993; Lovell et al., 2003). A recent study using ultra-high-resolution peptide and protein crystal structures has shown that the extent of sterically allowed (φ,ψ) space at a given residue position is drastically altered by subtle, simultaneous deviations in backbone bond lengths and angles (Ravikumar et al., 2019). Such subtle deviations are possible during vibrational motions of a protein, which could in turn affect the extent of accessible (φ,ψ) space at every residue position. This generates the possibility of the steric (φ,ψ) space itself varying with time. This could also explain how large transitions in (φ,ψ) values occurs. Hence, it is imperative that we study bond geometry-specific steric maps with the consciousness of underlying protein dynamics at the level of variations in bond lengths and angles.

In this work, we first show variations in (φ,ψ) angles in ensembles of protein structures, using multiple crystal structures of a given protein and long duration molecular dynamics (MD) simulations. We then use short timestep MD simulations of ultra high-resolution protein structures to understand the extent of variations in bond lengths and angles during dynamics, as well as the effect of such variations on the accessible

☆ This article is dedicated to Prof. N. Srinivasan who passed away on September 3rd, 2021.

* Corresponding author.
E-mail address: ns@iisc.ac.in (N. Srinivasan).

https://doi.org/10.1016/j.crstbi.2021.09.001
Received 26 June 2021; Received in revised form 10 September 2021; Accepted 11 September 2021

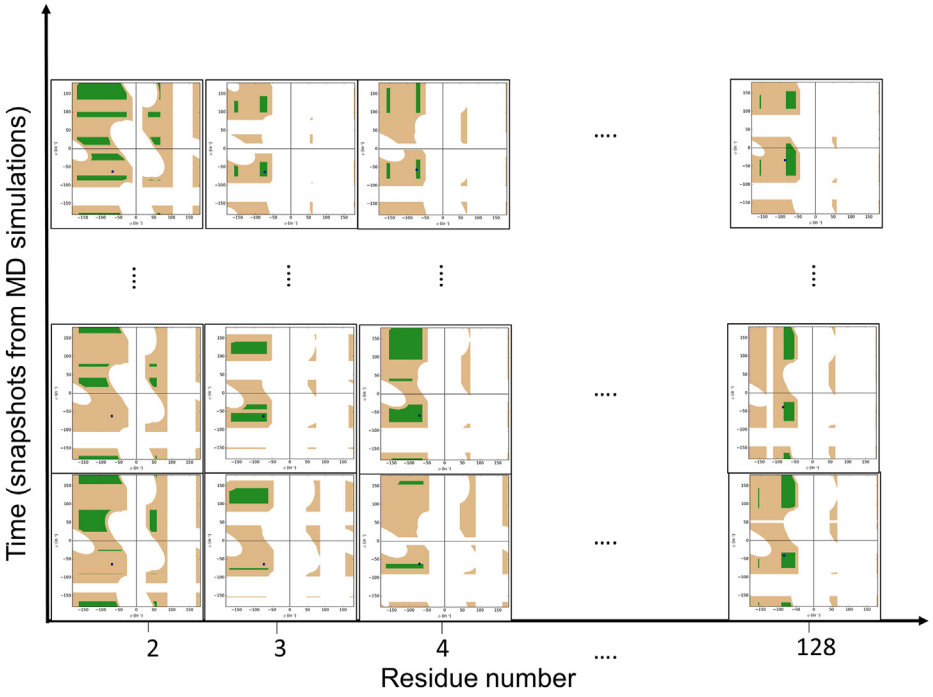

Fig. 1. A schematic on bond-geometry specific maps derived during protein dynamics.
Once the MD simulation of the protein is complete, at every residue position, for each snapshot of the MD trajectory, a bond geometry-specific (φ,ψ) steric map can be obtained, by considering the bond lengths and angles of the residue observed at the given snapshot. Green color indicates fully allowed regions and brown color indicates partially allowed regions.

conformational (φ,ψ) space, as depicted by bond geometry-specific steric map, at a residue position. We do this by generating a bond geometry-specific steric map for each residue at each simulation snapshot. This is shown as a schematic in Fig. 1. These maps are generated by isolating the residue of interest (say *i*) along with atoms (Cα, C, and O) of *i*-1st residue and atoms (N, H and Cα) of *i*+1st residue and varying only the φ and ψ angles while keeping all the other geometric parameters constant (see Methods for further details). We also study how well the steric map variations correspond to the (φ,ψ) potential energy landscape of the residue. We analyze examples which likely explain how large (φ,ψ) transitions during dynamics are accommodated.

It is important to realize that all the calculations performed are extremely sensitive to force field parameters used during MD simulations, since ultimately, these parameters determine the geometry of bond lengths and angles in the protein. However, we have employed multiple force fields and ensured consistency in the results.

MD simulations have been the preferred theoretical method for modeling the dynamics of biomolecules at the atomic level, especially when monitoring dynamics at the atomic level is extremely difficult or impossible using laboratory experiments (Baltzis and Glykos, 2016; Bermudez et al., 2016; Hollingsworth and Dror, 2018; Perilla et al., 2015; Wille et al., 2019). Hence, our observations based on MD simulations can provide realistic insights into the fine level dynamics of proteins.

2. Materials and Methods

2.1. Dataset of high-resolution protein structures

We selected protein structures without DNA/RNA bound, solved by X-ray crystallography at resolution ≤1.2 Å from Protein Data Bank (Berman et al., 2000). This resulted in a total of 3233 protein structures (see the list of structures provided in Supplementary Table 1). The sequences of these structures were clustered at 100% sequence identity using CD-HIT (Huang et al., 2010).

2.2. Ultra-high resolution protein structures chosen for MD simulations

Among all the protein crystal structures solved at a resolution <0.8 Å, we chose those with single chain, with no bound ligands and no modified residues. In order to ensure diversity in the nature and structure of proteins being studied, we chose a representative structure from each of the SCOP classes among these high-resolution structures. This resulted in the following five structures: (i) Hen egg-white Lysozyme (0.65 Å resolution, α+β SCOP class, PDB code 2VB1), (ii) Crambin (0.48 Å resolution, small proteins SCOP class, PDB code 3NIR), (iii) Mating pheromone (0.70 Å resolution, all α SCOP class, PDB code 6E6O), (iv) Non-structural protein 3 from SARS-Cov-2 (0.77 Å resolution, α/β SCOP class, PDB code 7KR0) and (v) Trypsin (0.78 Å resolution, α+β SCOP class, PDB code 5KWM). All these proteins were subject to 1-fs timestep MD simulations. The lysozyme MD trajectory was studied in greater detail.

Table 1
List of proteins that were subject to MD simulations along with timestep, duration and force field of simulation.

PDB Code	Resolution (in Å)	Timestep	Restrained/ Unrestrained	Duration of simulation	Force field used
2VB1	0.65	1 fs	Unrestrained	50 ps	Amber 99sb iLDN
2VB1	0.65	1fs	Unrestrained	50 ps	OPLS-AA
3NIR	0.48	1 fs	Unrestrained	50 ps	Amber 99sb iLDN
6E6O	0.70	1 fs	Unrestrained	50 ps	Amber 99sb iLDN
7KR0	0.77	1 fs	Unrestrained	50 ps	Amber 99sb iLDN
5KWM	0.78	1 fs	Unrestrained	50 ps	Amber 99sb iLDN
2VB1	0.65	2 fs	Restrained	300 ns	Amber 99sb iLDN

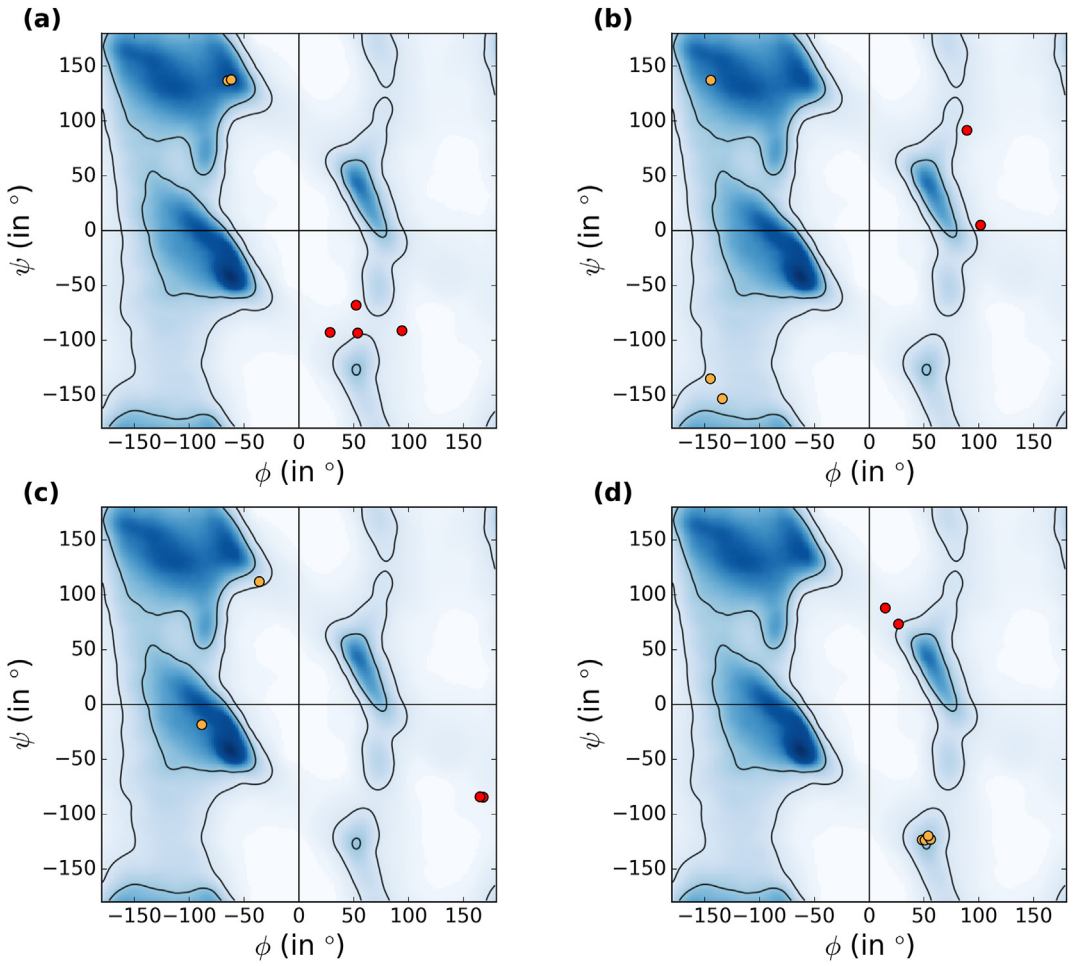

Fig. 2. Variations in (φ,ψ) values of residues in multiple crystal structures of the same protein.
(φ,ψ) values of (a) Arg 5 in small subunit of [NiFe]hydrogenase structures from Desulfovibrio vulgaris, (b) Ala 19 in Ribonuclease A structures of *Bos taurus*, (c) Ser 21 in Ribonuclease A structures of *Bos taurus* and (d) Lys 120 in Fatty acid binding protein structures of *Homo sapiens*. Acceptable (φ,ψ) values (Favored or Allowed according to MolProbity) are shown in yellow and outlier (φ,ψ) values are shown in red on the MolProbity Ramachandran map. Details about these points including the (φ,ψ) values and PDB codes are given in Supplementary Table 3.

2.3. Molecular dynamics simulation protocols

We performed molecular dynamics (MD) simulations on the five protein structures given in Section 2.2. All the systems that were simulated are listed in Table 1. We used GROMACS version 5.1.3 (Abraham et al., 2015) with AMBER99sb-ILDN force field (Lindorff-Larsen et al., 2010) to perform two kinds of simulation: in one of the simulations, the constraints on all bond lengths and angles were removed with a 1-fs timestep and a short duration (50 ps). This was done to analyze the variations in bond lengths and angles during dynamics. To ensure robustness, lysozyme was also simulated with OPLS force field. In the other simulation, the constraints were retained, with 2-fs timestep and performed for a longer duration (300 ns). The purpose of this simulation was to study large (φ,ψ) transitions in residues during dynamics. The 2-fs timestep simulation was performed on the lysozyme system. Further details of the simulations are presented below.

2.3.1. Unconstrained 1-fs timestep simulation

The protein was placed in a dodecahedron box with a minimum distance of 10 Å between protein and edge of the box and solvated with explicit TIP3P water. Counter ions were added to neutralize the excess charge of the system. Energy minimization was performed using steepest descent algorithm with 50,000 steps and a maximum force limit of 1000 kJ/mol/nm. After energy minimization, the system was heated to a temperature of 300 K using Berendsen thermostat at constant volume for 100 ps with position restraints on the protein. This was followed by constant pressure equilibration at 1 bar pressure for 100 ps? Particle Mesh Ewald (PME) method was used for long-range electrostatics, and distance cut-offs for short-range electrostatic and van der Waals interactions were set at 12 Å. We performed a 50-ps production run with 1-fs timestep by removing constraints on all bonds (we set the mdp file parameter "constraints" to "none"). Snapshots were stored at every femtosecond timestep, totaling to 50,000 snapshots.

2.3.2. Unconstrained 1-fs timestep simulation with OPLS force field

Apart from AMBER99sb-ILDN force field, simulation of lysozyme with the same protocols described in Section 2.3.1 was also performed with OPLS-AA force field (Robertson et al., 2015).

2.3.3. Constrained 2-fs timestep simulation of lysozyme

All the protocols until the production run are the same as those described for the unconstrained 1-fs timestep simulation. This production run was carried out at 2-fs timestep for a total duration of 300 ns. All bonds were constrained using LINCS algorithm (Hess et al., 1997) and snapshots were stored at 10 ps intervals, totaling to 30,000 snapshots.

2.4. Measurement of backbone bond lengths and angles during simulation

The length of Cα-C, C–O, C–N, N-Cα and Cα-Cβ bonds and Cα-C-O, Cα-C-N, O–C–N, C–N-Cα, N-Cα-Cβ, Cβ-Cα-C and N-Cα-C angles of all residues

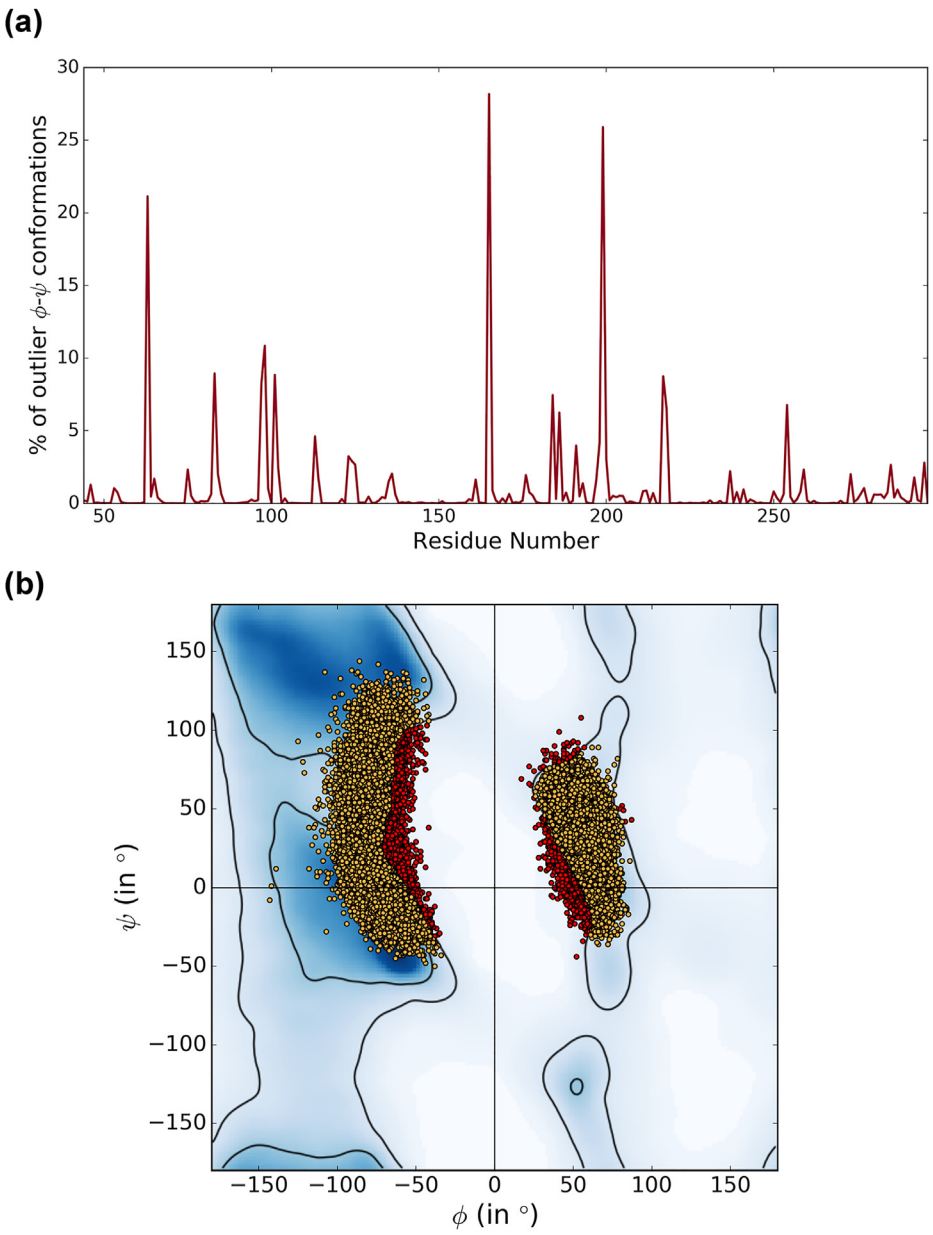

Fig. 3. Variations in (φ,ψ) angles captured during dynamics of mouse protein kinase A (PKA). (a) percent of snapshots out of the 50,000 snapshots of 1μs MD simulation of mouse PKA catalytic domain where the (φ,ψ) angles fall in outlier regions of MolProbity Ramachandran map in each residue. The percent is non-zero for 222 out of 253 residues. (b) Distribution of (φ,ψ) values from MD trajectory for Lys 83. Red points indicate outlier values and yellow points indicate acceptable values. . (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

during the entire 1-fs timestep simulation were measured using VMD (Humphrey et al., 1996).

2.5. Generation of steric maps

To obtain the bond geometry-specific steric map for a given residue at a given snapshot of the simulation, the dipeptide with the residue of interest in the middle was *in silico* isolated from the whole structure in the snapshot and the sidechain was truncated to Cβ atom (a model dipeptide is depicted in Fig. S2(A) in (Ravikumar et al., 2019)). The φ and ψ torsion angles were varied through every possible value at 1° interval between −180° and 180° degrees and the corresponding conformers were generated. The inter-atomic distances within the conformer were measured and compared against the contact criteria used by Ramachandran and co-workers (Supplementary Table 2). For a given combination of φ and ψ, if all the inter-atomic distances are more than the limiting distance, it is considered steric clash-free; if one or more distances is lower than the limiting distance, it has steric clash. In this way, decisions (allowed or disallowed) were obtained for every 1° × 1° grid point for the residues from the snapshots of the five protein structures.

2.6. Energy calculations

To obtain the potential energy value at every point in (φ,ψ) space at a given residue position for a given snapshot, the dipeptide containing the residue of interest was *in silico* isolated from the whole protein and its sidechain was truncated to Cβ atom. Hydrogen atoms were added to the first and third Cα atoms as well as the Cβ atom to make them complete methyl groups and 360 × 360 conformers were generated, each with a unique integral (φ,ψ) value ranging from −180° to 180°. The potential energy of a given (φ,ψ) conformer was calculated using GROMACS version 5.1.3 (Abraham et al., 2015) with AMBER99SB-ILDN force field (Lindorff-Larsen et al., 2010) by placing the isolated dipeptide in vacuum. The total potential energy at the zeroth step of energy minimization is taken as the energy of the given conformer of the dipeptide. The energy terms account for bond length and angle stretching, torsion, Lennard-Jones and electrostatic potential as well as hydrogen bond energy. This way, the potential energy at every integral (φ,ψ) value was

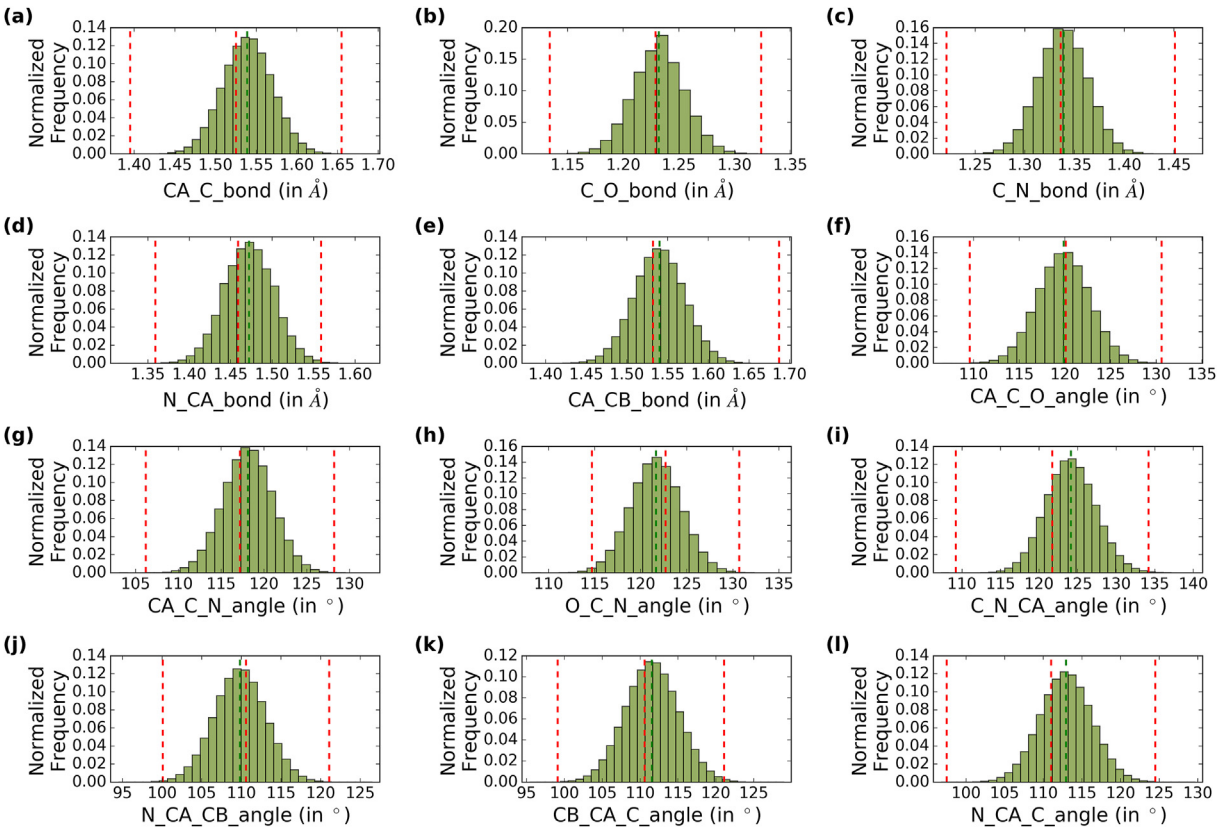

Fig. 4. Distribution of observed backbone bond lengths and angles during MD simulations of lysozme.
(a) to (e) shows the distribution of bond lengths of all residues in the 50,000 snapshots of the 1-fs timestep unconstrained MD simulation. (f) to (l) shows the bond angle distributions of the same. In each subplot, the mean of the distribution is shown as green dashed line. The red dashed line in the middle or close to the middle shows the ideal value of the respective parameter from Engh and Huber analysis and the red dashed lines to the left and right of the distribution, if present, correspond to the extreme limits of the ideal geometry. The y-axis shows the normalized number of observations in each bin of the histogram. The distribution for four other proteins are shown in Supplementary Figs. 1–4.

calculated for residues in specific snapshots of the MD trajectory.

3. Results

3.1. Variations in backbone torsion angles in ensembles of protein structures

The φ and ψ backbone torsion angles in a protein are often viewed as fixed values, as seen in the single, static structure of the protein. However, due to motions experienced by the protein, its (φ,ψ) angles could change. Also, the energy landscape of a given protein is viewed as having multiple minima, causing the existence of conformational ensembles of the protein, with the possibility of variations in (φ,ψ) angles (Frauenfelder et al., 1991; Wei et al., 2016). Hence, a residue which has unfavorable (φ,ψ) angles in one conformation of the protein could have favorable angles in another conformation and vice versa. To demonstrate these changes in (φ,ψ) values, we first analyzed the (φ,ψ) torsion angles of the same residue in multiple crystal structures of the same protein. These structures could represent an ensemble of conformations of the protein, occupying different minima in the energy landscape. We clustered the sequences of all super-high-resolution protein structures (resolution ≤1.2 Å) at 100% sequence identity and identified 871 clusters of proteins with at least two crystal structures in each cluster. We then classified the (φ,ψ) angles of all residues in these structures as Favored, Allowed or Outlier using the MolProbity Ramachandran map (Chen et al., 2010). Residues with Favored or Allowed (φ,ψ) angles were considered as acceptable and remaining were unacceptable (or) outliers. We identified residues from within the cluster of protein structures where, in some of the structures, a given residue has acceptable (φ,ψ) values and in the remaining structures of the protein, the same residue has unacceptable (φ,ψ) values.

There were 245 residues from 110 clusters of proteins which had acceptable and outlier (φ,ψ) angles among multiple structures of the same protein. In 13% of proteins, there is at least one residue which has (φ,ψ) angles falling in acceptable and unacceptable regions of the Ramachandran map. Four examples of such residues are shown in Fig. 2 (also see Supplementary Table 3). Along with change in acceptability of (φ,ψ) angles among the different structures, we can also notice the large change in absolute values of (φ,ψ) angles as well.

In order to characterize the variations in the (φ,ψ) angles during the natural motions of proteins, we utilized a long-duration (1 μs), 2fs timestep molecular dynamics simulation trajectory of mouse Protein Kinase A (PKA) in active state (Kalaivani et al., 2019). 50,000 snapshots taken at 20 ps interval was available. We calculated the (φ,ψ) values at every snapshot for all the residues and classified them using MolProbity as previously described. Fig. 3(a) shows the percent of snapshots at which the residues were having outlier (φ,ψ) values. We see that for 222 out of 253 residues the percent of outlier conformations was non-zero, indicating that protein could adopt varying (φ,ψ) conformations during its motions. Also, in all the 222 residues, (φ,ψ) move from acceptable regions, which are of low energy to outlier regions, which are of apparent high energy. Fig. 3(b) shows the distribution of (φ,ψ) values of Lys 83 in the PKA MD trajectory on the MolProbity Ramachandran map. Apart from certain (φ,ψ) values falling in outlier regions of the map, we also see a huge variation in the (φ,ψ) values being accessed by the residue; there are a large number of observations with $\varphi < 0^{\circ}$ as well as $\varphi > 0^{\circ}$, which are separated by inaccessible (φ,ψ) region.

Prior studies on ultra-high-resolution peptide and protein structures

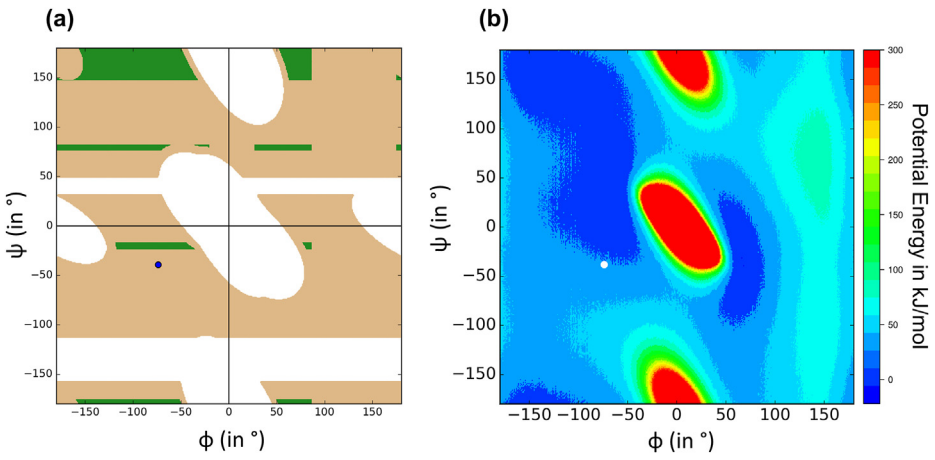

Fig. 5. Highly deviant bond geometry-specific steric map and its potential energy map.
(a) Bond geometry-specific (φ,ψ) steric map of Glu35 in lysozyme at time T = 10,278 fs in the 1-fs timestep unconstrained simulation. The light-brown regions are the partially allowed regions and green regions are the fully allowed regions. The blue point is the (φ,ψ) value observed at Glu35 in the 10278th snapshot. (b) Potential energy distribution (in kJ/mol) of Glu35 at time T = 10,278 fs. Colder to warmer colours indicate increasing energy values. The white point indicates the (φ,ψ) value observed at Glu35 in the 10278th snapshot.

have shown that the (φ,ψ) steric space is dramatically altered by subtle variations in backbone bond lengths and angles (Ravikumar et al., 2019). Such changes in bond geometry could occur due to small-scale vibrational motions in the protein, which could make outlier (φ,ψ) conformations accessible to the residue. In other words, by modulating the bond lengths and angles within acceptable limits, the protein could lower the overall energy of an outlier (φ,ψ) conformation, thereby making it accessible. The MD trajectory of the mouse PKA could not be utilized to test this hypothesis since it was performed at 2 fs timestep, whereas studying bond vibrations requires simulations performed at 1 fs timescale or lower. Hence we performed MD simulations on five ultra-high-resolution (<0.8 Å) protein crystal structures at 1 fs timestep with constraints on bond lengths and angles removed so that changes in bond lengths and angles could be sampled (see Materials and Methods for details on the five proteins and the MD protocols used). From these MD trajectories we generated the bond geometry-specific (φ,ψ) steric maps for every residue in every snapshot of the simulation (as shown in Fig. 1).

3.2. Extent of variations in bond lengths and angles during MD

Before generating the bond geometry-specific steric maps from the MD trajectory, we first studied the distribution of backbone bond lengths and angles of all residues during the unconstrained 1-fs timestep simulation (see Methods for the list of bond parameters measured). We compared the mean, minimum and maximum values observed during MD to the standard, ideal values obtained from small-molecule peptide structures (Engh and Huber, 2012). The maximum and minimum of the ideal geometry were defined as values which are +5σ and -5σ away from the mean. All the distributions for the MD trajectory of Hen egg-white lysozyme (PDB code: 2VB1) are shown in Fig. 4. The bond length and angle distributions for the remaining four proteins are shown in Supplementary Figs. 1–4. In almost all the bond parameters, the mean value observed from dynamics is very close to the mean values from static structures. Also, there are very few cases where the bond parameter values go beyond the maximum or minimum of ideal geometry range. It is to be noted that, despite the bond parameter distributions being obtained from dynamics simulations with no constraints on the bond lengths and angles, they are highly consistent with Engh & Huber values, which are obtained purely from static small molecule structures (Engh and Huber, 2012). This validates the stereochemical accuracy of the unconstrained MD simulations. In order to ensure the robustness of these observations against the influence of force field used, we also simulated the lysozyme structure using OPLS-all-atom force field using the same MD protocol (Robertson et al., 2015). The bond geometry distribution is shown in Supplementary Fig. 5. Here too we see for most of the parameters, the mean values from the simulation is very close to ideal values despite the removal of constraints.

3.3. Accessible (φ,ψ) space at a given residue position is drastically altered during dynamics

In the previous section, we have shown that there are variations in backbone bond parameters within acceptable limits of ideal peptide geometry during the simulation. Prior studies have also shown that such changes affect the extent of steric clash-free regions in (φ,ψ) steric map for a given residue (Ravikumar et al., 2019). These steric clash-free regions depict the accessible (φ,ψ) space for the residue. Here, we wanted to see the effect of such subtle changes in bond length and angle values on the accessible (φ,ψ) space while the protein is in motion. To illustrate this through an example, we chose Glu35 residue in the lysozyme structure, which is the active site residue. For each of the 50,000 snapshots from the unconstrained 1-fs timestep simulation, we generated the bond geometry-specific steric map for Glu35. Supplementary Video 1 shows the variations that occur in the bond geometry-specific steric maps of Glu35 during MD. The movie is a snippet of 1000 snapshots (frames 10,000 to 11,000) out of the 50,000 snapshots. The corresponding dynamics of the dipeptide containing Glu35 is also shown on the right.

Supplementary video related to this article can be found at https://doi.org/10.1016/j.crstbi.2021.09.001

The accessible (φ,ψ) space, as defined by steric-clash-free (φ,ψ) regions, is continuously changing, as shown in Supplementary Video 1. We note such variations in other proteins as well, as shown in Supplementary Videos 2–5. These correspond to the following residues of interest (residues in active site or evolutionarily conserved residues) in these proteins: Asn14 in Crambin (PDB code 3NIR) (Supplementary Video 2), Ser17 in Mating pheromone (PDB code 6E6O) (Supplementary Video 3), Phe131 in Non-structural protein 3 from SARS-Cov-2 (PDB code 7KR0) (Supplementary Video 4) and His42 in Trypsin (PDB code 5KWM) (Supplementary Video 5). (The variations of steric space in Glu35 of lysozyme, as seen in the simulation performed using the OPLS forcefield is shown in Supplementary Video 6).

Supplementary video related to this article can be found at https://doi.org/10.1016/j.crstbi.2021.09.001

Some of the steric maps also look very different from the classic steric map derived by Ramachandran and co-workers based on the ideal dipeptide dimensions (Ramachandran et al., 1963) (shown in Supplementary Fig. 6(a)). One such example is shown in Fig. 5(a). This steric map corresponds to Glu35 of lysozyme at 10278th femtosecond. There are significantly more (φ,ψ) values that are steric clash-free compared to the classical map. Interestingly, majority of the conformers with positive φ value are also steric clash-free. Also, several (φ,ψ) values close to φ = 0°, which are usually disallowed due to severe clashes between carbonyl oxygen of *i*-1 residue and carbonyl carbon or carbonyl oxygen of *i*th residue, are now allowed. These clashes are avoided by subtle but acceptable adjustments to the backbone bond length and angle values

Table 2
Backbone bond length and angles of Glu35 at time T = 12,078 fs of 1-fs timestep unconstrained MD simulation along with ideal geometry parameters and Z-scores.

Parameter	Observed Value	Ideal:Min:Max	Z-Score of observed value
CA-C	1.572 Å	1.525:1.395:1.655 Å	1.81
C–O	1.237 Å	1.229:1.134:1.324 Å	0.42
C–N	1.379 Å	1.336:1.221:1.451 Å	1.87
N-CA	1.473 Å	1.459:1.359:1.559 Å	0.7
CA-CB	1.606 Å	1.532:1.377:1.687 Å	2.39
CA-C-O	117.4°	120.1:109.6:130.6°	−1.29
CA-C-N	119.1°	117.2:106.2:128.2°	0.86
O–C–N	123.4°	122.7:114.7:130.7°	0.44
C–N-CA	129.3°	121.7:109.2:134.2°	3.04
N-CA-CB	116.4°	110.6:100.1:121.1°	2.76
CB-CA-C	108.7°	110.6:99.1:122.1°	−0.83
N-CA-C	116.4°	111.0:97.5:124.5°	2.00

rendering the (φ,ψ) values allowed. The bond length and bond angle values corresponding to this steric map are listed in Table 2. As can be noted, all the bond lengths and angles concerned are within 5σ from ideal values, beyond which they are usually considered as outliers (Read et al., 2011).

The steric map may be viewed as a crude representation of van der Waals energy. In order to see how the map looks when energetic contributions from bonded interactions such as bond length and angle stretching, and also non-bonded interactions such as electrostatics and hydrogen bonding, are considered, we calculated the potential energy associated with every (φ,ψ) conformer of Glu35 at 10278th femtosecond (see Materials and Methods for details on energy calculation). This energy map is shown in Fig. 5(b). It can be seen that there is a good correspondence between the steric map and energy map. This shows that, apart from the van der Waals repulsion being minimal/absent in the steric clash-free regions of the steric map, the unfavorable energetic contributions from slightly deviant bond lengths and angles are not significant enough in majority of the (φ,ψ) regions.

3.4. *Disallowed (φ,ψ) regions can become allowed at a different time and vice versa*

Another interesting phenomenon that can be seen in Supplementary Videos 1–5 is that some (φ,ψ) conformations which are disallowed due to steric clash at one time point say T1, becomes allowed (or) steric clash-free at a later time point, say T2. One such example is shown in Fig. 6. Fig. 6(a) corresponds to steric map of Glu35 of lysozyme at 10065th fs (say T1) from the unconstrained 1-fs timestep simulation. The (φ,ψ) value of Glu35 at this time point is (−69°,-69°) (shown as a green dot). This conformation is allowed at T1 as it is steric clash-free. At another timepoint T2 = 21,121 fs, the (φ,ψ) value of Glu35 is (−103°,-25°). This

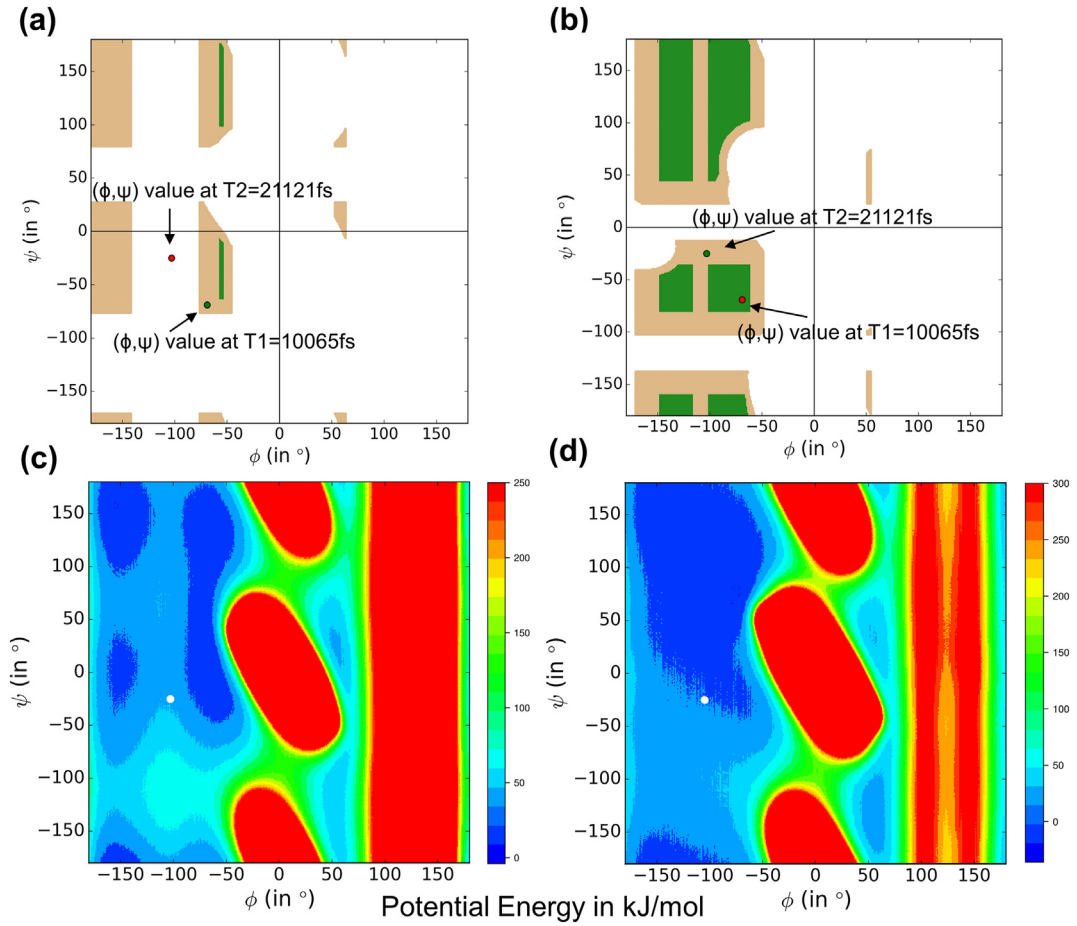

Fig. 6. Changes in accessible and inaccessible (φ,ψ) space during dynamics.
(a) Steric map of Glu35 in lysozyme at time T1 = 10,065 fs from 1-fs timestep unconstrained simulation. The green point shows the (φ,ψ) values of Glu35 observed at T1 = 10,065 fs. The red point shows (φ,ψ) value of the same residue at time T2 = 21,121 fs. This (φ,ψ) value is disallowed at time T1 due to presence of steric clash. (b) Steric map of Glu35 at time T2 = 21,121 fs from the same MD trajectory. The green point shows the (φ,ψ) value of Glu35 observed at time T2. In this steric map, this point is allowed due to absence of steric clash. The red point shows the (φ,ψ) value at time T1. (c) Energy map of Glu35 at time T1 corresponding to (a). The white point is the (φ,ψ) value of Glu35 at time T2. Here, the energy value of the (φ,ψ) point is 34 kJ/mol. (d) Energy map of Glu35 at time T2 corresponding to (b). The white point is the (φ,ψ) value of Glu35 at time T2. Here, the energy value of the (φ,ψ) point is −9 kJ/mol.

Table 3
Backbone bond length and angles of Glu35 at time T1 = 10,065 fs and T2 = 21,121 fs of 1-fs timestep unconstrained MD simulation along with ideal geometry parameters and Z scores.

Parameter	Observed Values at time T1	Observed Values at time T2	Ideal:Min:Max	Z score for values at T1	Z score for values at T2
CA-C	1.601 Å	1.621 Å	1.525:1.395:1.655 Å	2.92	3.69
C–O	1.225 Å	1.228 Å	1.229:1.134:1.324 Å	−0.21	−0.05
C–N	1.321 Å	1.284 Å	1.336:1.221:1.451 Å	−0.65	−2.26
N-CA	1.503 Å	1.441 Å	1.459:1.359:1.559 Å	2.2	−0.9
CA-CB	1.548 Å	1.531 Å	1.532:1.377:1.687 Å	0.52	−0.03
CA-C-O	120.2°	118.8°	120.1:109.6:130.6°	0.05	−0.62
CA-C-N	116.9°	116.8°	117.2:106.2:128.2°	−0.14	−0.18
O–C–N	122.9°	123.6°	122.7:114.7:130.7°	0.13	0.56
C–N-CA	118.9°	121.2°	121.7:109.2:134.2°	−1.12	−0.2
N-CA-CB	111.9°	108.4°	110.6:100.1:121.1°	0.62	−1.05
CB-CA-C	103.8°	112.4°	110.6:99.1:122.1°	−2.96	0.78
N-CA-C	117.5°	107.5°	111.0:97.5:124.5°	2.41	−1.3

(φ,ψ) combination is disallowed at time T1 since the (φ,ψ) point falls in steric clash region of the T1 map (shown as a red point). However, when we plot the steric map of Glu35 at time T2, this point becomes allowed as it is now in steric clash-free region (Fig. 6(b)). Once again, these changes in steric maps are caused due to subtle variations in backbone bond geometry and none of them are outliers (within 5σ from ideal values). The bond length and angle values for Glu35 at time T1 and T2 are listed in Table 3.

The potential energy maps for the two steric maps are shown in Fig. 6(c) and (d). First of all, there is some overall correspondence between the steric maps (in Fig. 6(a) and (b)) and distribution of energy values. For instance, the steric clash region which is seen around φ = −120° to φ = −80° in Fig. 6(a), roughly corresponds to relatively higher energy in Fig. 6(c). The energy value of the (−103°,-25°) (φ,ψ) point, which is the value for Glu35 at time T2, in the energy map of Glu35 at time T1 is 34 kJ/mol (shown as a white dot in Fig. 6(c)). The energy value for the same conformation at time T2 is −9 kJ/mol. Hence the once relatively higher energy of the conformer has been brought down at a different time, by minute adjustments in backbone bond parameters.

3.5. *(φ,ψ) transitions during dynamics are enabled through accompanying changes in accessible (φ,ψ) space*

The steric (φ,ψ) space, as well as the energy landscape of the classical (φ,ψ) space consists of islands of accessible (and often corresponding to low energy) (φ,ψ) values. Both are shown in Supplementary Fig. 6. Transitions in (φ,ψ) values of a given residue during dynamics from one island to another can occur. However, regions with severe steric clashes which usually correspond to high-energy barrier regions must be crossed

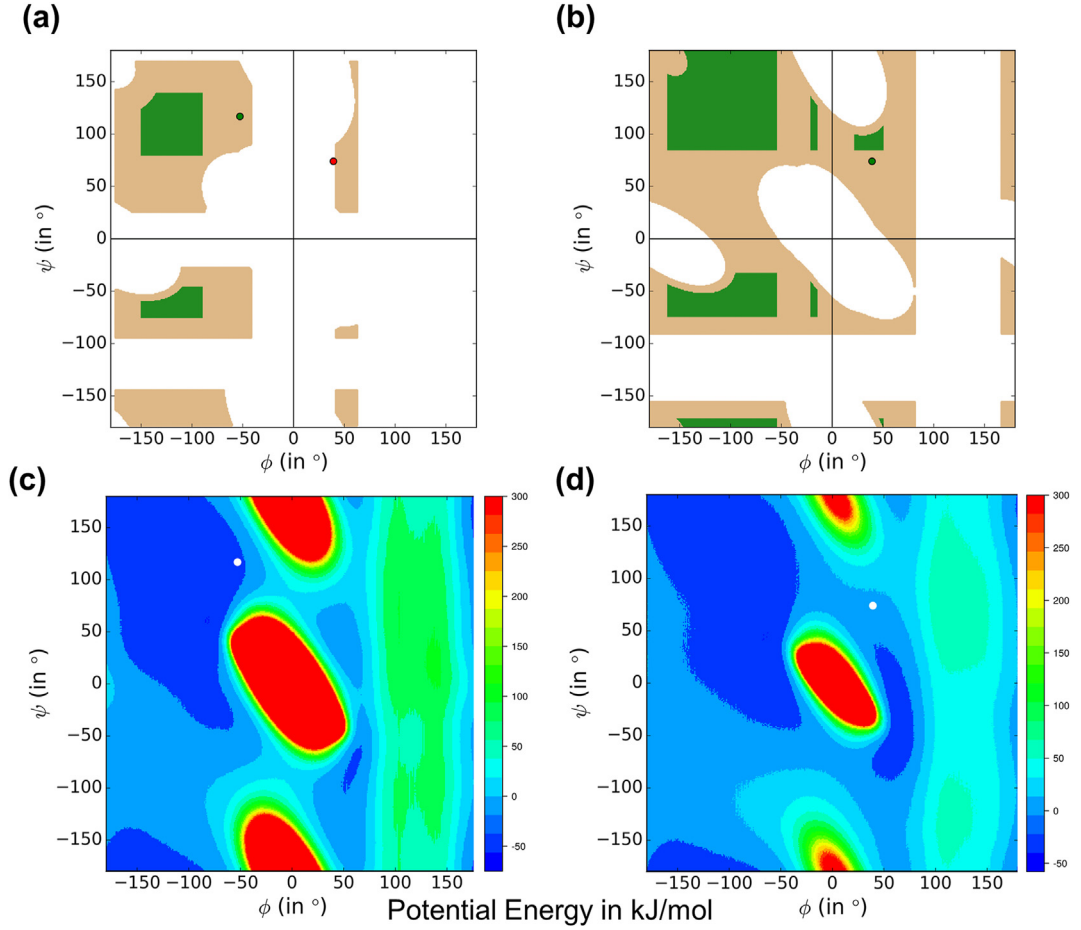

Fig. 7. (φ,ψ) transitions during dynamics enabled by changes in bond geometry-specific steric maps.
(a) Steric map of Ser72 in lysozyme at time T1 = 16.68 ns from 2-fs timestep constrained simulations. The green point shows the observed (φ,ψ) value of Ser72 at T1. The red point shows the observed (φ,ψ) value of Ser72 at time T2 = 16.69 ns. Between T1 and T2 the backbone torsion has transitioned from the left to right quadrant. (b) Steric map of Ser72 at time T2 = 16.69 ns. The green point shows the observed (φ,ψ) value of Ser72 at time T2. This is same as the red point shown in (a). A bridging steric-clash-free region can be seen between the left and right quadrants. (c) Energy distribution of Ser72 at time T1, corresponding to (a). The white point is the observed (φ,ψ) value of Ser72 at time T1. (d) Energy distribution of Ser72 at time T2, corresponding to (b). The white point is the observed (φ,ψ) value of Ser72 at time T2.

over for these transitions to happen. To understand how such transitions are likely accommodated during dynamics, we show an example from the 2-fs timestep simulation of lysozyme (see Materials and Methods for details on MD protocol). Fig. 7(a) shows the steric map of Ser72 at T1 = 16.68 ns of the simulation. The (φ,ψ) value of the residue at T1 is marked as a green circle ((φ,ψ) = (−53°,117°)). At the next snapshot of the simulation, which is taken at time T2 = 16.69 ns (10 ps after T1), the (φ,ψ) value of the residue transitions to the right top quadrant of the (φ,ψ) space, which is marked as a red circle in Fig. 7(a) ((φ,ψ) = (39°,74°)). This value is also disallowed at time T1. Fig. 7(b) shows the steric map of Ser72 at time T2 = 16.69 ns. Here, apart from the (φ,ψ) value becoming allowed (marked as a green circle), a bridge of sterically accessible (φ,ψ) values can be observed between the left and right quadrants of the map. The energy maps of both these steric maps are shown in Fig. 7(c) and (d). Between Fig. 7(c) and (d), a noticeable reduction in energy values in regions around φ = 0° and ψ = 100° can be seen. Hence it is likely that the transition in (φ,ψ) from the left to right quadrant could have been enabled by bringing down the high energy barrier in the bridging region between the two quadrants through minor deviation in bond lengths and angles from standard values. This observation paves way to further investigate the mechanism of such (φ,ψ) jumps using more sophisticated simulation methods and energy calculations.

4. Discussion

In this work we have shown that (φ,ψ) values are not static but vary during protein dynamics. To understand how such changes in (φ,ψ) values are accommodated, and to characterize the effect of subtle bond stretching and bending motions on the accessible (φ,ψ) space, we studied the dynamics of five different protein structures and showed that backbone bond lengths and angles show variations from ideal values within acceptable limits as defined by observations in small molecule peptide structures. We generated bond geometry-specific steric maps for residues in these structures at every timestep of the simulation. Viewing these maps in sequence showed continuous and drastic changes in acceptable (φ,ψ) space with time, irrespective of nature of protein structure or the force field used. Hence the accessible and acceptable (φ,ψ) space themselves are not static for a given residue but vary with time due to vibrational movements in backbone bond lengths and angles. We noted that (φ,ψ) regions which are inaccessible due to steric clashes at one time point can become steric clash-free at a future point and vice versa within a given residue. From a longer duration 2-fs timestep simulation of the lysozyme structure, we showed that large transitions in observed (φ,ψ) values within a residue, which requires crossing over high energy barriers in the (φ,ψ) space, is made possible by the formation of bridging steric-clash-free low energy regions. This is enabled by minor adjustments to the backbone geometry. Hence, through this work, we have demonstrated the idea of dynamically varying nature of acceptable and accessible (φ,ψ) steric space in a protein. Proteins could sample (φ,ψ) space which is originally considered to be inaccessible through minor adjustments to their backbone bond geometry. Indeed, majority of MD simulations, in general, are performed with bond constraints, which means that true bond vibrations cannot be sampled reliably. While in our work, we have performed short-duration, unconstrained simulation and noted plausible lowering of high-energy barriers, more reliable, robust observations can be made using advanced energy calculations, such as quantum mechanical calculations. Ideally, all simulations should be performed with no constraints on bonds and with integration timestep equal to or lower than 1 fs. However, due to the exorbitant need for computational resources and storage for such simulations, this may be practically difficult. But our observations open up interesting questions to explore in future, where one can potentially look at the large-scale effect that apparent minor bond length/angle variations have on the dynamics and transitions between different conformational ensembles of the entire protein.

CRediT authorship contribution statement

Ashraya Ravikumar: Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Narayanaswamy Srinivasan:** Conceptualization, Methodology, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Research in our group is supported by the FIST program sponsored by the Department of Science and Technology (DST) and the Department of Biotechnology (DBT), Government of India in the forms of IISc-DBT partnership program and Bioinformatics and Computational Biology centre. Support from UGC, India – Centre for Advanced Studies and Ministry of Human Resource Development, India is gratefully acknowledged. NS is a J. C. Bose National Fellow. AR is supported by the Bioinformatics and Computational Biology centre grant.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crstbi.2021.09.001.

References

Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., Lindahl, E., 2015. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. Software 1–2, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

Baltzis, A.S., Glykos, N.M., 2016. Characterizing a partially ordered miniprotein through folding molecular dynamics simulations: comparison with the experimental data. Protein Sci. 25 (3), 587–596. https://doi.org/10.1002/pro.2850.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The protein Data Bank. Nucleic Acids Res. 28 (Issue 1), 235–242. https://doi.org/10.1093/nar/28.1.235. Springer.

Bermudez, M., Mortier, J., Rakers, C., Sydow, D., Wolber, G., 2016. More than a look into a crystal ball: protein structure elucidation guided by molecular dynamics simulations. Drug Discov. Today 21 (11), 1799–1805. https://doi.org/10.1016/j.drudis.2016.07.001.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C., 2010. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. Sect. D Biol. Crystallogr. 66 (1), 12–21. https://doi.org/10.1107/S0907444909042073.

Engh, R.A., Huber, R., 2012. Structure quality and target parameters. In: second ed.Arnold, E., Himmel, D.M., Rossmann, M.G. (Eds.), International Tables for Crystallography, pp. 474–484. https://doi.org/10.1107/97809553602060000857.

Frauenfelder, H., Sligar, S.G., Wolynes, P.G., 1991. The energy landscapes and motions of proteins. Science 254 (5038), 1598. https://doi.org/10.1126/science.1749933. LP – 1603.

Haran, G., Mazal, H., 2020. How fast are the motions of tertiary-structure elements in proteins? J. Chem. Phys. 153 (13), 130902. https://doi.org/10.1063/5.0024972.

Henzler-Wildman, K.A., Kern, D., 2007. Dynamic personalities of proteins. Nature 450 (7172), 964–972. https://doi.org/10.1038/nature06522.

Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M., 1997. LINCS: a linear constraint solver for molecular simulations. J. Comput. Chem. 18 (12), 1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

Hollingsworth, S.A., Dror, R.O., 2018. Molecular dynamics simulation for all. Neuron 99 (6), 1129–1143. https://doi.org/10.1016/j.neuron.2018.08.011.

Huang, Y., Niu, B., Gao, Y., Fu, L., Li, W., 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics 26 (5), 680–682. https://doi.org/10.1093/bioinformatics/btq003.

Humphrey, W., Dalke, A., Schulten, K., 1996. VMD: visual molecular dynamics. J. Mol. Graph. 14 (1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

Kalaivani, R., Narwani, T.J., de Brevern, A.G., Srinivasan, N., 2019. Long-range molecular dynamics show that inactive forms of Protein Kinase A are more dynamic than active forms. Protein Sci. 28 (3), 543–560. https://doi.org/10.1002/pro.3556.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 26 (2), 283–291. https://doi.org/10.1107/S0021889892009944.

Levinson, N.M., Kuchment, O., Shen, K., Young, M.A., Koldobskiy, M., Karplus, M., Cole, P.A., Kuriyan, J., 2006. A Src-like inactive conformation in the Abl tyrosine kinase domain. PLoS Biol. 4 (5), 753–767. https://doi.org/10.1371/journal.pbio.0040144.

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O., Shaw, D.E., 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins: Structure, Function and Bioinformatics 78 (8), 1950–1958. https://doi.org/10.1002/prot.22711.

Lovell, S.C., Davis, I.W., Arendall, W.B., de Bakker, P.I.W., Word, J.M., Prisant, M.G., Richardson, J.S., Richardson, D.C., 2003. Structure validation by Cα geometry: φ,ψ and Cβ deviation. Proteins: Structure, Function, and Bioinformatics 50 (3), 437–450. https://doi.org/10.1002/prot.10286.

Oruganty, K., Talathi, N.S., Wood, Z. a, Kannan, N., 2013. Identification of a hidden strain switch provides clues to an ancient structural mechanism in protein kinases. Proc. Natl. Acad. Sci. U. S. A. 110 (3), 924–929. https://doi.org/10.1073/pnas.1207104110.

Perilla, J.R., Goh, B.C., Cassidy, C.K., Liu, B., Bernardi, R.C., Rudack, T., Yu, H., Wu, Z., Schulten, K., 2015. Molecular dynamics simulations of large macromolecular complexes. Curr. Opin. Struct. Biol. 31, 64–74. https://doi.org/10.1016/j.sbi.2015.03.007.

Ramachandran, G.N., Ramakrishnan, C., Sasisekharan, V., 1963. Stereochemistry of polypeptide chain configurations. J. Mol. Biol. 7 (1), 95–99. https://doi.org/10.1016/S0022-2836(63)80023-6.

Ravikumar, A., de Brevern, A.G., Srinivasan, N., 2021. Conformational strain indicated by ramachandran angles for the protein backbone is only weakly related to the flexibility. J. Phys. Chem. B 125 (10), 2597–2606. https://doi.org/10.1021/acs.jpcb.1c00168.

Ravikumar, A., Ramakrishnan, C., Srinivasan, N., 2019. Stereochemical assessment of (ϕ,ψ) outliers in protein structures using bond geometry-specific ramachandran steric-maps. Structure 27 (12), 1875–1884. https://doi.org/10.1016/j.str.2019.09.009.

Read, R.J., Adams, P.D., Arendall, W.B., Brunger, A.T., Emsley, P., Joosten, R.P., Kleywegt, G.J., Krissinel, E.B., Lütteke, T., Otwinowski, Z., Perrakis, A., Richardson, J.S., Sheffler, W.H., Smith, J.L., Tickle, I.J., Vriend, G., Zwart, P.H., 2011. A new generation of crystallographic validation tools for the Protein Data Bank. Structure 19 (10), 1395–1412. https://doi.org/10.1016/j.str.2011.08.006.

Robertson, M.J., Tirado-Rives, J., Jorgensen, W.L., 2015. Improved peptide and protein torsional energetics with the OPLS-AA force field. J. Chem. Theor. Comput. 11 (7), 3499–3509. https://doi.org/10.1021/acs.jctc.5b00356.

Teague, S.J., 2003. Implications of protein flexibility for drug discovery. Nat. Rev. Drug Discov. 2 (7), 527–541. https://doi.org/10.1038/nrd1129.

Wei, G., Xi, W., Nussinov, R., Ma, B., 2016. Protein ensembles: how does nature harness thermodynamic fluctuations for life? The diverse functional roles of conformational ensembles in the cell. Chem. Rev. 116 (11), 6516–6551. https://doi.org/10.1021/acs.chemrev.5b00562.

Wille, H., Dorosh, L., Amidian, S., Schmitt-Ulms, G., Stepanova, M., 2019. Combining molecular dynamics simulations and experimental analyses in protein misfolding. Advances in Protein Chemistry and Structural Biology 118, 33–110. https://doi.org/10.1016/bs.apcsb.2019.10.001.