Chapter 5

Molecular Dynamics Simulations of Hydrophobic Moments and Hydrogen Bonds in Intrinsically Disordered Proteins

5.1 Introduction

Many native proteins lack well-defined three dimensional structures and fluctuate between a dynamic ensemble of interconvertible conformations resembling unfolded proteins under physiological conditions [142, 209, 210]. Numerous evidences of structural disorder is prevalent in the functional protein domains, which establish that well-defined three dimensional structures is not a necessary prerequisite for the functional specificity of these proteins [211]. These proteins are the intrinsically disordered (ID) proteins predominantly characterized by the absence of suitable tertiary interactions. Intrinsically disordered proteins can be either extended like a random coil or compact like a molten globule [37]. The latter type of disorder includes conformational fluctuations of the secondary structures. Disordered segments in a protein may range from a...
few amino acid residues to a part of the domain or even an entire domain of a protein. Completely disordered proteins are also found in nature and their structural plasticity has significant functional implications. Important biological functions of intrinsically disordered proteins include regulation of transcription and translation, cellular signal transduction, protein phosphorylation and cell cycle control [164–168, 212]. Disordered regions also provide sites for chromosomal translocation which are responsible for neurodegenerative diseases and cancer [145, 171, 173, 174, 213, 214]. The absence of a well-defined structure in disordered proteins is encoded in the specific features of their constituent amino acid sequences. Statistical analysis reveals that the amino acid sequences of the ID proteins or domains are considerably different from those of the ordered ones with respect to sequence composition, flexibility, hydrophobicity, charge, secondary structure propensity, coordination number and other intrinsic attributes [156, 215]. The amino acid compositions in the disordered regions are characterized by low occurrences of bulky hydrophobic amino acids and a high preponderance of polar and charged amino acids. The weighted contribution of these input parameters are applied in various disorder predictor algorithms to distinguish the disordered local sequence region from the ordered one [149–151]. A widely used disorder prediction method is based on the distribution of the mean net charge of a protein as a function of its mean net hydrophobicity [185–187]. A combination of low mean net hydrophobicity and high mean net charge is an important prerequisite for the non-compact disordered structures in a protein [188–190]. A comparison with the globular proteins reveals that the ID proteins occupy an unique region of the charge-hydrophobicity space, distinctly separated from the compact globular proteins by a linear boundary.

The high conformational flexibility of the ID proteins enable them to use the disordered regions for the molecular recognition of a specific target or to switch functions by adopting different conformations upon binding respective targets [209, 216, 217]. The mechanism by which the disordered proteins recognize or bind with their respective targets is not well understood [218, 219]. Since the ID proteins exist in a dynamic
conformational ensemble, an atomic-level characterization of their crystal structures is quite challenging. NMR spectroscopy is the primary technique for the structural characterization of this class of proteins. The fast conformational fluctuations of disordered proteins may be captured by NMR by averaging over the entire ensemble of conformations sampled by the protein [183]. Hence unlike globular proteins, the determination of a single conformation is not sufficient to characterize the disordered state. Thus molecular dynamics simulations may be employed to determine the entire conformational ensemble of ID proteins, as a complement to the NMR experiments, for understanding the dynamic structural changes and its impact on the protein function. The molecular dynamics simulations are performed at nanosecond time scale as it is well-established that proteins undergo fast motion in picoseconds, though little is known about longer time scales [220].

This work aims to characterize the effect of different structural features of ID proteins, such as the persistence length, hydrophobic moments, accessible surface area, intramolecular hydrogen bonds, packing density and contact density of $\text{C-\alpha}$ atoms on its dynamic conformational ensemble in order to unravel the mechanism of their respective interactions with specific targets. Persistence length is used as a measure of residue flexibility of the disordered proteins. Disordered regions in a protein crystal have higher average flexibility index (B-factor), lower persistence lengths, higher average absolute net charge and higher average hydrophilicity. Missing electron density in X-ray crystallographic structures is often correlated with the disordered regions in a protein. The absence of interpretable electron density occur due to the increased mobility of atoms in these regions leading to incoherent X-ray scattering, thus rendering the atoms invisible. The second order ellipsoidal hydrophobic moment characterizes the spatial distribution of the hydrophobic residues from the core to the surface of the protein, which is markedly different from that of globular proteins with a defined hydrophobic core. The time dependent hydrogen bond autocorrelation function provides the dynamics of intramolecular hydrogen bonds by considering the hydrogen bonds at any instant of time with respect to the initial time. The first order hydrophobic
moment provides a measure of the degree of hydrophobic imbalance involved in specific protein-protein interactions and identifies regions of functional relevance when correlated with the accessible surface area of proteins.

The chapter is organized as follows: Section 5.2 outlines the method to calculate the above mentioned physical properties and molecular dynamics simulations. Section 5.3 comprises of the results accompanied by discussions. Section 5.4 briefly summarizes the main conclusions of this work.

## 5.2 Methods

The disordered proteins in this study are classified into two groups: (a) proteins which lack well-defined secondary structures and typically behave as a random coils (b) compact proteins, where the well-defined secondary structures co-exist with highly flexible disordered regions [160–162]. Representative candidates from both groups are included in the present study. A metallothionine protein of 61 residues, 4MT2 and a 74 residue neurotoxin, 2ABX, are selected from group (a), while a 105 residue ribonucleoprotein, 1HQ1 and a 120 residue scaffolding protein, 1CD3B, represent group (b) proteins [221–224]. The three dimensional structures of these proteins are obtained from the Protein Data Bank (PDB) [120]. The coordinates of the disordered regions of group (b) proteins are missing from the PDB files [147]. 1HQ1 is a RNA binding protein, while 1CD3B is complexed with single stranded DNA and other scaffolding proteins. The initial template structure of the uncomplexed protein is generated by the following procedure [225]. The method involves the creation of a 3D structure in PyMOL with defined secondary structures as found in the complexed protein [226]. The energy of this structure is minimized and the structure is made compact by simulating in vacuum at 300 K using AMBER 9.0. The radius of gyration ($R_G$) of the ordered part of the resultant structure should be constant and similar to that of the complexed protein. This yields the initial structure for simulation in explicit solvent with TIP3P water model. The initial structure for the group (a)
proteins is obtained from PDB as the position coordinates of all residues are well-defined.

The molecular dynamics simulations for all these proteins was performed by using AMBER 9.0 package with TIP3P water model as the explicit solvent [130]. The missing hydrogen atoms were added with LEAP subroutine. Each protein structure was solvated in the cubic box with TIP3P water model. Charge of the system was neutralized by adding the sodium or chloride ions respectively depending upon the charge of ID proteins. The leaprc.ff99SB force field was used with the periodic boundary conditions. This force field presents a careful reparameterization of the backbone torsion terms in ff99 and achieves a better balance of four basic secondary structure elements [227]. This force field also show the best agreement with experimental data [228]. Each system was energy minimized twice, the solvent is energy minimized first by keeping the protein constrained followed by minimizing the energy of the whole system. The solvated protein is simulated initially at low temperature of 100 K and the temperature is gradually raised upto 300 K for 10 ps at constant volume. Another equilibration for 100 ps is performed at constant temperature 300 K and constant pressure of 1 bar. Constant temperature is maintained through a weak coupling to Berendsen temperature bath with coupling constant of 2 ps while constant pressure is maintained through weak coupling to isotropic pressure bath with a coupling constant of 1 ps [229]. The last conformation after the equilibration is used to start the extended simulation at constant temperature and pressure for 25 ns and the coordinates are saved after every 2 ps for analysis.

Persistence length \( (l_p) \) is a mechanical property which directly measures the intrinsic elasticity/flexibility of polypeptides/nucleic acids. It corresponds to the typical correlation length scale of thermal fluctuations and modulates the interactions between different proteins or proteins and nucleic acids. Persistence length may be measured by various experimental techniques like electron microscopy, light scattering, linear flow birefringence and dichroism [18,66]. Theoretically the worm-like chain model (WLC) [16] may be used to evaluate the elastic and statistical properties of
both proteins and nucleic acids especially those of double stranded DNA. Two input parameters are required for the theory: (i) contour length \((L)\), (ii) persistence length \((l_p)\). The persistence length \((l_p)\) of the ID proteins may be estimated from the probability distribution function of the end-to-end distance \((R_e)\) of a protein.

The distribution function for the end-to-end distance, \(P(R_e)\), of the WLC model is given by a mean-field approximation [67]

\[
P(R_e) = 4\pi R_e^2 N_c \frac{1}{(1 - (R_e/L)^2)^2} \exp \left( \frac{-3L}{4l_p(1 - (R_e/L)^2)} \right) \tag{5.1}
\]

where \(N_c\) is the normalization constant, \(L\) and \(l_p\) are the contour length and persistence length of the chain respectively. \(P(R_e)\) may be experimentally determined from single molecule fluorescence resonance energy transfer (FRET) measurements. For small \(l_p\), \(P(R_e)\) is Gaussian, while for large \(l_p\), \(P(R_e)\) approaches the rigid rod limit and exhibits a non-Gaussian behavior. For molecules of large contour length(\(L)\), the end-to-end distance may be approximated by \(\langle R_e^2 \rangle \approx \langle R_G^2 \rangle \approx Ll_p\). The inter-monomer distance distribution function \(P(r)\) for large \(r\) i.e. \(r >> R_G\) follows an exponential decay given by [67]

\[
P(r) = \beta \exp \left( \frac{-1}{1 - x^2} \right) \tag{5.2}
\]

where \(\beta\) is an arbitrary constant and \(x = l_p r / R_G^2\). The persistence length \(l_p\) may be calculated from the fit of \(P(r)\) to the above equation.

The ellipsoidal representation of the overall shape of a protein may be used to characterize the hydrophobicity profile of ID proteins. The center of mass is chosen as the origin with the principal axes denoted by three components of the radius of gyration, \(g_1, g_2\) and \(g_3\). These three components of \(R_G\) define an ellipsoidal representation given by

\[
x^2 + g_2'y^2 + g_3'z^2 = d^2 \tag{5.3}
\]

where \(g_2' = g_2/g_1, g_3' = g_3/g_1\), \(x, y\) and \(z\) coordinates are expressed in terms of the principle geometric axis and \(d\) is the major principal axis of the ellipsoid, which
is considered as the generalized ellipsoidal radius. Concentric ellipsoids generated with increasing values of $d$ encompass the amino acids on/within the surface of the protein. The hydrophobicity distribution of the amino acid residues may be obtained from the hydrophobicity values of the Eisenberg hydrophobicity consensus scale provided in Table 1 [33]. This distribution is shifted to obtain zero net hydrophobicity and normalized to yield an unit value of the standard deviation. This shift and normalization eliminates the zeroth order moment from the distribution, which removes the dependence of the second order hydrophobic moment on the differences in net protein hydrophobicity. This scaling enables the comparison of the hydrophobicity profiles of different proteins.

The first order hydrophobic moment [32, 33, 192] is a measure of the degree and direction of the linear hydrophobic imbalance about the mean value of hydrophobicity calculated with respect to the protein centroid as the origin of moment expansion and is given by

$$H_1 = \sum_i h_i (r_i - r_c)$$  \hspace{1cm} (5.4)

where $r_i$ is the vector to the centroid of the $i^{th}$ amino acid residue with hydrophobicity consensus value $h_i$ and $r_c$ is the geometric center of protein calculated by

$$r_c = 1/n \sum_i r_i$$  \hspace{1cm} (5.5)

with $n$ being the total number of amino acid residues present in the protein.

The ellipsoidal coordinates are mapped onto a sphere with radius equal to the major principal axis such that all residues in the ellipsoid contribute equally to the magnitude of the moment irrespective of its distance from the geometric center. Thus the Eqn. 5.4 may be recasted as

$$H_1 = 1/n \sum_i h_i (x_i \hat{i} + \sqrt{g_{33}} y_j \hat{j} + \sqrt{g_{33}} z_k \hat{k})$$  \hspace{1cm} (5.6)

The value of the first order hydrophobic moment may be useful for predicting
5.2 Methods

the spatial organization of residue hydrophobicity for a particular tertiary protein structure. Such measures may estimate the cell surface-binding affinity of the protein, which identifies regions of functional interest [192].

The spatial profile of the amino acid hydrophobicity from the protein interior to exterior for disordered proteins is determined by the second order ellipsoidal hydrophobic moment. For residues located within the ellipsoidal surface, the second order hydrophobic moment per residue measured in terms of the generalized ellipsoidal radius $d$, is given by

$$H_2(d) = \frac{1}{n_d} \sum_{i \leq d} h_i' d_i^2$$

$$= \frac{1}{n_d} \sum_{i \leq d} h_i' \left( x_i^2 + g_2' y_i^2 + g_3' z_i^2 \right)$$

(5.7)

where $x_i$, $y_i$, $z_i$ represent the position coordinates of the $i^{th}$ residue. The values of $H_2(d)$ are calculated as a function of various values of $d$ for each protein.

The accessible surface area of proteins (ASAP) was calculated by using DSSP.

Molecular dynamics simulations quantitatively characterize the dynamics of hydrogen bonds in ID proteins in terms of different hydrogen bond time correlation functions as proposed initially by Stillinger [230] and developed later by Luzar and Chandler [231]. Hydrogen bond between an acceptor and donor pair in the ID protein is defined in terms of the geometric constraint determined by the maximum and minimum distance between them. According to this criterion, the maximum and minimum distance between the acceptor and donor should lie between 2.4 to 3.5 Å, while the distance between the hydrogen and the acceptor should be in the range of 1.5 – 2.5 Å [129]. The mean bond lifetime of the hydrogen bonds is estimated by calculating the time dependent autocorrelation function which reflect the existence/nonexistence of hydrogen bonds between all possible donor-acceptor residue pairs [232]. Such autocorrelation functions may be defined as

$$C_x(t) = \frac{\sum_{i,j} S_{ij}(t_0)S_{ij}(t_0 + t)}{\sum_{i,j} S_{ij}(t_0)}$$

(5.8)
where $S_{ij}(t)$ is a quantity which measures the existence of hydrogen bond at time $t$. When hydrogen bond is present between the $i^{th}$ and $j^{th}$ residues at time $t$, then the value of $S_{ij}(t)$ is unity and zero if the hydrogen bond is absent. The set of values $S_{ij}$ for all donor-acceptor pairs specifies all hydrogen bonds present at any time instant $t$. The value of the autocorrelation function is unity at the initial time $t_0$. Two different types of autocorrelation functions, continuous ($x = C$) and intermittent ($x = I$), are calculated in this work. The continuous autocorrelation function evaluates the complete duration of the hydrogen bond until its first breakage and does not consider the same bond, even if it is subsequently reformed at a later time instant. The intermittent correlation function accounts only for those hydrogen bonds which were present at the initial time $t_0$ ignoring all new bonds formed subsequently in the duration of simulation.

Packing in tertiary protein structures is an important determinant of their shape, stability, function and flexibility. The packing of the ID proteins is calculated by using Voronoia, a database for the analysis of atomic-scale packing density, based on an improved Voronoi Cell algorithm [119]. It uses hyperboloid interfaces instead of the planar ones, to access the atomic volumes and has a better resolution of the solvent-accessible and inaccessible regions of the residues in the protein.

The contact density of the proteins is also calculated to measure the extent of change compared to the uncomplexed proteins. The contact density is determined by the number of non hydrogen atoms surrounding the C-$\alpha$ atoms within the mean cutoff radius $R_c$. For ID proteins, the value of $R_c$ is chosen to be 7.3 Å as it marginally deflects from the mean value [131]. The lower cutoff is fixed at 3.5 Å to exclude the covalent neighbours. For globular proteins, the mean contact density is 67.5.
5.3 Results and Discussions

5.3.1 Persistence Length

Fig. 5.1(a) displays the distribution of persistence length, $l_p$, as a function of simulation time. For 4MT2, the range of fluctuation of $l_p$ varies from 3.21 to 4.46 Å. At initial times, the value of $l_p$ is 3.7 Å, which gradually increases to 4 Å at 1.5 ns with large fluctuations in $l_p$ values. This may be attributed to the spatial inhomogeneities in the packing density. From 1.5 to 20 ns, the fluctuations of the $l_p$ values decrease considerably and after 20 ns, the value of $l_p$ remains almost constant. For 2ABX the plot shows that the value of $l_p$ varies from 3.7 to 4.15 Å with almost similar fluctuations, while $l_p$ increases sharply from 4.3 to 6.7 Å for 1HQ1 with lower average fluctuations compared to other ID proteins. The magnitude of $l_p$ for 1CD3 ranges between 5.7 to 6.2 Å in the entire simulation duration, while the fluctuations are more pronounced upto 11 ns. Large fluctuations in $l_p$ values with time lend support to the fact that these molecules exist in a dynamical ensemble of large interconvertible conformations and undergo fast picosecond-timescale motions.

The change in $l_p$ may be directly correlated to the change in size, as measured by

\[ l_p = \left( \frac{3}{8} \frac{\sigma^2}{\rho} \right)^{1/3} \]

where $\sigma$ is the root mean square deviation and $\rho$ is the mass density of the molecule. The change in persistence length is also directly correlated to the change in the size of the molecules, as measured by

\[ \text{size} = \left( \frac{1}{2} \frac{M}{\rho} \right)^{1/3} \]

where $M$ is the mass of the molecule. The change in size is also directly correlated to the change in the persistence length, as measured by

\[ \frac{\Delta l_p}{\Delta \text{size}} = \frac{1}{2} \frac{\rho}{\sigma^2} \]

The change in persistence length is also directly correlated to the change in the size of the molecules, as measured by

\[ \frac{\Delta l_p}{\Delta \text{size}} = \frac{1}{2} \frac{\rho}{\sigma^2} \]

The change in persistence length is also directly correlated to the change in the size of the molecules, as measured by

\[ \frac{\Delta l_p}{\Delta \text{size}} = \frac{1}{2} \frac{\rho}{\sigma^2} \]

The change in persistence length is also directly correlated to the change in the size of the molecules, as measured by

\[ \frac{\Delta l_p}{\Delta \text{size}} = \frac{1}{2} \frac{\rho}{\sigma^2} \]

The change in persistence length is also directly correlated to the change in the size of the molecules, as measured by

\[ \frac{\Delta l_p}{\Delta \text{size}} = \frac{1}{2} \frac{\rho}{\sigma^2} \]

The change in persistence length is also directly correlated to the change in the size of the molecules, as measured by

\[ \frac{\Delta l_p}{\Delta \text{size}} = \frac{1}{2} \frac{\rho}{\sigma^2} \]

The change in persistence length is also directly correlated to the change in the size of the molecules, as measured by

\[ \frac{\Delta l_p}{\Delta \text{size}} = \frac{1}{2} \frac{\rho}{\sigma^2} \]

The change in persistence length is also directly correlated to the change in the size of the molecules, as measured by

\[ \frac{\Delta l_p}{\Delta \text{size}} = \frac{1}{2} \frac{\rho}{\sigma^2} \]

The change in persistence length is also directly correlated to the change in the size of the molecules, as measured by

\[ \frac{\Delta l_p}{\Delta \text{size}} = \frac{1}{2} \frac{\rho}{\sigma^2} \]
5.3 Results and Discussions

by the radius of gyration, $R_G$ of the protein. $l_p$ increases with the size of the protein with a concomitant increase in the accessible surface area. For group (a) protein 4MT2, $l_p$ value changes considerably due to the change in $R_G$ of the protein which may be due to the formation of intermittent secondary structures. For 2ABX, the variation in $l_p$ is small as the size of protein does not change considerably throughout the simulation. The sharp increase in $l_p$ for the group (b) protein, 1HQ1 may be attributed to the decrease in the secondary structure content. Similarly for 1CD3 a prominent change in $l_p$ occurs due to the change in secondary structure content upto 11 ns and subsequently the fluctuations plateau off resulting in almost constant $l_p$ values.

In Fig. 5.1(b), the calculated values of $l_p$ for ID proteins vary sub-linearly with the chain length ($N$) of the protein according to the relation $l_p = (0.74 \pm 0.01)N^{(0.43 \pm 0.02)}$, where the value of the scaling exponent ($\alpha$) is 0.43 $\pm$ 0.02. For the globular proteins the value of $\alpha$ is 0.34 $\pm$ 0.02, implying that $\alpha$ reflects the difference in the type of interactions which stabilize the tertiary structure of globular protein as compared to the disordered ones, at length scales similar to the magnitude of $l_p$. For 4MT2, the time averaged value of the exponent is $\alpha = 0.39$. The average value of the exponent for 2ABX is found to be $\alpha = 0.40$. For 1HQ1, the value of $\alpha$ exhibits considerable fluctuations throughout the entire simulation with an average value of $\alpha = 0.43$, which is quite similar to that of the average value $\alpha = 0.44$ for 1CD3. This difference in the scaling exponent reflects the difference in the intrinsic flexibility of these ID proteins. Thus ID proteins in group (a) are found to be more flexible than those in group (b), due to the absence of well-defined secondary/tertiary structures.

5.3.2 Second Order Hydrophobic Moment

Fig. 5.2 shows the $H_2(d)$ profile of disordered proteins as a function of the distance, $d$ at various simulation times. Only the highlighted time instants depict a considerable change in $H_2(d)$ profile. Different time profiles of $H_2(d)$ for 4MT2 are shown in Fig. 5.2(a), which demonstrates that there is no significant distinction between the
core and the surface with respect to the distribution of hydrophobic residues. The hydrophobicity of 4MT2 changes continuously from hydrophilic to hydrophobic and vice versa from core to the surface throughout the simulation time. This fluctuation in hydrophobicity is due to the presence of larger number of polar and charged residues distributed throughout from core to the surface as compared to the globular proteins, where hydrophobic residues reside predominantly in the core of the protein. The first change in the $H_2(d)$ time profile occurs between 1 and 4 ns where the core changes from hydrophobic to hydrophilic with a smaller ellipsoidal distance, $d_0$. The value of $d_0$ decreases again at 6 ns showing a skewed Gaussian profile of $H_2(d)$. The profile changes significantly at 8 ns even though the value of $d_0$ remains almost constant. At 12 ns, the value of $d_0$ decreases further and the $H_2(d)$ values approach zero, while at 14 ns the profile changes significantly towards the surface with concomitant decrease in $d$ value. The ellipsoidal distance, $d_0$ changes due to the change in the size of the protein throughout the simulation. The $H_2(d)$ profile changes consecutively at 18 ns and 20 ns before it becomes constant till 25 ns with similar values of $d_0$.

The $H_2(d)$ profile for 2ABX exhibits continuous fluctuations throughout the entire simulation trajectory. At 1, 3 and 5 ns the shape of $H_2(d)$ profile with almost similar $d$ values changes significantly towards the surface after $d = 10$ Å. For similar values of $d_0$, $H_2(d)$ profile changes till 12 ns, while at 18 ns the $d_0$ value decreases and the profile changes towards the surface after $d = 12$ Å. Another conspicuous change in the profile is noted at 20 ns, where $H_2(d)$ changes considerably near the core, instead of the surface with $d < 12$ Å, resembling the 1 ns profile. The profile repeatedly changes till 25 ns with similar values of $d_0$. The $d_0$ values for 2ABX shows that the size of the protein does not change appreciably, while the hydrophobicity changes continuously from core to the surface of the protein.

The change in $H_2(d)$ profile for group (b) proteins is not as prominent as compared to group (a) proteins. Fig. 5.2(c) represents the $H_2(d)$ profile for 1HQ1, which depicts that the protein does not have distinct core or surface hydrophobicity. The values of $d_0$ increases continuously during the simulation trajectory till 20 ns, which
5.3 Results and Discussions

Figure 5.2 (a) The second order ellipsoidal hydrophobicity profile for disordered proteins at different time.

gradually plateaus off afterwards. From 1 to 5 ns the hydrophobicity profile changes from hydrophilic to hydrophobic after $d = 12 \, \text{Å}$, while a more pronounced change in hydrophobicity is observed around $d = 18$ to $20 \, \text{Å}$, marked by a significant change in the $d_0$ value from $d_0 = 30 - 40 \, \text{Å}$. The shape of the profile changes again from 10 to 13 ns with an increase in the $d_0$ value from 40 to 45 Å, while the 17 ns profile shows that the shape of the profile changes markedly even though the $d_0$ value remains almost constant. From 14 to 20 Å, the hydrophobicity varies throughout the distance $d$ accompanied by an increase in $d_0$ value. The shape of $H_2(d)$ profile changes even beyond 20 ns inspite of the fact that $d_0$ value remains almost constant.

Fig. 5.2(d) represents the $H_2(d)$ profile at different time snap shots of the simulation run for 1CD3. The value of $d_0$ does not change appreciably till 20 ns, beyond which it increases markedly from 30 to 36 Å in 5 ns. The shape of $H_2(d)$ profile remains constant from 1 to 3 ns till $d = 17.5 \, \text{Å}$, beyond which the change in hydrophobicity alters the hydrophobicity profile. At 9 ns hydrophobicity of the core
shows a significant switch to hydrophilic residues till $d = 17 \text{ Å}$. The hydrophobicity of this region changes successively at 13, 15 and 20 ns with a decrease in the value of $d_0$. The shape of $H_2(d)$ profile remains unchanged from the core to $d = 19 \text{ Å}$ from 20 to 25 ns, beyond which the hydrophobicity changes significantly towards the surface, with an increase in the value of $d_0$ from 29 to 37 Å.

The changes in $H_2(d)$ profile for both groups (a) and (b) proteins shows that there is no significant distinction of hydrophobicity from core to the surface for the ID proteins. This indicates that polar and charged residues are present indiscriminately throughout the protein core and surface providing specific stabilizing interactions for complexation through hydrogen bonds as opposed to the non-specific hydrophobic interactions. The profile changes throughout the simulation time and the change is more pronounced in the group (a) proteins as compared to the group (b) proteins. This implies that ID proteins may provide different binding sites for specific recognition and complexation of different molecules, which characterizes their diverse biological functions.

### 5.3.3 Hydrogen Bonds

Fig. 5.3(a) represents the exponential decay of the continuous autocorrelation function $C_C(t)$ for the intramolecular hydrogen bonds of the ID proteins. $C_C(t)$ denotes the probability that a hydrogen bond formed between a specific pair of donor and acceptor at $t = 0$ is retained upto time $t$. $C_C(t)$ is calculated by averaging over all pairs of intramolecular hydrogen bonds over the entire simulation interval. The exponential fit of $C_C(t)$ may be given by

$$C_C(t) = A \exp(-t/\tau_C) \quad (5.9)$$

where $\tau_C$ is the relaxation time which measures the lifetime of the tagged hydrogen bond. The values of $\tau_C$ for 4MT2, 2ABX, 1HQ1 and 1CD3 are 0.038, 0.042, 0.297 and 0.476 ns respectively. The increased values of the relaxation times in the group (b) proteins imply that the hydrogen bonds are more persistent in group (b) proteins.
5.3 Results and Discussions

Figure 5.3 (a) Distribution of continuous autocorrelation function for hydrogen bonds in different ID proteins. (b) Distribution of intermittent autocorrelation function for hydrogen bonds in different ID proteins.

as compared to those of group (a). Higher values of $\tau_c$ in group (b) proteins may be due to the presence of substantial amount of secondary structures.

5.3.4 First Order Hydrophobic Moment

Fig. 5.3(b) depicts the intermittent time autocorrelation function of the intramolecular hydrogen bonds with respect to time. $C_I(t)$ describes the probability of a particular tagged hydrogen bond to remain intact at time $t$ provided it was intact at $t = 0$. The decay curves behave sluggishly at long times. Such slow time decay cannot be described by a single exponential and may be fitted to a sum of an exponential and a stretched exponential according to the given equation,

$$C_I(t) = A \exp(-t/\tau_{I1}) + B \exp(-(t/\tau_{I2})^{0.02})$$

$$C_I(t) = C_{I1}(t) + C_{I2}(t) \quad (5.10)$$

where $A$ and $B$ are the amplitude factors and $\tau_{I1}$ and $\tau_{I2}$ are the relaxation times for the different regions of $C_I(t)$. The initial decay of $C_I(t)$ at short times may be fitted to an exponential. For 4MT2, 2ABX, 1HQ1 and 1CD3 the values of the relaxation times $\tau_{I1}$ are 2, 1.67, 0.77 and 1.92 ns respectively. It is evident from the values of $\tau_c$
Figure 5.4 The first order hydrophobic moment and accessible surface area of proteins for ID proteins plotted at different times.

and $\tau_{I1}$ that the decay of $C_I(t)$ is slower compared to that of $C_C(t)$, which implies that the hydrogen bonds are not continuously present throughout the simulation time but reform/break to stabilize the protein. The decay of the correlation function $C_I(t)$ is given by a stretched exponential function at longer times corresponding to the relaxation time $\tau_{I2}$. The values of $\tau_{I2}$ for all the proteins is very high i.e. $\tau_{I2} \approx 10^{32}$ compared to the $\tau_{I1}$ values, indicating that $C_I(t)$ values are almost constant at longer times. This implies that the hydrogen bonds, which break and reform only during the initial 6.8, 5.1, 1.7 and 6.2 ns for 4MT2, 2ABX, 1HQ1 and 1CD3 respectively are weak and have short lifetimes. At longer times, the dynamics of hydrogen bonds remains almost unaffected throughout the entire duration of simulation.

Fig. 5.4 shows the distribution of the three components $H_1(x)$, $H_1(y)$ and $H_1(z)$ of the first order hydrophobic moment, $H_1$, with the accessible surface area of proteins (ASAP) as a function of simulation time. $H_1$ depends on the spatial distribution of the amino acid hydrophobicity and provides a measure of degree of imbalance of
the hydrophobic forces involved in protein-protein interactions. The direction of $H_1$ assists in identifying regions of functional interest, while the correlation of ASAP with $H_1$ may be used to predict the cell surface affinity. The direction of $H_1$ is determined by Eqn. 5.6. From Fig. 5.4 it is clear that the magnitude of $H_1(x)$ is almost constant and the change in the magnitude of other two components of $H_1(y)$ and $H_1(z)$ account for the change in direction of $H_1$.

Fig. 5.4(a) illustrates the $H_1$ and ASAP profile of the disordered proteins as a function of the simulation time. For 4MT2 the magnitude of $H_1$ exhibits an initial increase at 1 ns and remains almost constant upto 2.5 ns. The first order moment follows a periodic pattern of decrease and increase till 10 ns. After a maxima at 14 ns, the same trend is continued upto 19 ns and thereafter $H_1$ remains almost constant till 25 ns. Similar behavior is depicted in the ASAP even though after 19 ns, the magnitude of ASAP decreases due to the change in the direction of $H_1$ moment influenced by a change in $H_1(z)$ component.

The $H_1$ moment of 2ABX is portrayed in Fig. 5.4(b). The plot shows two closely spaced maxima, one between 8 to 9 ns and the other at 11.5 ns implying a local imbalance of the hydrophobic interactions due to a significant conformational change. The time evolution of $H_1$ follows a similar periodic behavior as 4MT2 upto 20 ns beyond which it remains almost constant. A similar trend is reflected in the ASAP profile throughout the entire duration of simulation except between 9 – 10 ns where even though the $H_1$ moment magnitude decreases the ASAP value increases due to a change in the direction of $H_1$ moment caused by a change in the value of $H_1(z)$.

Fig. 5.4(c) and (d) shows that the magnitude of $H_1$ moment increases with time throughout the simulation duration for both group (b) proteins. For 1HQ1, the magnitude of ASAP reflects a similar trend as the $H_1$ moment and increases throughout the entire simulation. For 1CD3, there is a slight change in the magnitude of $H_1$ during the first 2 ns, while the ASAP exhibits a marked increase during the same time. The direction of $H_1$ changes in the time interval between 2 – 14 ns although its magnitude remains appreciably constant. The ASAP increases during 2 – 14 ns and
5.3 Results and Discussions

Figure 5.5 Packing density as a function of simulation time.

The average packing density of the protein molecules is an important physical property for evaluating the compactness of the protein structure [20, 139]. For globular proteins, the average packing density of protein interiors is similar to that of a crystalline solid [133], with a value of 0.74. The average packing density of ID proteins

thereafter follows the same increasing trend as that of the $H_1$ moment. The increase in the magnitude of $H_1$ and ASAP reflects the direct correlation of the first moment with the accessible surface area resulting in an increase in cell surface affinity for interactions with other biomolecules. The change in the direction of $H_1$ affects the increase/decrease of ASAP and constitutes a key component for facilitating specific interactions required for molecular recognition.

5.3.5 Packing Density

The average packing density of the protein molecules is an important physical property for evaluating the compactness of the protein structure [20, 139]. For globular proteins, the average packing density of protein interiors is similar to that of a crystalline solid [133], with a value of 0.74. The average packing density of ID proteins
is plotted with respect to time in Fig. 5.5. The average packing density of 4MT2 fluctuates continuously in the intermediate time interval upto 20 ns and attains a constant value of 0.72 at 25 ns. The time-averaged packing density of the protein is 0.73. The packing density distribution of 2ABX fluctuates upto 17 ns and reaches a steady value thereafter except for the slight decrease at 22.5 ns. For 1HQ1 the packing density records a significant continuous increase from 0.72 to 0.74. The packing density of 1CD3 has a constant value of 0.72 at the initial phase of the simulation, fluctuates in the intermediate time interval and reaches a constant value of 0.73 at 25 ns. The average contact density of the proteins increases with time for all the ID proteins showing the ID proteins acquire more compact structure with respect to time. The change in packing is more pronounced in group (b) proteins as compared to group (a) proteins. The ID proteins can be distinguished from globular proteins by solely considering the average packing density as the difference in the average packing density of these proteins is not significant compared to the globular proteins.

*Figure 5.6* Contact density distribution as a function of simulation time.
The contact density of disordered proteins provide a direct measure of the differences in packing density of ID proteins. The C-\(\alpha\) contact density is calculated with a cutoff distance \(R_c = 7.3\ \text{Å}\) for the chosen group (a) and group (b) proteins. The contact density plot, Fig. 5.6(a) for 4MT2, shows that the contact density for all residues changes significantly throughout simulation due to the rapid conformational changes and the average contact density for each residue ranges between 28 to 58. The average contact density of whole protein fluctuates continuously till 19 ns at which the contact density is 40. At 25 ns, the contact density reaches a constant value of 43. The fluctuations in the values of contact density for 2ABX are pronounced for each residue even though its magnitude is smaller compared to 4MT2. The average contact density per residue ranges from 20 to 60 with maximum residues having contact density less than 50. The contact density for different residues in 1HQ1 ranges from 10 – 90 and the average contact density per residue ranges from 28 – 68 with maximum residues having less than a contact density of 55. Similarly for 1CD3 the contact density for each residue ranges from 10 – 90, with an average contact density of 23 – 67 and maximum residues have a contact density less than 50. The average contact density of the proteins of both groups varies between 42 – 45 which is less than the average contact density of globular proteins i.e. 67.5 [131] indicating a comparatively loose packing with reduced compactness.

5.4 Conclusions

This study aims to characterize the different physico-chemical properties like the persistence length, hydrophobic moments, intramolecular hydrogen bonding, average packing density and contact density of group (a) and group (b) ID proteins as a function of time. Group (a) proteins are found to be more flexible compared to the group (b) proteins reflecting the difference in the type of interactions stabilizing their respective tertiary structures. The first order hydrophobic moment determines the hydrophobic imbalance across the entire protein and is a measure of its amphiphilicity.
An increase in the magnitude of first order hydrophobic moment implies deviation from the spherical geometry with the first order hydrophobic moment pointing away from the binding site. The first order hydrophobic moment is directly correlated with the accessible surface area of ID proteins and provides a measure of their binding affinity. For group (a) proteins the fluctuations in the magnitude of the first order hydrophobic moment is related to the changing position of the binding site, while for group (b) proteins it increases consistently with less fluctuations resulting in fewer possibilities for the intermolecular interactions. The hydrogen bond dynamics characterized by the continuous autocorrelation function shows that the hydrogen bonds are more persistent in group (b) proteins as compared to those of group (a). The change in average packing density and contact density of C-α atoms with respect to time implies that the change in packing density is more pronounced in the group (b) proteins as compared to those of group (a), while the average contact density of C-α atoms is higher for the group (b) proteins as compared to those of group (a), but is considerably lower compared to the globular proteins.