THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

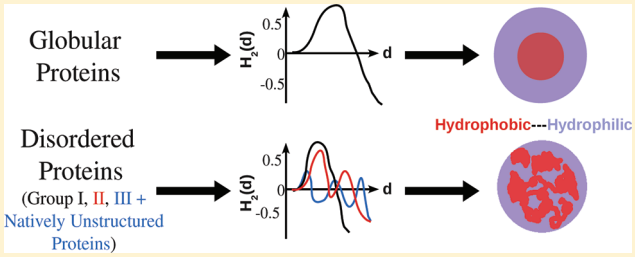
pubs.acs.org/JPCB

Hydrophobic Moments, Shape, and Packing in Disordered Proteins

Nidhi Rawat and Parbati Biswas*

Department of Chemistry, University of Delhi, Delhi-110007, India

ABSTRACT: Disordered proteins play a significant role in many biological processes and provide an attractive target for biophysical studies under physiological conditions. Disordered proteins may be classified as (a) proteins with overall well-defined secondary structures, interspersed with regions of missing residues, or (b) natively unstructured proteins which lack definite secondary structure. The spatial profile of second order hydrophobic moment for disordered proteins depicts the distribution of hydrophobic residues from the interior to the surface of the protein and indicates the lack of a well-formed hydrophobic core unlike that of the globular proteins. This trend is independent of the size or position of the disordered region in the sequence. The hydrophobicity profile of the ordered regions of the disordered proteins differ considerably from that of globular proteins implying the role of disordered parts and the significance of hydrophobic interactions in the folding of proteins. The shape asymmetry of the two classes of disordered proteins is determined by calculating the asphericity and shape parameters, derived from the Cartesian components of radius of gyration tensor. Disordered proteins of group a are more spherical as compared to the natively unstructured proteins (group b), which are more prolate. Both groups of proteins exhibit similar types of side-chain backbone contacts, as that of the globular proteins. While disordered proteins contains few hydrophobic residues natively unstructured proteins are characterized by a residues of low mean hydrophobicity and high mean net charge.



I. INTRODUCTION

Intrinsically disordered proteins (IDP) are highly flexible noncompact proteins or protein domains with minimal or no ordered tertiary and/or secondary structures under physiological conditions, which may be represented by dynamic ensembles of interconvertible conformations.[1−5] In contrast to proteins with well-defined three-dimensional structures, these disordered proteins exhibit noncooperative conformational changes.[6] The conformational behavior of this class of proteins typically resemble the non-native states of globular proteins, which may exist in any of the three different structures represented by molten globule, premolten globule, and the unfolded conformation.[7] These states are structurally distinct with increasing disorder and may be easily characterized by various physicochemical techniques.[8−11] Disordered proteins are widespread among all functional proteins but are mostly prevalent in the eukaryotes.[12−19] Disorder in these proteins may be of two types: (a) Despite highly flexible disordered parts, proteins are compact with well-defined secondary structures, exhibiting the properties of a molten globule[20,21] (b) proteins lack well-defined secondary structures and typically behave as random coils.[22] This group is also referred to as the "natively unstructured proteins".

The existence of a unique three-dimensional structure may not be a necessary prerequisite for the functional specificity of a protein. Indeed, the lack of a stable three-dimensional structure under physiological conditions imparts increased plasticity to the disordered proteins to bind specifically with various targets. This order−disorder transition induced due to target-specific binding regulate various cellular processes like DNA-recognition, transcription and translation, cellular signal transduction, cell cycle control, structural uncoupling of two or more domains by flexible linkers, protein folding inhibition, protein phosphorylation etc.[23−28] These proteins remain unstructured in solution in absence of physiological partners, but folds in presence of suitable partners forming stable complexes that can be appropriately characterized. Biomedical study of disordered proteins is also important as their accumulation may be linked to many neurodegenerative disorders like Alzheimer's disease, Down's syndrome, Parkinson's disease, dementia, etc.,[29−34] while the interactions of disordered proteins with other proteins, nucleic acids, membranes or small molecules provide new drug targets for discovering novel drugs.[35−37]

The functional diversity of the IDP's and their significant role in various debilitating diseases requires a complete characterization of their sequences and conformational properties. Though previous experimental and theoretical studies[38−42] have emphasized different aspects of their sequence-structure-function correlation yet a consistent picture is still lacking. Disorder in protein structures may be either local or global. Locally disordered regions are usually observed in numerous X-ray crystallographic and NMR structures.[43] Examples of global disorder are found in many protein domains, sometimes even the entire protein is unstructured.[44] Examination of the amino acid sequence data reveal that many proteins contain extended

Received: February 20, 2012
Revised: May 11, 2012
Published: May 14, 2012

dx.doi.org/10.1021/jp3016529 | J. Phys. Chem. B 2012, 116, 6326−6335

segments of low sequence complexity and amino acid compositions that are not conducive to the formation of a globular protein fold. The sequences usually contain a low frequency of hydrophobic amino acids and a high proportion of charged residues.[45−47] However, the disordered proteins cannot be distinguished from globular proteins by solely considering the relation between the net charge and the net hydrophobicity.[48−50]

For globular proteins, the spatial distribution of hydrophobic residues distinctly varies from the core to the surface of the protein and the protein hydrophobicity is quantitatively measured by the second order ellipsoidal moment using the hydrophobicity scale of Eisenberg.[51] The second order hydrophobic moment profile of most globular proteins depicts the spatial transition from hydrophobic core to the hydrophilic surface with an invariant hydrophobic ratio of 0.71 ± 0.08, which is the simple ratio of the distance at which the second order moment and zeroth order moment of hydrophobicity vanishes.[52−55]

This article compares various structural features of disordered proteins with those of the globular ones. The distinct differences in shape parameters, hydrophobicity profile, two-body contacts, charge vs net hydrophobicity pattern serves as a benchmark to distinguish different groups of disordered proteins from the globular ones, providing a structural insight into characterize their specific functions and diverse interactions. The shape parameters are calculated from the three-dimensional coordinates of the X-ray crystallographic protein structures available from the Protein Data Bank (PDB) and Database of Protein Disorder (DisProt). Our earlier results establish that the natively unstructured proteins are less compact but more flexible compared to the globular proteins.[56] The hydrophobicity profile of the disordered proteins varies as a function of the length of the protein. The second order ellipsoidal hydrophobic moment for most disordered proteins is found to lie between the values of globular proteins and those of the natively unstructured ones in group b. The paper is organized as follows: Section II outlines the procedure followed to calculate the hydrophobicity, shape parameters and charge. Section III presents the results and discussions. Section IV briefly summarizes the conclusions of our study.

II. METHOD

Both groups of the disordered proteins (a and b) are considered in this study. Proteins in group a have intermittent disordered regions along with the well-defined secondary structures. The supplementary data set in DisProt database gives 2321 monomeric proteins which are characterized by X-ray crystallography with disordered regions of ≥3 residues. The atomic coordinates of these proteins are obtained from PDB.

The shape of any object with arbitrary geometry may be measured by two rotationally invariant quantities, the asphericity parameter (Δ) and the shape parameter (S) which are determined from the inertial tensor[57−59]

$$T_{(\alpha,\beta)} = \frac{1}{2M^2}\sum_{i}^{N}\sum_{j}^{N} m_i m_j (r_{i\alpha} - r_{j\alpha})(r_{i\beta} - r_{j\beta}) \quad (1)$$

where $r_{i\alpha}$ is the αth component of the position of the ith residue and $\alpha, \beta = x, y, z$. The eigen values of the matrix $T_{(\alpha,\beta)}$, denoted by λ_1, λ_2, and λ_3, are the squares of the three principal radii of gyration R_G. The square of the radius of gyration R_G^2 is given by $R_G^2 = trT = \sum_{i=1}^{3}\lambda_i$.

Asphericity parameter characterizes the average deviation of the chain conformation from spherical symmetry. In three dimensions, the extent of asphericity is measured by Δ, defined as

$$\Delta = \frac{3}{2}\frac{\sum_{i=1}^{3}(\lambda_i - \lambda)^2}{(trT)^2} \quad (2)$$

where

$$\lambda = \frac{\sum_{i=1}^{3}\lambda_i}{3} \quad (3)$$

The value of Δ varies between zero (the value for a sphere) and one (the value for a rod).

The shape parameter (S) characterizes the ellipticity of the chain by designating it as either prolate or oblate.

$$S = 27\frac{\prod_{i=1}^{3}(\lambda_i - \lambda)}{(trT)^3} \quad (4)$$

S is a bounded parameter which obey the inequality $-1/4 \le S \le 2$. The value $S = 0$ implies that the molecule is completely symmetric, i.e., spherical. Positive values of S correspond to prolate ellipsoids resembling the shape of a rod, while negative values indicate oblate ellipsoids resembling the shape of a disk. Δ and S are directly calculated from the atomic coordinates of the three-dimensional structures of both groups of disordered proteins obtained from the PDB. No tacit assumptions about the extent of packing and steric overlaps are considered,[60] which include a definite degree of arbitrariness.

We enumerate the number of two-body contacts for understanding the extent of packing in both groups of disordered proteins. The number and nature of two body contacts is determined by classifying the amino acids into nine groups based on their charge and van der Waals volume.[61] All heavy atoms of the side chains of each amino acid are considered for calculating the *ss* contacts. For any two side-chains to be in contact, at least a pair of heavy atoms from the two amino acids should be separated by a distance (d), $d \le 5.2$ Å. Backbones (b) are represented by C_α atoms and *bb* contacts are considered by fixing the distance between two backbones as ≤6.5 Å. The minimum distance between the backbone and side-chain residues for *sb* contacts should be ≤5.5 Å.[58] The *sb* contacts play a dominant role in the side chain packing, thus stabilizing the conformations of globular proteins.[62]

The overall shape of any globular protein may be generalized using the ellipsoidal representation. This shape may be generated from the molecular moments of inertia such as the moments of inertia tensor. The center of mass of the protein is considered as the origin, while the three components of R_G are designated as g_1, g_2 and g_3, with $g_1 > g_2 > g_3$. The ellipsoidal representation generated by the three components of R_G may be written as

$$x^2 + g_2'y^2 + g_3'z^2 = d^2 \quad (5)$$

where $g_2' = g_2/g_1$, $g_3' = g_3/g_1$, x, y, and z coordinates are expressed in the frame of the principle geometric axis and d is the major principal axis of the ellipsoid and may be considered as generalized ellipsoidal radius. Gradual increase of the d values generate concentric ellipsoids which include the amino acid residues lying on/within the surface. The values of the hydrophobicity of the residues is obtained from the Eisenburg hydrophobicity consensus scale provided in Table 1 in ref 52.

dx.doi.org/10.1021/jp3016529 | J. Phys. Chem. B 2012, 116, 6326−6335

The hydrophobicity distribution of amino acids is shifted to obtain zero net hydrophobicity and normalized to yield a standard deviation value of unity. This shifting and normalization helps to eliminate the zeroth order moment from the distribution and consequently removes the dependence of the second-order moment on the differences in net protein hydrophobicity. This scaling provides a basis for the comparison of the hydrophobic moment profiles of different proteins including their respective hydrophobic ratio.

The zeroth order hydrophobic moment (H_0) or the net hydrophobicity of a protein is given by

$$H_0 = \sum_i h_i \tag{6}$$

where h_i is the hydrophobicity of the ith residue in the protein.

The normalized zero order moment of protein within an ellipsoidal surface when value of d is just sufficient to enclose the whole protein is given as

$$H_0(d) = 1/n_d \sum_i h_i' = 1/n_d \sum (h_i - \overline{h})/\langle (h_j - \overline{h})^2 \rangle^{1/2} \tag{7}$$

where (h_i') is the normalized hydrophobicity of residues, $\overline{h}$ is the mean hydrophobicity, $\langle (h_j - \overline{h})^2 \rangle^{1/2}$ is the standard deviation and n_d is the number of residues.

The first order hydrophobic moment[52,53,63] is a measure of the degree and direction of the linear hydrophobic imbalance about the mean value of hydrophobicity calculated with the protein centroid as the origin of moment expansion, which is given by

$$\vec{H}_1 = \sum_i h_i(\vec{r}_i - \vec{r}_c) \tag{8}$$

where, $\vec{r}_i$ is the vector to the centroid of the ith amino acid residue with hydropbobicity consensus value h_i and $\vec{r}_c$ is the geometric center of protein calculated as

$$\vec{r}_c = 1/n \sum_i \vec{r}_i \tag{9}$$

where n is the total number of amino acid residues in the protein. The value of the first order hydrophobic moment yields useful information with respect to the spatial organization of residue hydrophobicity for a given tertiary protein structure. Such measures may be important for estimating the cell surface-binding affinity of the protein, which may identify regions of functional interest.[53]

However, the focus of the present study is to determine the spatial distribution of amino acid hydrophobicity from protein interior to exterior for the disordered proteins, which is determined by the second order ellipsoidal hydrophobic moment and the hydrophobic ratio.

Second order hydrophobic moment per residue among residues located within the ellipsoidal surface measured in terms of the generalized ellipsoidal radius d may be defined by

$$H_2(d) = 1/n_d \sum_{i \le d} h_i' d_i^2 = 1/n_d \sum_{i \le d} h_i'(x_i^2 + g_2' y_i^2 + g_3' z_i^2) \tag{10}$$

where x_i, y_i, z_i represent the position coordinates of the ith residue. The values of $H_2(d)$ are calculated for increasing values of d for each protein.

Hydrophobic ratio gives the ratio of the two distances measured from the protein core at which the second order and zeroth order hydrophobic moments vanish. For globular proteins both moments start from the positive value, initially increasing and subsequently decreasing with increasing values of d. Zeroth order moment $H_0(d)$ is always positive and at the maximum value of d, denoted as d_0, $H_0(d)$ becomes zero. The second order moment becomes zero at an intermediate distance d_2 and is negative for all higher values of $d < d_0$. The hydrophobic ratio is defined as

$$R_H = d_2/d_0 \tag{11}$$

The hydrophobic ratio is relatively invariant for the globular proteins which assists in validating the predicted native and near-native protein structures and in predicting protein folding pathways. R_H cannot be defined for proteins which do not show the expected smooth spatial profile for the second order moment and thus are considered to have a decoy profile. Most disordered proteins exhibit an irregular spatial profile of the second order moment.

A consideration of both factors of mean hydrophobicity and mean net charge demarcates the intrinsically unstructured proteins from the native(folded) ones. The normalized mean hydrophobicity plotted as a function of the absolute mean net charge of the protein show that disordered proteins lie in the region of high charge and low hydrophobicity. The normalized mean hydrophobicity is calculated by dividing the sum of zero order moments by the number of residues, which is determined by using the Eisenburg hydrophobicity consensus values for each amino acid. For a given protein, the net charge (Q) at any pH is evaluated by using[64]

$$Q = \sum Q^- + \sum Q^+ \tag{12}$$

where

$$Q^- = \frac{(-1)}{1 + 10^{-(\mathrm{pH} - \mathrm{p}K_a)}} \tag{13}$$

and

$$Q^+ = \frac{(+1)}{1 + 10^{+(\mathrm{pH} - \mathrm{p}K_a)}} \tag{14}$$

are the fractions of negative and positive charges, respectively. The pK_a values for the side chain groups of amino acids are those commonly reported for the respective amino acids, i.e., for Asp(pK_a = 3.9), Glu(pK_a = 4.35), His(pK_a = 6.5), Try(pK_a = 9.9), Lys (pK_a = 10.35), and Arg(pK_a = 12.5). The pK_a value of C-terminal and N-terminal groups in general vary from protein to protein but the values used in this calculation are the approximate average values, i.e., 3.3 and 8.5 for the C-terminal and N-terminal, respectively. The absolute mean net charge is defined as the ratio of the absolute value of the net charge to the total number of residues.

III. RESULTS AND DISCUSSION

A. Protein Selection. The database of 2321 monomeric proteins having disordered regions is selected from the supplementary data set of DisProt, i.e., Database of Protein Disorder,[44] where the disordered regions containing ≥3 residues are listed in the Fasta Header. The coordinates of these proteins are extracted from the (August-2003 list for disordered proteins and July-2009 list for globular and natively unstructured protein) Protein Data Bank (PDB)[65,66] deter-

dx.doi.org/10.1021/jp3016529 | J. Phys. Chem. B 2012, 116, 6326−6335

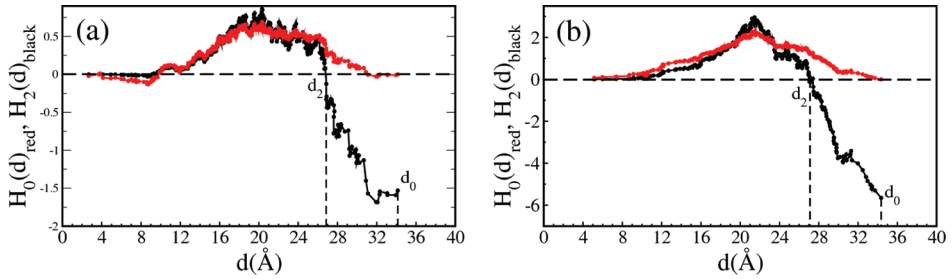


Figure 1. Zero (red) and second order (black) ellipsoidal hydrophobicity profile for (a) the globular proteins and (b) group I proteins.

mined by X-ray crystallography and comprises of the three-dimensional coordinates of the ordered regions. The X-ray diffraction structures of these 2321 proteins have resolution less than 3.7 Å and *R* factor less than 0.27. Various interactions among the residues of protein chain play a significant role in protein folding to attain the functional native structure of which the hydrophobic interactions are recognized to play a key role for the initiation of folding.[67−69] It is also noted that large polar and charged groups also participate in stabilizing the initial folded state, while changing the protein sequence may affect the initial folding state. A set of 3967 monomeric proteins constitute the database of globular proteins.

The database of natively unstructured proteins comprise of 215 monomeric proteins with <5% of secondary structure.[71−74] The structural analysis in our previous study reveals that these proteins exhibit an intermediate scaling regime between a collapsed and a random walk, are less compact, more aspherical and show higher flexibility which closely resemble the conformational characteristics of the natively unstructured proteins.[56,75] Specifically, these proteins are enriched with polar and charged residues as compared to the globular proteins. Analysis of the two-body contacts show that their structures are primarily stabilized due to the large number of side-chain backbone contacts, predominantly between the polar and charged residues.[56]

Natively unstructured proteins are characterized by a combination of low mean net hydrophobicity and relatively high mean net charge, which is an important prerequisite for the absence of compact structure under physiological conditions.[48,76] Thus, these proteins are localized within a specific region of charge-hydrophobicity phase space, which is different compared to that of the folded globular proteins. The proteins constituting this data set exhibit low compactness, low secondary structure content, higher flexibility, predominance of polar and charged residues, high mean net charge, and low mean net hydrophobicity. However, these proteins have stable structures either due to various interactions with specific targets like natural ligands in vivo.[77] or they are frozen in stable conformations by the local context.[48] A comparative study of the sequences of these proteins and those of the known natively unstructured proteins obtained from the current DisProt database released on February 23, 2012, justifies the choice of such a data set as representative members of the accessible states of the natively unstructured proteins, discussed in subsection E of the Results section.

Most globular protein structures (except 10 structures which are determined by NMR) are determined by X-ray diffraction with a resolution of less than 3.5 Å and *R* factor less than 0.35. The atomic coordinates of the three-dimensional structures of the natively unstructured proteins are determined by NMR, X-ray crystallography and cryoelectron microscopy. Of the 215 proteins present in the database of natively unstructured proteins, 132 have X-ray diffraction structures of resolution ranging from 1.9−3.7 Å and *R* factor less than 0.39, 33 are NMR resolved structures and 50 structures are determined by cryoelectron microscopy with resolution ranging from 6 to 13 Å.

For this study, we calculated the zero-order and second-order hydrophobic moments for the selected data set of 2321 proteins. The spatial profile of the second-order hydrophobic moment is generated with respect to the generalized ellipsoidal radius and the proteins with disordered regions are classified into three groups. Group I consists of 614 proteins exhibiting similar spatial profile for the second-order moment as that of the globular proteins, while group II consists of 1409 proteins displaying considerable deviation from that of the globular proteins. Group III comprises of 298 proteins with hydrophobicity profile same as that of the natively unstructured proteins.

The classification of proteins as group I, group II, and group III is independent of the length or position of the disordered regions present. Among 614, 1409, and 298 proteins in group I, group II, and group III respectively, 377, 753, and 139 proteins have single disordered regions, while 237, 656, and 159 proteins have multiple disordered regions. The majority of the proteins, i.e., 256, 490, and 85 proteins in group I, group II, and group III respectively have single disordered regions of <10 residues. Disordered regions of 10−30 residues occurs in 109, 216, and 39 proteins in the respective groups, while 14, 45, and 15 proteins have disordered regions of more than 30 residues with a maximum length of 100, 87, and 113 residues in group I, group II, and group III, respectively.

B. Hydrophobic Moments. Figure 1a represents the detailed $H_0(d)$ and $H_2(d)$ profile for the 485 residue globular protein, 1AUK, which is a representative example for the whole database for 3967 globular proteins. Profile of $H_0(d)$ is magnified to a comparable scale as that of $H_2(d)$. From the figure it is evident that the value of $H_0(d)$ increases with d starting from d = 0 Å until a maximum value at d = 20.25 Å. The value then decreases and at d_0 = 34.5 Å, $H_0(d)$ = 0. For $H_2(d)$ the maximum value is attained at the same distance d = 20.25 Å. The first negative value is obtained at d_2 = 27 Å, consequently it decreases continuously and reaches a minimum at d_0 = 34.5 Å. Thus, the hydrophobic ratio R_H is found to be 27/34.5 = 0.78.

Figure 1b depicts the profile for group I proteins which has similar $H_0(d)$ and $H_2(d)$ profile as that of the globular proteins. 1*AA*4, a 291 residue protein is a representative example from group I. For 1AA4, the value of $H_0(d)$ starts from d = 0 Å and shows a maximum at d = 21.5 Å. The value then decreases until it becomes zero at d_0 = 34.4 Å. The value of $H_2(d)$ reaches a maximum at 21.5 Å followed by a decrease. At d_2 = 27 Å, $H_2(d)$

dx.doi.org/10.1021/jp3016529 | J. Phys. Chem. B 2012, 116, 6326−6335

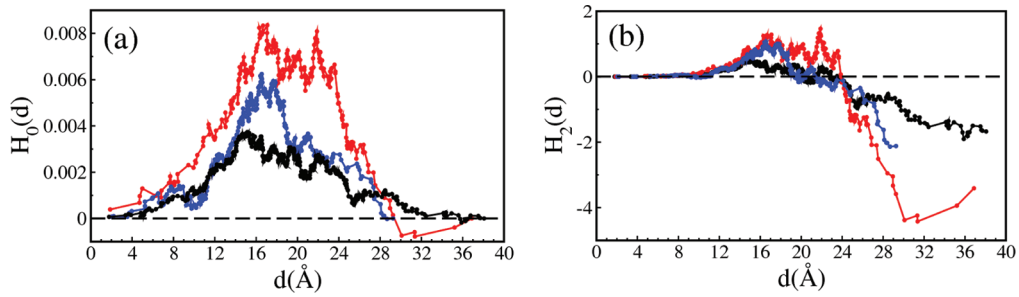


Figure 2. Zero (a) and second order (b) ellipsoidal hydrophobicity profile for group II proteins,1A6R (blue), 1CBY (black), and 1FO8 (red).

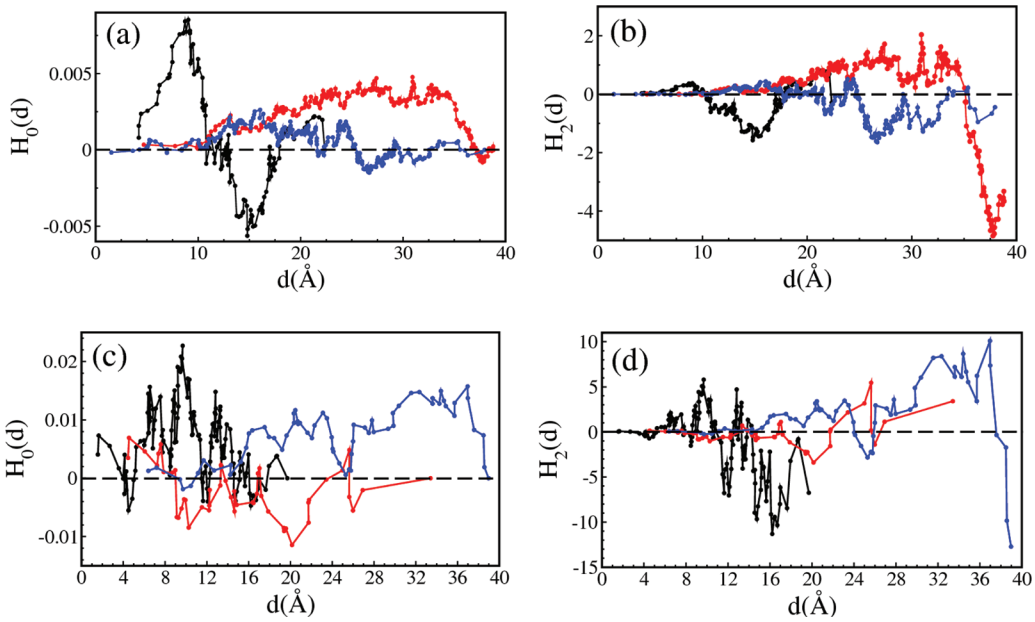


Figure 3. Zero (a and c) and second order (b and d) ellipsoidal hydrophobicity profile for group III proteins and natively unstructured proteins, respectively.

becomes negative and then decreases continuously until d_0 = 34.4 Å. Thus, the hydrophobic ratio is 27/34.4 = 0.78. These two examples also illustrate that the second-order hydrophobic moment profile is relatively independent of the size of the proteins since both 485 and 291 residue proteins exhibit a similar trend.

Parts a and b of Figure 2 represent the $H_0(d)$ and $H_2(d)$ profiles for the group II proteins. Here, $H_0(d)$ and $H_2(d)$ profiles for 1A6R, 1CBY, and 1FO8 are plotted as representative examples of proteins belonging to group II. The profile of $H_0(d)$ for all the three proteins show that the values start from zero, increase with increasing values of d with a maxima at 15.2, 20, and 16.7 Å for 1A6R, 1CBY, and 1FO8 respectively; consequently, the values decrease and near d_0 values are negative. For 1A6R, the initial values of $H_2(d)$ profile are negative crossing over to positive values followed by a less demarcated crossover from positive to negative values after attaining the maximum value of $H_2(d)$. For 1CBY, the crossover from positive to negative values of $H_2(d)$ profile is not well-defined, while there are more than one values of d for which the value of $H_2(d)$ vanishes. The $H_2(d)$ profile for 1J08 depicts a well-defined crossover from positive to negative values. The value of $H_2(d)$ increases until $d = d_0$ = 37 Å after a minimum value of $H_2(d)$ at d = 30 Å. For calculating the hydrophobic ratio ($R_H = d_2/d_0$) of group II proteins whose crossover is not well-defined, the value of d_2 is the first negative value after which $H_2(d)$ remains negative until $d = d_0$.

Parts a and b of Figure 3 depict the hydrophobicity profile for group III proteins, typically represented by 3 proteins 1A6R, 1CBY, and 1FO8. Unlike globular proteins, the hydrophobic profile does not exhibit any significant distinction between the core and surface. The profile of the zeroth order moment displays that the value of $H_0(d)$ fluctuates around zero and the profile is skewed compared to a Gaussian distribution. $H_0(d)$ becomes negative near d_0 for which $H_0(d) = 0$. Similarly, the second order moment values fluctuate around zero throughout the distance d. The hydrophobicity profile exhibits a crossover from positive values to small negative values. $H_2(d)$ shows an increase from the most negative values until d_0. Similar spatial profile of $H_0(d)$ and $H_2(d)$ is also found in the natively unstructured proteins as shown in Figure 3, parts c and d. No definite crossover from the positive to the negative values is observed. Hence the hydrophobic ratio for these proteins cannot be defined and due to the absence of a specific hydrophobic profile we cannot generalize any shape for the hydrophobicity profile of group III and the natively unstructured proteins.

The irregular profile of $H_2(d)$ for group III and natively unstructured proteins may be attributed to the increase in polar and charged residues at the expense of hydrophobic residues. Polar and positively charged residues are necessary for the

dx.doi.org/10.1021/jp3016529 | J. Phys. Chem. B 2012, 116, 6326−6335

stability of protein complexes in providing specific recognition and strong binding via the formation of hydrogen bonds as opposed to the nonspecific hydrophobic interactions.[70,78] The fluctuating $H_2(d)$ profile of the group III and natively unstructured proteins indicate the presence of polar residues at specific positions throughout the protein which facilitate specific stabilizing interactions. The values of $H_2(d)$ at the cross over region denote the presence of positively charged residues on the surface as they have the least hydrophobicity. Similar conclusions may be inferred from the $H_0(d)$ profile where the positively charged residues appear near d_0 promoting specific stabilizing contacts and reducing the risk of aggregation of the unfolded protein.

The $H_2(d)$ profile of 1A6R in group II is presented as an example. Figure 4 represents the $H_2(d)$ profile of 1A6R, which

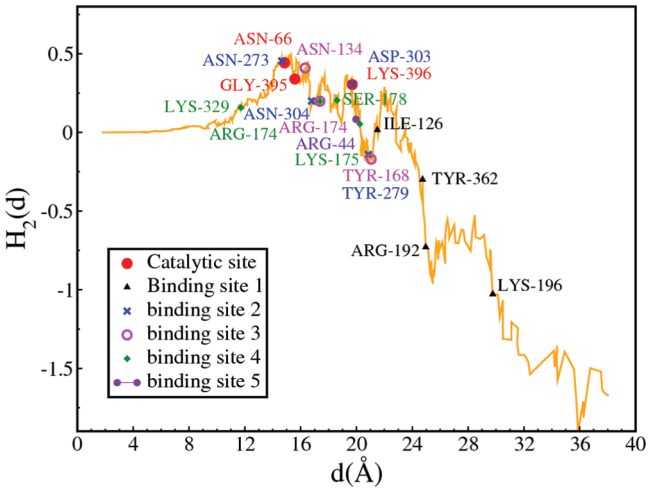


Figure 4. Different active sites represented in second order hydrophobic profile of group II protein 1A6R.

consists of one catalytic and five binding sites. All catalytic and binding sites, except the first binding site, are located from 14.75 to 21.0 Å of the proteins and the first binding site is located in the distance range 21.59−29.84 Å. All these sites mainly comprises of polar and charged residues. Specific binding sites 1, 4, and 5 are mostly composed of the charged residues, which help in stabilizing the specific contacts, while the binding sites 2 and 3 along with the catalytic sites contain more polar residues, which help in specific recognition by forming stabilizing hydrogen bonds. It is also evident from figure that polar and charged residues are widely distributed throughout the protein as the value of $H_2(d)$ fluctuates throughout the length d, while an abrupt decrease of $H_2(d)$ values correspond to the presence of positively charged residues, *ARG* and *LYS*, which participates in the specific binding of the protein. It is also evident from the profile that the disordered region comprising of mostly charged residues occur on the surface (and the disordered region which is generally toward the surface also contains large number of charged residues) reducing the risk of aggregation. The charged residues present in the intermediate region stabilizes specific contacts. The broad profile of $H_2(d)$ highlights a large region for specific recognition due to a wide distribution of polar residues throughout the protein.

Figure 5 compares the distribution of the hydrophobic ratio R_H of the globular proteins to that of group I and group II proteins. The average value of the hydrophobic ratio for 3967

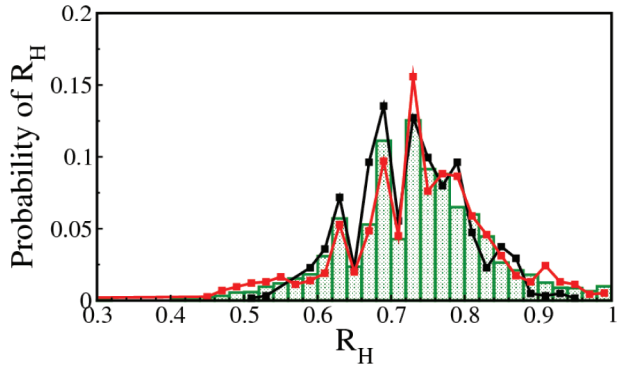


Figure 5. Hydrophobic ratio for group I (black), group II (red), and globular proteins (green).

globular proteins is 0.72, which is similar to the value calculated in a previous study of globular proteins.[55] For group I proteins, R_H value ranges from 0.58−0.88 with an average of R_H = 0.728. For group II proteins, R_H has a broader range of values lying between 0.45−0.9 with an average of R_H = 0.73. Group I proteins show a similar $H_2(d)$ profile as that of globular proteins with a higher average value of R_H. Group II proteins show a broader distribution of R_H with a higher average value as compared to the globular proteins. For all globular, group I and group II proteins the hydrophobic residues are mainly preferred in the core region of the protein as compared to the hydrophilic residues which are favored toward the surface. The increase in the number of hydrophilic residues toward the surface results in the crossover of $H_2(d)$ values. The R_H values or the ratio of crossover distance to the total distance, are primarily determined by the decrease in the hydrophobic residue density from the protein core to the surface with a marginal dependence on the differental accumulation of the hydrophobic and hydropilic residues. The decrease of this residual density yields different ranges of R_H.

C. Shape Parameters. *Asphericity.* Asphericity parameter quantifies the mean deviation of the chain conformation from spherical symmetry. The distribution of asphericity for globular, group I, group II, group III, and natively unstructured proteins is illustrated in Figure 6. For globular monomeric proteins, the value of Δ varies from 0−0.1 for 80.56% proteins, while for 16.69% proteins, it lies between 0.1−0.3. Distribution of the asphericity parameter for group I and group II proteins is similar to that of the globular proteins with Δ varying from 0−

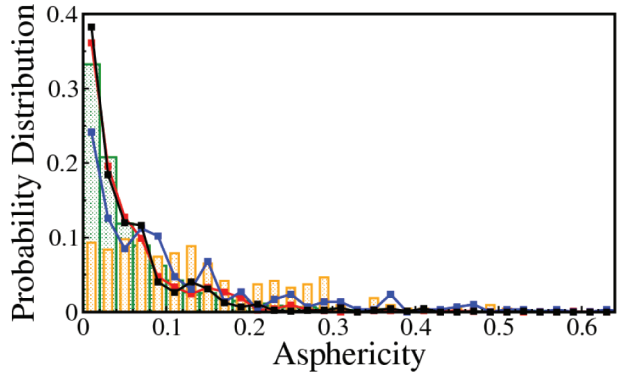


Figure 6. Distribution of asphericity parameter for globular (green), group I (black), group II (red), group III (blue), and natively unstructured (orange) proteins.

dx.doi.org/10.1021/jp3016529 | J. Phys. Chem. B 2012, 116, 6326−6335

0.1 for 84.33% and 83.1% proteins respectively and from 0.1−0.3 for 13.6% and 15.85% proteins, respectively. The distribution for group III shows some broadening as only for 66.67% proteins Δ varies from 0−0.1, while for 25.17% proteins Δ ranges between 0.1−0.3 and remaining 8.61% proteins have Δ values lying between 0.3−0.63. For the natively unstructured proteins, 44.65% and 48.83% proteins are with Δ values ranging between 0−0.1 and 0.1−0.3 respectively. This shows that irrespective of the considerable differences in their spatial hydrophobic profile, group I and group II proteins have similar topology as compared to that of the globular proteins. The spatial hydrophobic profile of group III proteins closely resemble the natively unstructured proteins, even though the distribution of the asphericity parameter is narrower compared to that of the natively unstructured proteins. The proteins belonging to group III exhibits maximum deviation from the nearly spherical shape of the globular proteins.

The distribution of asphericity parameter plotted as a function of the mean hydrophobicity of the protein is illustrated in Figure 7. The figure shows no perceptible difference of the

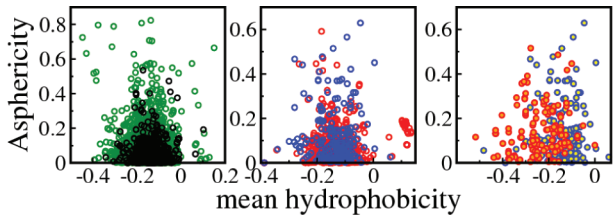


Figure 7. Asphericity vs mean net hydrophobicity distribution for globular (green), group I (black), group II (red), group III (blue), and natively unstructured(orange) proteins.

asphericity distribution of group I, group II and group III proteins compared to that of the globular proteins with similar range of the mean hydrophobicity. Group III proteins differ from the natively unstructured proteins in the range of mean hydrophobicity and asphericity. The natively unstructured proteins record a lower value of mean hydrophobicity as compared to the group III proteins and a higher value of asphericity. This observation implies that the natively unstructured proteins are more aspherical and less hydrophobic compared to the group III proteins, irrespective of similar $H_2(d)$ profile.

Shape. Figure 8 depicts the distribution of the shape parameter for globular, group I, group II, group III and natively unstructured proteins. For globular, group I, and group II proteins 35%, 33% and 36.2% proteins, respectively, have S values lying between −0.02 and 0, while 45.35%, 50.98%, and 48.4% proteins, respectively, have values of S varying from 0 to 0.02. There are no proteins in group I and group II with $S < -0.02$, while for the globular proteins, only 3% proteins have $S < -0.02$. This shows that the shape of group I and group II proteins are spherically symmetrical closely resembling that of the globular proteins. Shape distribution for group III proteins show that only 2.25% proteins have $S < -0.02$, while 26.87% and 35.71% proteins have S values ranging from −0.02−0 and 0−0.02 respectively and 30% proteins have $0.02 < S < 0.3$. Distribution for group III is broader compared to group I and group II proteins, while $S > 0$ indicate predominantly prolate shapes. The shape parameter distribution for the natively unstructured proteins is broad compared to the group III proteins, indicating the occurrence of both prolate and oblate

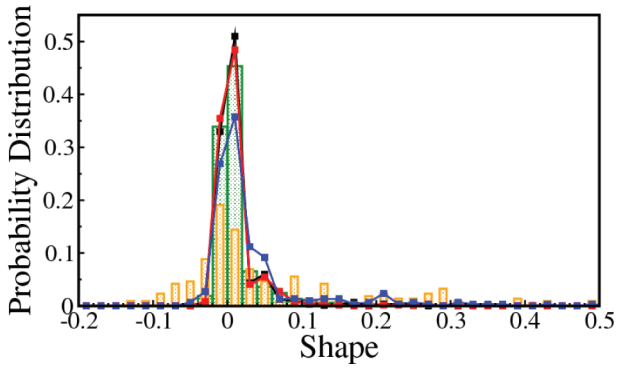


Figure 8. Distribution of shape parameter for globular (green), group I (black), group II (red), group III (blue), and natively unstructured (orange) proteins.

shapes. Approximately 40% proteins are oblate with $S < 0$, of which 19.1% proteins have $-0.02 \leq S \leq 0$. This distribution exhibits maximum deviation from spherical symmetry.

D. Two Body Contact Analysis. The two body contacts are calculated for the different classes of proteins viz. globular, group I, group II, group III and natively unstructured proteins to evaluate the degree of packing which finally determines the shape of the protein. Side-chain side-chain (*ss*), side-chain backbone (*sb*) and backbone−backbone (*bb*) contacts are calculated in this study with the cutoff for the number and type of two body contacts as D_{ss} = 5.2 Å, D_{sb} = 5.5 Å, D_{bb} = 6.5 Å using nine classes of amino acids.[61] Figure 9 shows the

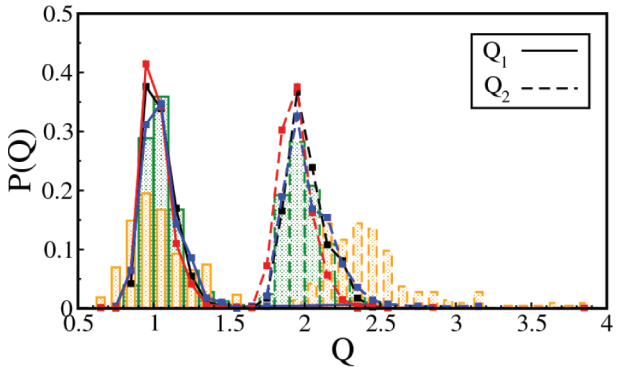


Figure 9. Distributions of $Q_1 = N_{bb}/N_{sb}$ and $Q_2 = N_{ss}/N_{sb}$ in globular (green), group I (black), group II (red), group III (blue) and natively unstructured (orange) proteins.

distribution, $P(Q_1)$ and $P(Q_2)$ for $Q_1 = N_{bb}/N_{ss}$, and $Q_2 = N_{sb}/N_{ss}$ where N_{bb}, N_{sb}, and N_{ss} are the numbers of *bb*, *sb*, and *ss* contacts, respectively. The distributions $P(Q_1)$ and $P(Q_2)$ are almost similar for the globular, group I, group II, and group III proteins, displaying a slight deviation from each other with peaks at 1 and 2 respectively. This implies that *sb* contacts are predominant among the two body contacts, while *ss* and *bb* contacts have similar frequency. Presence of higher *sb* contacts imply that the molecules have maximum short-range interactions, while the diminished *ss* contact frequency indicates that the molecule is less densely packed. Both $P(Q_1)$ and $P(Q_2)$ distributions of globular proteins are found to be similar to those of the disordered proteins (group I, group II, and group III) implying that irrespective of the presence of disordered regions and differences in the second order spatial hydrophobic moment profile, the ordered regions of disordered proteins

dx.doi.org/10.1021/jp3016529 | J. Phys. Chem. B 2012, 116, 6326−6335

show similar packing and therefore the overall shape of disordered proteins closely resemble the globular proteins. For the natively unstructured proteins, $P(Q_2)$ distribution has a broader range with a maximum at 2.5, showing a significant difference from that of the globular proteins. This implies that the *sb* contacts are higher in natively unstructured proteins with 36.05% hydrophobic, 22.55% polar residues while 31.7% are charged amino acids. The increase in polar and charged residues promotes specificity of recognition and stability of the set of unfolded states of disordered proteins since these interactions are dominated by the hydrogen bonds and ion pairs.[79] The *sb* contacts for globular and the three groups of disordered proteins show that approximately 41.81% residues are hydrophobic, 21.69% are polar and 25.02% are charged residues which constitute these contacts.

E. Charge. Figure 10 depicts the mean net charge vs the mean net hydrophobicity distribution for 408 known natively

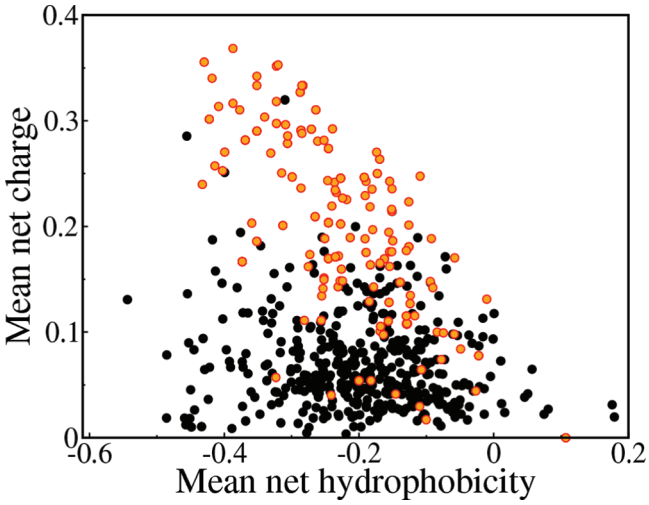


Figure 10. Mean net charge vs mean net hydrophobicity distribution for natively unstructured proteins: (a) DisProt database (black) and (b) Proteins (orange) with <5% secondary structure.

unstructured proteins from DisProt database as compared to our data set of 215 proteins with <5% secondary structure. Proteins with least secondary structure are located in a specific region defined by the mean net charge and the mean net hydrophobicity similar to the natively unstructured proteins. Most proteins in our data set are present in the region with low mean net hydrophobicity and high mean net charge. Hence this set of proteins may represent accessible states of natively unstructured proteins.

The mean net charge plotted as a function of the mean net hydrophobicity for the globular, group I, group II, group III, and natively unstructured proteins is depicted in Figure 11. Figure 11, parts a_1, b_1. and c_1, shows the respective distribution for the globular proteins, natively unstructured proteins and the ordered regions of the three classes of disordered proteins. Figure 11, parts a_2, b_2, and c_2, portrays the distribution for the entire disordered protein compared to those of the globular and natively unstructured proteins. The net charge vs net hydrophobicity of the disordered proteins with and without disordered parts clearly show a small decrease in hydrophobicity accompanied by a significant increase of charge for all disordered proteins. This observation is in accord with the fact that the disordered sequences contain predominantly charged

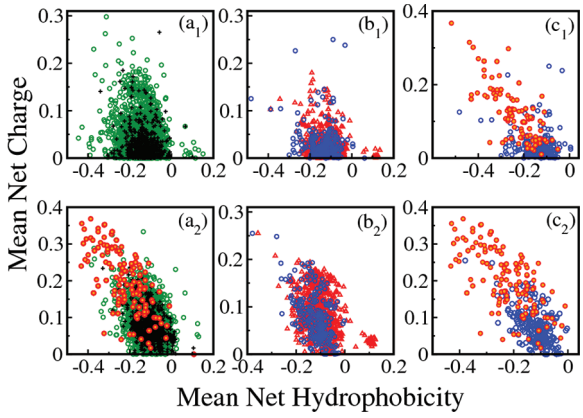


Figure 11. Mean net charge vs mean net hydrophobicity distribution for globular (green), group I (black), group II (red), group III (blue), and natively unstructured (orange) proteins. a_1, b_1, and c_1 denote the ordered part of disordered proteins and a_2, b_2, and c_2 for the entire disordered proteins including the disordered regions.

residues with very few hydrophobic groups. For all three classes of disordered proteins, the plot of charge vs hydrophobicity is similar to that of the globular proteins implying that these proteins cannot be markedly distinguished from the globular proteins in the presence of ordered regions.

Figure 11c_2 shows the group III proteins (which have similar second order hydrophobic profile as that of natively unstructured proteins) have higher mean hydrophobicity and lower mean charge as compared to the natively unstructured proteins. With exception of few proteins which are in the intermediate region, there is a clear distinction between the two regions. This implies that mean net charge vs mean hydrophobicity value is not able to distinguish the disordered proteins from the ordered ones and the ordered parts play a dominant role in defining the physical properties of disordered proteins. The mean net charge vs mean hydrophobicity value distribution, Figure 11a_2 shows that most of the natively unstructured proteins have high mean net charge with low mean hydrophobicity under physiological conditions compared to the globular proteins. With few exceptions, two different regions may be demarcated in the distribution for the mean net charge vs the mean hydrophobicity corresponding to the globular and the natively unstructured proteins.

IV. CONCLUSIONS

This study focuses on the comparison of the spatial hydrophobic profile, shape, packing, and mean charge of disordered proteins with those of globular proteins. Natively unstructured proteins exhibit a clear distinction from the globular proteins in all these physical properties. Natively unstructured proteins are more aspherical and display an irregular spatial profile of second order hydrophobic moment. The shape parameter distribution indicates the dominant preference toward prolate structures, even though 40% proteins are oblate. The profile of the two body contacts differs considerably from the globular proteins with a significant increase in the *sb* contacts implying an increase in the number of polar and charged residues. Under physiological conditions, these proteins are characterized by a combination of low mean hydrophobicity and relatively high net charge.

All groups of disordered proteins closely resemble the globular proteins. Except a small group of disordered proteins

dx.doi.org/10.1021/jp3016529 | J. Phys. Chem. B 2012, 116, 6326−6335

(group III proteins), the shape parameters, two body contacts and the mean charge vs mean hydrophobicity reveal a similar trend as those of globular proteins. The only notable difference is exhibited in the zeroth order and second order hydrophobicity profile of disordered proteins where 71% disordered proteins (i.e., group II and group III proteins) exhibit a considerable deviation from the spatial profile of the globular proteins irrespective of the length or position of disordered regions. The spatial profile does not distinguish the hydrophobic core from the surface of the protein. Both moments fluctuate around zero and the second order moment exhibits a crossover from positive values to small negative values. This study provides a structural characterization of the disordered proteins and distinguishes them from the globular ones which may help in elucidating their specific functions.

V. APPENDIX I

List of natively unstructured proteins.

1ANSA, 1AQRA, 1BCVA, 1FP0A, 1G1SC, 1GNAA, 1GNBA, 1K5KA, 1KB7A, 1KB8A, 1KC4B, 1KKDA, 1KL8B, 1LXGB, 1LXHB, 1MEAA, 1MJ1L, 1MJ1O, 1MJ1P, 1NIMA, 1PAJA, 1PAKA, 1PANA, 1S1I0, 1S1I9, 1S1IJ, 1S1IS, 1S1IV, 1S1IY, 1S1IZ, 1TACA, 1TBCA, 1TORA, 1TVTA, 1U3NA, 1VOY1, 1VOY2, 1VOY3, 1VOY5, 1VOY6, 1VOYD, 1VOYF, 1VOYH, 1VOYI, 1VOYJ, 1VOYL, 1VOYM, 1VOYQ, 1VOYS, 1VOYW, 1VOYX, 1VP01, 1VP02, 1VP05, 1VP06, 1VP0D, 1VP0F, 1VP0H, 1VP0I, 1VP0J, 1VP0L, 1VP0Q, 1VP0R, 1VP0W, 1VP0X, 1VS61, 1VS64, 1VS6C, 1VS6D, 1VS6L, 1VS6P, 1VS6R, 1VS6U, 1VS6W, 1VS81, 1VS84, 1VS8C, 1VS8P, 1VS8U, 1VS8W, 1VSAP, 1VSAS, 1VSPS, 1WENA, 1WESA, 1WG2A, 2ABXA, 2AW44, 2AW4C, 2AW4L, 2AW4P, 2AW4R, 2AW4U, 2AW4W, 2AWB1, 2AWB4, 2AWBC, 2AWBL, 2AWBP, 2AWBR, 2AWBU, 2AWBW, 2B660, 2B9N0, 2B9P0, 2GYA1, 2GYAD, 2GYAH, 2GYAJ, 2GYAN, 2GYAR, 2GYAU, 2GYAZ, 2GYC1, 2GYCH, 2GYCJ, 2GYCN, 2GYCR, 2GYCU, 2I2TW, 2J016, 2J01C, 2J036, 2J03C, 2J281, 2J284, 2J28C, 2J28L, 2J28P, 2J28R, 2J28U, 2J28W, 2JL6C, 2JL8C, 2JZWA, 2K03D, 2K04D, 2K05D, 2MADH, 2V494, 2V496, 2V49C, 2VHN1, 2VHN4, 2VHNC, 2VHNI, 2VHNL, 2VHNU, 2VHNW, 2YS2A, 2Z4LW, 2ZJP1, 2ZJP3, 2ZJQ3, 2ZJQ5, 2ZJR3, 2ZKR2, 2ZKRY, 3BBO1, 3BBO2, 3BBO3, 3BBOE, 3BBOF, 3BBON, 3BBOT, 3BBOW, 3BBOX, 3BBOY, 3BBX0, 3BBX1, 3BBX4, 3BBXC, 3BBXD, 3BBXP, 3BBXR, 3BBXU, 3BBXW, 3CF51, 3CF53, 3DLL1, 3DLL3, 3DLLI, 3FIN6, 3FINC, 4MT2A.

Of 215 proteins, 185 proteins are listed above as the natively unstructured proteins. Among those 185 proteins, 92 proteins have less than 40% sequence similarity with the other proteins in the whole database and 93 proteins have approximately 90% sequence similarity to some of these proteins. But all 93 proteins distinctly differ in the physicochemical properties such as size, asphericity, persistence length, two-body contacts, and $H_2(d)$ profile from their homologous counterparts. The remaining 30 proteins, which are not listed above, have similar sequence and approximately similar physicochemical properties, viz., size, asphericity, persistence length, $H_2(d)$ profile, and two body contacts, as compared to the natively unstructured proteins listed above.

AUTHOR INFORMATION

Corresponding Author

*E-mail: pbiswas@chemistry.du.ac.in.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Authors gratefully acknowledge the financial assistance from Delhi University. N.R. acknowledges the financial support from CSIR India in the form of a Senior Research Fellowship.

REFERENCES

(1) Wrabl, J. O.; Gu, J.; Liu, T.; Schrank, T. P.; Whitten, S. T.; Hilser, V. J. *Biophys. Chem.* **2011**, *159*, 129−141.
(2) Tompa, P. *Curr. Opin. Stru. Biol.* **2011**, *21*, 419−425.
(3) Uversky, V. N.; Dunker, A. K. *Biochim. Biophys. Acta* **2010**, *1804*, 1231−1264.
(4) Tompa, P.; Fersht, A. *Structure and function of intrinsically disordered proteins*; Chapman and Hall/CRC Taylor and Francis Group: Boca Raton, 2010.
(5) Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197−208.
(6) DisProt, *release 5.0* 2010 (http://www.disprot.org).
(7) Daughdrill, G. W.; Pielak, G. J.; Uversky, V. N.; Cortese, M. S.; Dunker, A. K. *Handbook of Protein Folding (Natively disordered proteins)*; Wiley-VCH Verlag GmbH: Weinheim, Germany, 2005, pp 271−353.
(8) Dosztányi, Z.; Bálint Meszáros, B.; Simon, I. *Brief. Bioinfom.* **2009**, *11*, 225−243.
(9) Dosztányi, Z.; Csizmok, V.; Tompa, P.; Simon, I. *Struct. Bioinform.* **2005**, *21*, 3433−3434.
(10) Linding, R.; Jensen, L. J.; Diella, F.; Bork, P.; Gibson, T. J.; Russell, R. B. *Structure* **2003**, *11*, 1453−1459.
(11) Iakoucheva, L. M.; Dunker, A. K. *Structure* **2003**, *11*, 1316−1317.
(12) Uversky, V. N. *Int. J. Biochem. Cell Biol.* **2011**, *43*, 1090−1103.
(13) Uversky, V. N. *Protein Sci.* **2002**, *11*, 739−756.
(14) Dunker, A. K.; Obradovic, Z.; Romero, P.; Garner, E. C.; Brown, C. J. *Genome Inform. Ser. Workshop Genome Inform.* **2000**, *11*, 161−171.
(15) Brown, C. J.; Takayama, S.; Campen, A. M.; Vise, P.; Marshall, T. M.; Oldfield, C. J.; Williams, C. J.; Dunker, A. K. *J. Mol. Evol.* **2002**, *55*, 104−110.
(16) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; et al. *J. Mol. Graph. Model.* **2001**, *19*, 26−59.
(17) Dunker, A. K.; Silman, I.; Uversky, V. N.; Sussman, J. L. *Curr. Opin. Struct. Biol.* **2008**, *18*, 756−764.
(18) Turoverov, K. K.; Kuznetsova, I. M.; Uversky, V. N. *Prog. Biophys. Mol. Biol.* **2010**, *102*, 73−84.
(19) Seetharaman, J. K.; Oikawa, M.; Grimshaw, S. B.; Wirmer, J.; Duchardt, E.; Ueda, T.; Imoto, T.; Smith, L. J.; Dobson, C. M.; Schwalbe, H. *Science* **2002**, *295*, 1719−1722.
(20) Ptitsyn, O. B. *Adv. Protein Chem.* **1995**, *47*, 83−229.
(21) Plaxco, K. W.; Gross, M. *Nat. Struct. Biol.* **2001**, *8*, 659−660.
(22) Tanford, C. *Adv. Protein Chem.* **1968**, *23*, 121−282.
(23) Tantos, A.; Han, K.; Tompa, P. *Mol. Cel. Endocrin.* **2012**, *348*, 457−465.
(24) Iakoucheva, L. M.; Radivojac, P.; Brown, C. J.; ÓConnor, T. R.; Sikes, J. G.; Obradovic, Z.; Dunker, A. K. *Nucleic Acids Res.* **2004**, *32*, 1037−1049.
(25) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. *J. Mol. Recog.* **2005**, *18*, 343−384.
(26) Tompa, P.; Fuxreiter, M.; Oldfield, C. J.; Simon, I.; Dunker, A. K.; Uversky, V. N. *Bioessays* **2009**, *31*, 328−335.
(27) Radivojac, P.; Iakoucheva, L. M.; Oldfield, C. J.; Obradovic, Z.; Uversky, V. N.; Dunker, A. K. *Biophys. J.* **2007**, *92*, 1439−1456.
(28) Permyakov, S. E.; Bakunts, A. G.; Denesyuk, A. I.; Knyazeva, E. L.; Uversky, V. N.; Permyakov, E. A. *Proteins* **2008**, *72*, 822−836.
(29) Uversky, V. N. *Focus On Structural Biology*; Springer: Berlin, 2009.

(30) Uversky, V. N.; Oldfield, C. J.; Midic, U.; Xie, H.; Xue, B.; Vucetic, S.; Iakoucheva, L. M.; Obradovic, Z.; Dunker, A. K. *BMC Genomics* **2009**, *10*, S7.
(31) Glenner, G. G.; Wong, C. W. *Biochem. Biophys. Res. Commun.* **1984**, *120*, 885−890.
(32) Masters, C. L.; Multhaup, G.; Simms, G.; Pottgiesser, J.; Martins, R. N.; Beyreuther, K. *EMBO J.* **1985**, *4*, 2757−2763.
(33) Lee, V. M.; Balin, B. J.; Otvos, L., Jr.; Trojanowski, J. Q. *Science* **1991**, *251*, 675−678.
(34) Weinreb, P. H.; Zhen, W.; Poon, A. W.; Conway, K. A.; Lansbury, P. T., Jr. *Biochemistry* **1996**, *35*, 13709−13715.
(35) Cheng, Y.; LeGall, T.; Oldfield, C. J.; Mueller, J. P.; Van, Y. Y.; Romero, P.; Cortese, M. S.; Uversky, V. N.; Dunker, A. K. *Trends Biotechnol.* **2006**, *24*, 435−442.
(36) Wallace, J.; Daman, O. A.; Harris, F.; Phoenix, D. A. *Theor. Biol. Med. Model* **2004**, *1*, 5.
(37) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. *Annu. Rev. Biophys.* **2008**, *37*, 215−246.
(38) Demarest, S. J.; Zhou, S.; Robblee, J.; Fairman, R.; Chu, B.; Raleigh, D. P. *Biochemistry* **2001**, *40*, 2138−2147.
(39) Hebrard, E.; Bessin, Y.; Michon, T.; Longhi, S.; Uversky, V. N.; Delalande, F.; Dorsselaer, A. V.; Romero, P.; Walter, J.; Declerck, N.; et al. *Virol. J.* **2009**, *6*, 23.
(40) Blavatska, V.; Janke, W. *J. Chem. Phys.* **2010**, *133*, 184903.
(41) Oroguchi, T.; Ikeguchi, M.; Sato, M. *J. Phys. (Paris)* **2011**, *272*, 012005.
(42) Duvignaud, J.; Savard, C.; Fromentin, R.; Majeau, N.; Leclerc, D.; Gagn, S. M. *Biochem. Biophys. Res. Commun.* **2009**, *378*, 27−31.
(43) Lobanov, M. Y.; Garbuzynskiy, S. O.; Galzitskaya, O. V. *Biochemistry (Moscow)* **2010**, *75*, 192−200.
(44) Sickmeier, M.; Hamilton, J. A.; LeGall, T.; Vacic, V.; Cortese, M. S.; Tantos, A.; Szabo, B.; Tompa, P.; Chen, J.; Uversky, V. N. et al. *Nucleic Acids Res.* **2007**, *35 (Database issue)*, D786−D793.
(45) He, B.; Wang, K.; Liu, Y.; Xue, B.; Uversky, V. N.; Dunker, A. K. *Cell Res.* **2009**, *19*, 929−949.
(46) Goh, G. K.; Dunker, A. K.; Uversky, V. N. *BMC Genom.* **2008**, *9* (Suppl 2), S4.
(47) Maoa, A. H.; Crick, S. L.; Vitalisa, A.; Chicoine, C. L.; Pappua, R. V. *Proc. Natl. Aacd. Sci. U.S.A.* **2010**, *107*, 8183−8188.
(48) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. *Proteins* **2000**, *41*, 415−427.
(49) Mller-Spath, S.; Soranno, A.; Hirschfeld, V.; Hofmann, H.; Regger, S.; Reymond, L.; Nettels, D.; Schuler, B. *Proc. Natl. Aacd. Sci. U.S.A.* **2010**, *107*, 14609−14614.
(50) Sandelin, E. *Biophys. J.* **2004**, *86*, 23−30.
(51) Silverman, B. D. *J. Theor. Biol.* **2002**, *216*, 139−146.
(52) Silverman, B. D. *Proc. Natl. Aacd. Sci.* **2001**, *98*, 4996−5001.
(53) Silverman, B. D. *Proteins* **2003**, *53*, 880−888.
(54) Zhou, R.; Silverman, B. D.; Royyuru, A. K.; Athma, P. *Proteins* **2003**, *52*, 561−572.
(55) Zhou, R.; Silverman, B. D. *Pac. Symp. Biocomp.* **2002**, *7*, 673−684.
(56) Rawat, N.; Biswas, P. *J. Chem. Phys.* **2009**, *131*, 165104.
(57) Aronovitz, A.; Nelson, D. R. *J. Phys. (Paris)* **1986**, *47*, 1445−1456.
(58) Dima, R. I.; Thirumalai, D. *J. Phys. Chem. B* **2004**, *108*, 6564−6570.
(59) Rawdon, E. J.; Kern, J. C.; Piatek, M.; Plunkett, P.; Stasiak, A.; Millett, K. C. *Macromolecules* **2008**, *41*, 8281−8287.
(60) Rudnick, J.; Gaspari, G. *Science* **1985**, *237*, 384−389.
(61) Creighton, T. E. *Proteins: Structures and molecular properties*: W.H. Freeman and Company: New York, 1993.
(62) Spassov, V. Z.; Yan, L.; Flook, P. K. *Protein Sci.* **2007**, *16*, 494−506.
(63) Eisenberg, D.; Weiss, R. M.; Terwilligaenrd, T. C.; Wolcox, W. *Faraday Symp. Chem. Soc.* **1982**, *17*, 109−120.
(64) Moore, D. S. *Biochem. Educ.* **1985**, *3*, 10−11.
(65) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(66) Gall, T. E.; Romero, P. R.; Cortese, M. S.; Uversky, V. N.; Dunker, A. K. *J. Biomol. Struct. Dyn.* **2007**, *24*, 325−341.
(67) Rose, G. D. *Adv. Protein Chem.*: Unfolded proteins; Academic: San Diego, CA, 2002, 62.
(68) Anfinsen, C. B. *Science* **1961**, *181*, 223−230.
(69) Dyson, H. J.; Wright, P. E.; Harold, A.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 13057−13061.
(70) Berezovsky, I. N. *Phys. Biol.* **2011**, *8*, 035002.
(71) Dunker, A. K.; Brown, C. J.; Obradovic, Z. *Adv. Protein Chem.* **2002**, *62*, 25−49.
(72) Liu, J.; Rost, B. *Nucleic Acids Res.* **2003**, *31*, 3833−3835.
(73) Vucetic, S.; Obradovic, Z.; Vacic, V.; Radivojac, P.; Peng, K.; Iakoucheva, L. M.; Cortese, M. S.; Lawson, J. D.; Brown, C. J.; Sikes, J. G.; et al. *Bioinformatics* **2005**, *21*, 137−140.
(74) Schlessinger, A.; Schaefer, C.; Vicedo, E.; Schmidberger, M.; Punta, M.; Rost, B. *Curr. Opin. Stru. Biol.* **2011**, *21*, 412−418.
(75) Blavatska, V.; Janke, W. *J. Chem. Phys.* **2010**, *133*, 184903.
(76) Uversky, V. N. *Eur. J. Biochem.* **2002**, *256*, 2−12.
(77) Uversky, V. N.; Narizhneva, N. V. *Biochemistry (Moscow)* **1998**, *63*, 420−433.
(78) Tuncbag, N.; Gursoy, A.; Keskin, O. *Phys. Biol.* **2011**, *8*, 035006.
(79) Garbuzynskiy, S. O.; Lobanov, M. Y.; Galzitskaya, O. V. *Protein Sci.* **2004**, *13*, 2871−2877.