RESEARCH PAPER

Intrinsically Disordered Proteins 1, e24848; 2013; © 2013 Landes Bioscience

Functional fragments of disorder in outer membrane β barrel proteins

Kavitha Kurup,[1] A Keith Dunker[2] and Sankaran Krishnaswamy[1,*]

[1]Centre of Excellence in Bioinformatics; School of Biotechnology; Madurai Kamaraj University; Madurai, Tamil Nadu, India; [2]Centre for Computational Biology and Bioinformatics; Indiana University School of Medicine; Indianapolis, IN USA

Keywords: bacterial outer membrane, BtuB protein, intrinsic disorder, OpcA protein, transmembrane protein, functional fragments of disorder, dual personality fragments

Abbreviations: ID, intrinsically disordered; DP, dual personality; ICloop, periplasmic loop; ECloop, extracellular loop; Disin, disorder in intracellular region; Disout, disorder in extracellular region; TM, transmembrane

The traditional view of "sequence–structure–function" has been amended by the discovery of intrinsically disordered proteins. Almost 50% of PDB structures are now known to have one or more regions of disorder, which are involved in diverse functions. These regions typically possess low aromatic content and sequence complexity as well as high net charge and flexibility. In this study, we examined the composition and contribution of intrinsic disorder in outer membrane β barrel protein functions. Our systematic analysis to find the dual personality (DP) fragments, which often function by disorder–order transitions, revealed the presence of 61 DP fragments with 234 residues in β barrel transmembrane protein structures. It was found that though the disorder is more prevalent in the periplasmic regions, most of the residues which undergo disorder–order transitions are found in the extracellular regions. For example, the calcium binding sites in BtuB protein are found to undergo disorder to order transition upon binding calcium. The conformational change in the cell receptor binding site of the OpcA protein, which is important in host cell interactions of *N. meningitidis,* was also found to be due to the disorder–order transitions occurring in the presence of the ligand. The natively disordered nature of DP fragments makes it more appropriate to call them "functional fragments of disorder." The present study provides insight into the roles played by intrinsically disordered regions in outer membrane protein functions.

Introduction

Many regions of proteins or whole protein structures are found to be intrinsically disordered (ID) under native conditions.[1,2] These regions may adopt a non-globular structure or even remain unfolded in solution. While ID regions typically exhibit high sequence variability,[3] some of these regions show sequence conservation,[3,4] and the disordered regions themselves are often conserved.[5,6] Furthermore, these ID regions possess a variety of biological functions[7-9] such as displaying a binding site for associating with another protein[10] or displaying a site for post-translational modification such as phosphorylation[11] or other modifications.[12] The absence of a well-defined structure allows disordered binding sites to interact with several different targets.[13-16] Discovery of protein intrinsic disorder in organisms with different complexities, i.e., from viruses[17] to eukaryotes,[18] have revealed that eukaryotes possess a larger portion of ID compared with bacteria and archaea, while viral proteomes have the widest range of predicted disorder.[19,20] Despite the smaller fraction of ID in prokaryotic proteins, intrinsic disorder has shown to play significant roles in these organisms.[21-23] As the intrinsically disordered regions are mostly studied in regulating and signaling proteins,[24] great interest has been generated in studying their roles in prokaryotic and highly structured proteins.

The prokaryotic transmembrane β barrel proteins (TMBETA) are found in the outer membranes of gram negative bacteria and they execute the function of import or export of molecules across the membrane.[25] The TM (Transmembrane) regions of the β barrel membrane proteins are found to be highly structured within the membrane bilayer.[26] These surface exposed proteins are found to play a critical role in pathogenic processes, such as adherence, motility and colonization of the host cells, formation of channels for the removal of antibiotics as well as the injection of toxins and cellular proteases.[27,28] Due to these functional properties, outer membrane proteins are also attractive targets for the development of antimicrobial drugs and vaccines.[29,30] The existence of intrinsic disorder has been reported recently in these highly structured TMBETA proteins through computational approaches.[31]

Many ID fragments undergo disorder-to-order transitions upon interaction with proteins.[32] In general, such fragments are identified as segments that become structured upon binding to a partner and are called "molecular recognition features" or

*Correspondence to: Sankaran Krishnaswamy; Email: mkukrishna@gmail.com
Submitted: 03/12/13; Revised: 04/28/13; Accepted: 04/28/13
Citation: Kurup K, Dunker AK, Krishnaswamy S. Functional fragments of disorder in outer membrane β barrel proteins. Intrinsically Disordered Proteins 2013; 1:45 - 55; http://dx.doi.org/10.4161/idp.24848

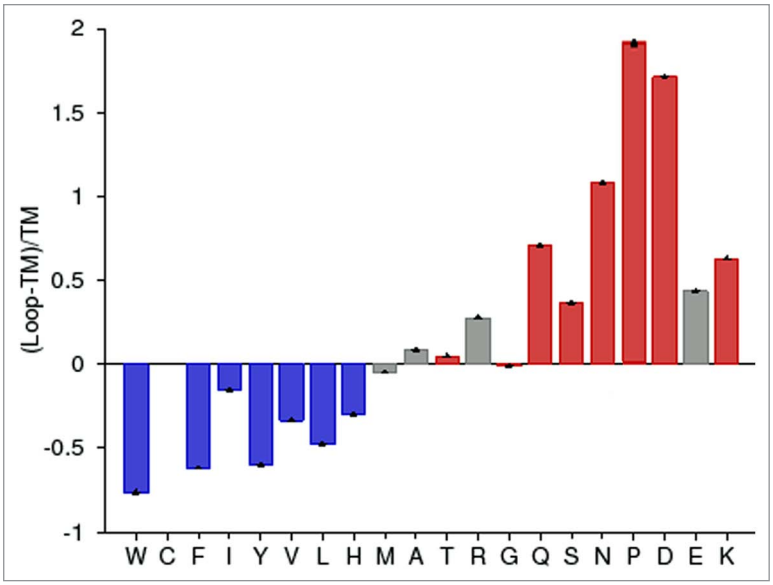

Figure 1. Amino acid profiles are calculated as (Loop-TM)/TM where, Loop is the composition ratio of a given type of amino acid in the loop data set set, and TM is the composition of the same amino acid in the data set being compared, thus giving the fractional difference for each amino acid for the data sets being compared. The X-axis shows the arrangement of the amino acids from the most structure-promoting on the left to the most disorder-promoting on the right.

"MoRF"s.[10] They have also been termed as preformed structural elements (PSEs) and pres-structured motifs (PreSMos) in the literature.[33,34] A second way to identify such fragments was suggested by Zhang et al.,[35] who showed that different crystal structures of the same protein often contain regions that are ID in one crystal but structured in a different crystal. These fragments were referred to as dual personality (DP) fragments because they exist in disordered as well as ordered states.[35] They are mostly enriched next to functional regions (active sites, post-translational modification sites).[34] MoRFs and DP fragments appear to have similar sequence features.[36] If a DP region is involved in partner recognition, it may be termed as a MoRF. In order to understand the functionality of disordered regions in TMBETA proteins, fragments that undergo disorder–order transmission were screened in the present study. The distinct disorder compositions in different regions of TMBETA may also provide a better training set for developing disorder predictors.

Results

Analysis of disordered regions. The compositions of disorder data sets of TMBETA proteins with partially disordered, fully disordered and fully ordered data sets of water soluble globular proteins have earlier been compared.[31] Further, comparison was also made with disorder predictions from different algorithms, as a preliminary step toward checking the need for developing a new disorder predictor for membrane proteins.[31] Distinct amino acid compositions was found in the disordered regions of TMBETA while compared with globular proteins. In our study, the ordered and the different disordered data set from TMBETA proteins were analyzed extensively in order to understand the compositions and functions of intrinsic disorder in these proteins.

The loop regions in TMBETA are found mainly on the periplasmic and extracellular regions. There is an assumption that these "exterior to the bilayer" segments are likely to be more similar to the disordered region than the TM segments.[31] In order to check this, the amino acid compositions of loop regions to that of the reference TM segment data set were compared (**Fig. 1**). It was confirmed that the disorder promoting residues R, Q, S, P, E, K and the other flexible amino acids N and D were highly enriched in the loop data set; whereas, the order promoting residues W, F, I, Y, V, L and H were found to be depleted.

The amino acid compositions of periplasmic loop regions were compared with the reference extracellular loop regions (**Fig. 2A**) to distinguish the composition bias in these data sets. Periplasmic loop regions were found to be enriched in the residues I, V, L, A, Q, and P while extracellular loop data set was enriched in residues W, Y, H, T, R, S, N, D, E, and K, showing distinct amino acid compositions among them.

Subsequently, the compositions of intrinsic disorder present in both periplasmic and extracellular loop regions were compared (**Fig. 2B**). For this, the disorder fragments of these proteins were mapped to the inner (periplasmic) and outer (extracellular) regions. In total, about 699 disordered residues were found to be completely mapped to the periplasmic region. This is also in agreement with an earlier observation that disorder is more frequent in the periplasmic regions in integral membrane proteins.[21] There were only 470 disordered residues positioned in the extracellular regions. Since the extracellular loops are longer than the periplasmic loops/turns, the ratio between the percentages of disordered residues of extracellular to periplasmic loops was calculated and found to be 1:2. The major residues presented in disordered data sets of periplasmic and extracellular regions were L, A, Q and Y, T, R, G, D, respectively.

Clustering of disordered data set with the 3 different parts of TMBETA was performed to clarify and examine the preference in disordered data set (**Fig. 3**). The TM region was found to form a separate cluster apart from the disordered regions. Supporting the composition analysis, **Figure 3** shows the clustering of disorder with loop regions, especially with the periplasmic regions.

Analysis on DP fragments in TMBETA. The initial work on DP by Zhang et al.[35] analyzed 19,858 PDBs and found that 45% of clusters contain DP fragments. They have reported only two clusters of outer membrane proteins in their redundant PDB and none of these clusters were found to possess DP fragments in their structures. Hence, it was interesting to identify the presence of DP fragments in TMBETA and to further characterize their composition and functional roles.

Twenty-four CD-HIT clusters with identical protein chains were found to possess DP fragments in TMBETA. The number of DP residues of length < 10 AA were 147 (62.82%), and

the number of DP residues of length > 10 AA were 87 (37.18%) (**Fig. 4**). Most of the DP fragments were of length between 1–3 AA and the longest DP fragment length found was 18AA. Although an attempt to categorize DP fragments according to their length and to find their specific roles were performed, the small size of the data set and lack of literature makes it difficult to analyze with statistical significance.

The secondary structure of DP fragments in outer membrane proteins were found to be different from that of non-membrane proteins mentioned by Zhang et al.[35] wherein, they show the order of secondary structure possessing DP fragments as loop > turns > helix > β. In the case of TMBETA, a complete absence of turns were observed, to make the order loop > helix > β. This was also different from the number of secondary structures in PDB (helix > sheet > turn > loop) (**Fig. 5**). The number of residues forming C (loop/irregular), H (Helix) and E (Extended β strand) structures were 206 (88.03%), 18 (7.70%) and 10 (4.27%), respectively.

The frequencies of amino acids in the disordered and DP fragment data sets were further compared with that of the background ordered data set and disordered fragments (**Fig. 6**). It was found that the disordered data set in TMBETA are enriched in the residues H, Q, S, P, and E, whereas the ordered data set of TMBETA are enriched in W, F, I, Y, and V. The enrichment of the residue H, which is not seen in the disordered data set of globular proteins, is found to be a remarkable feature of intrinsic disorder in TMBETA proteins. The DP fragment data set was also found to be enriched in residues H and P, which is found to be a subset of residues enriched in disordered data set.

Clustering of amino acids based on the ratio between their relative abundance in order, disorder and DP regions were performed using the GAP program.[37] It was found that DP fragments tend to cluster with the ordered fragments in TMBETA though their properties are clearly shown in between that of the ordered and disordered regions (**Fig. 7**). Likewise, when DP residues and disordered residues were clustered with periplasmic and extracellular loops, disordered residues tend to cluster with periplasmic regions while the DP fragments clustered with the extracellular regions.

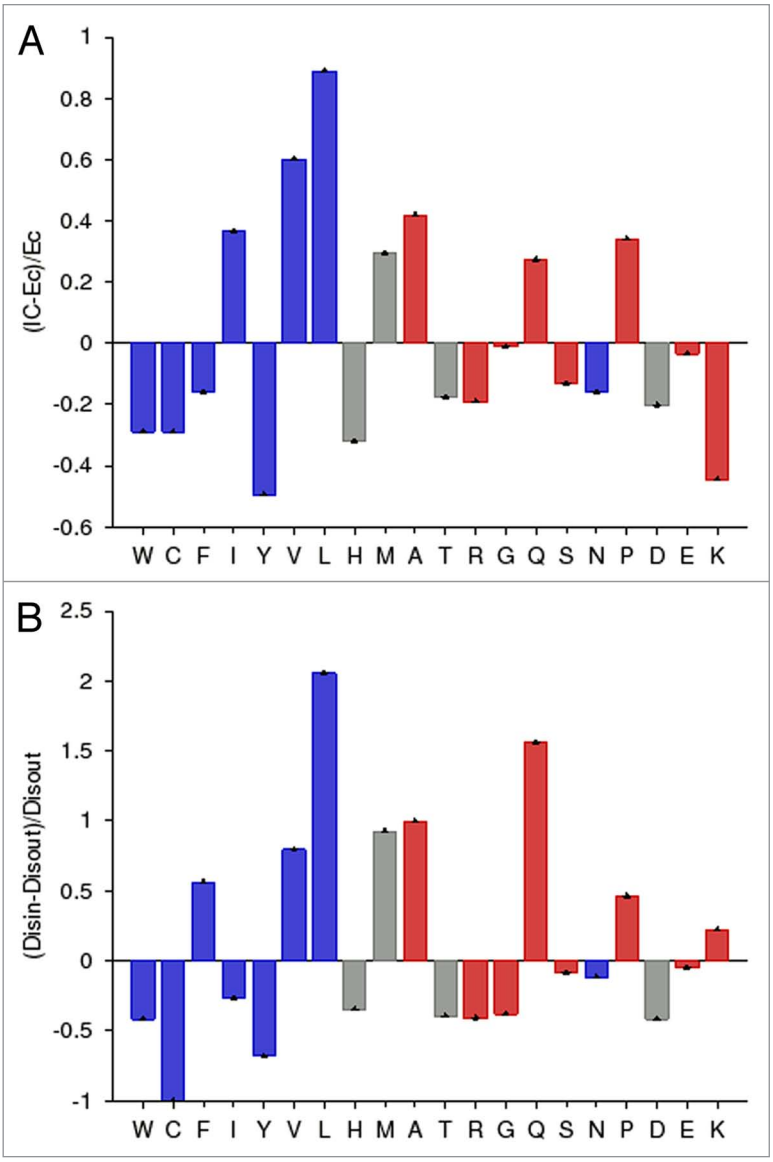

Figure 2. (**A**) The fractional difference between the amino acid frequencies in the periplasmic loop regions (IC) and extracellular loop regions (EC) of TMBETA proteins. The X-axis shows the arrangement of the amino acids from the most structure-promoting on the left to the most disorder-promoting on the right. (**B**) The fractional difference for disordered amino acids in the periplasmic loop regions (disin) and extracellular loop regions (disout) are compared.

The distribution of DP residues and all residues in the periplasmic (IC), extracellular (EC) and TM regions are shown in **Table 1**. The number of DP residues in EC region (138) is most compared with the other regions. The logs-odd ratio is highest for the DP residues in the EC region (0.64 bits). This shows that most of the residues, which undergo disorder–order transitions are present in the extracellular regions.

The observation that most of the residues, which undergo disorder-order transitions are found in the extracellular regions was a motivation to investigate more on the functionality of disordered regions in TMBETA proteins. The Scansite[38] search was performed to find the presence of functional motifs onto which disorder regions can be mapped. Not a single region was mapped to the current motifs in Scansite possibly indicating unique functions of these regions in the membrane proteins. Two examples of protein intrinsic disorder function we could observe in TMBETA are explained in detail further.

DP fragments in BtuB protein. Calcium-binding sites in proteins play major roles in stabilizing protein structures.[39] It is known that, in the absence of ligands, binding sites in the apo state are often found to be in disordered state.[40] Here, we have used 2 identical structures of spin-labeled BtuB V10R, one with bound calcium and the other in the apostate to show the disorder-to-order transitions of this protein upon binding to calcium.

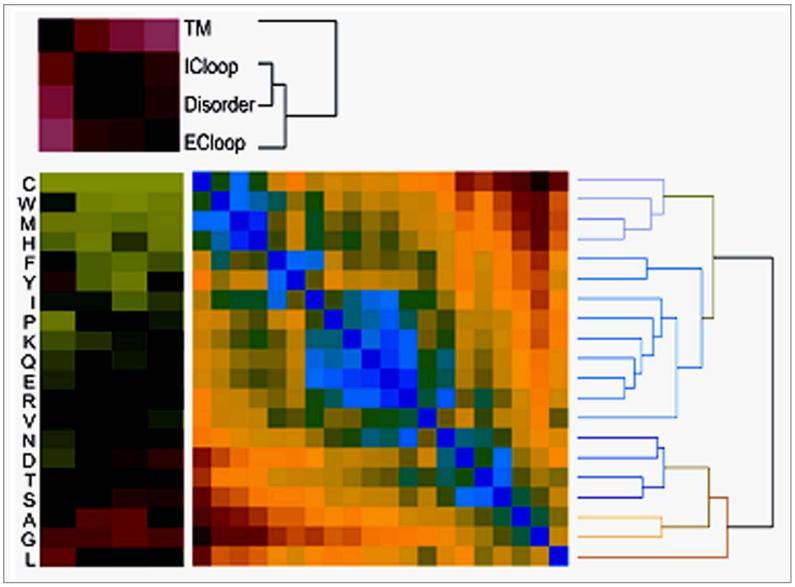

Figure 3. Clustering of amino acids in the TM (Trans membrane), Icloop (periplasmic loop), Ecloop (Extracellular loop) and disorder data sets of TMBETA based on their frequency shows the grouping of disorder data set to the Icloop data set. The topmost matrix is the column proximity matrix. Just below it is the raw data matrix and the right one is row proximity matrix. The raw data matrix together with 2 proximity matrices are projected through color spectrums to create matrix maps. The raw data matrix uses green-black-red color scheme to show low to high frequency of amino acids in the 4 data sets used. The distance of the amino acids in the 4 data sets used are shown in row proximity matrix (color spectra - Rainbow 130).

The crystal structure of spin-labeled BtuB V10R1 with bound calcium and cyanocobalamin vitamin (3M8D) have fragments, 178–195, 229–240 and 278–287, in ordered state (**Fig. 8A**). These fragments are disordered in the crystal structure of spin-labeled BtuB V10R1 in the apostate (3M8B). The two extracellular loops, 178–195 and 229–240, play important roles in signal transduction after the substrate binding.[41] The calcium binding site in these loops was formed by 4 negatively charged residues: D178, D193, D195 and D230. The B12 binding site includes the top of the hatch domain and parts of these loops in this structure. With the bound calcium ions, near the loops 178–195 and 229–240, the motion of these loops are found to be more correlated with the motion of hatch domain (**Fig. 8B**). Therefore, calcium binding to these loops has stabilized the substrate binding site, which increases the binding affinity to B12. The enhanced correlation between these loops and the hatch domain also indicates stiffening of the protein structure, which may promote signal transduction through the hatch domain.[41]

As the well-conserved DP fragments across different species are biologically meaningful, the conservation of DP fragments in orthologous proteins of the *Escherichia coli* BtuB was investigated. It was found that, aspartate, the key functional residue in the BtuB DP fragments (**Fig. 9**), is highly conserved in orthologous proteins.

DP fragments in OpcA protein. OpcA, the outer membrane adhesin protein from *Neisseria meningitidis* causes meningitis and septicimia in humans.[42] The interactions of OpcA with the proteoglycan surface receptors of the mammalian epithelial cells promote the binding of *N. meningitidis*. The understanding on the structural levels of adhesion to the host cells by *N. meningitidis* has been made possible by the crystal structure of OpcA (PDB ID: 1K24).[42] This structure was solved using crystals of OpcA grown in surfactant based vapor diffusion method, with inclusion of zinc ions, which was mandatory for crystal formation. The crystal structure revealed that the 2 zinc ions are bound to the extracellular loop regions. The second structure of OpcA (PDB ID: 2VDF) was solved using in meso method which is different from the in surfo method used in the original structure.[43] The 2 crystal structures clearly demonstrate the high motion of extracellular loops, which may play a role in ligand binding by induced fit. While investigating the disordered residues in the original in surfo crystal structure of OpcA, apart from the residues 1–4 in the N-terminal, no disordered residues were identified. In the meso structure of OpcA, the residues A1 - T6, K124 - T129, G174 - K180, E203 - S204 and V226 - S232 were found to be disordered. This provided the DP fragment data of the 1K24–2VDF cluster using CD-HIT. The topology of these DP fragments is given in **Table 2**.

The loops L3, L4 and L5, which exhibit disorder (in 2VDF) to order (in 1K24) transitions are found to play significant roles in the formation of crevice, which is the proposed location of heparin/proteoglycan binding site (**Fig. 10A**). The presence of the zinc ion, which acted as a ligand, have made the disordered loops of the in meso structure to become ordered in the in surfo crystal structure of OpcA (**Fig. 10B**). The same disorder–order transition of loops may occur in the presence of proteoglycan ligands. Cherezov et al.[43] has reported that information on the disparities between the two structures at the loops L3, L4, and L5 are not available due to the weak and untraceable densities in the in meso maps of those regions. The present study suggests that the presence of the ligand is responsible for the formation of the binding pocket mainly by the disorder–order transitions of L3, L4, and L5 and the closing of pocket by L2. The conservation studies of these DP fragments as performed in Btub were not possible in the case of OpcA due to the lack of orthologous proteins of OpcA.

Disordered residue classification. In this study, it is noticed that though the DP fragments have properties in between those of the ordered and disordered residues, natively they are found in the disordered state. The natively disordered nature of DP fragments makes it more appropriate to call them "functional fragments of disorder." It should also be noted at this point that not only the DP fragments, but also the MoRFs, SLiMs and even linker regions that do not undergo induced folding may be called as functional fragments of disorder. As the DP residues are considered as natively disordered residues, the disordered residues can be classified into three categories: (1) disordered residues that are only found in one structure, (2) disordered residues conserved in two or more structures and (3) the DP residues. In TMBETA, the residues which are always disordered were found mostly in

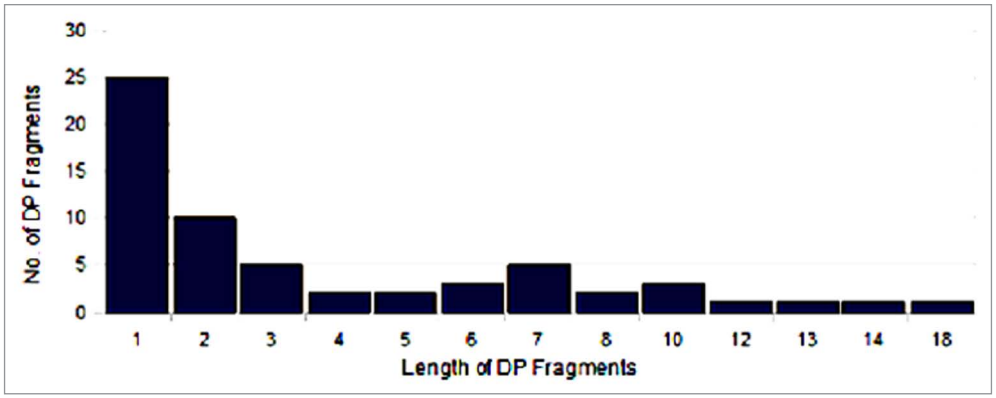

Figure 4. The distribution of DP fragments with varying amino acid lengths. The x axis denotes the length of DP fragments, and the y axis is the number of DP fragments with given length. All DP fragments has shown length below 20 amino acids.

single structures. The number of disordered residues conserved among 2 or more structures was found to decrease as the number of structures increases (**Table 3**).

Discussion

Although several computational studies have shown the significance of disorder in transmembrane α-helical proteins,[21,31] the occurrence of disorder in TMBETA proteins and a comparison of their characteristics to the structured and disordered regions of globular proteins is reported only recently.[31] Therefore, the present study aims to give a more detailed vision of intrinsic disorder in TMBETA proteins.

During the disordered data set preparation, it was observed that the number of PDB structures for TMBETA is less in number and the disorder fraction in the TMBETA is even lesser in these PDBs. So the first attempt made was to include the maximum number of PDB chains for the study. Clustering PDBs with 100% sequence identity and also taking sequence similar, but structurally dissimilar PDBs helped to increase the disorder data set.

The amino acid composition analysis has shown that disordered residues are more prevalent in the loop regions of TMBETA. Hence, the question was which loop region, whether extracellular or periplasmic loops display more disordered regions. We observed that, although the extracellular loops are longer in TMBETA the disordered content was found to be significantly higher in periplasmic loops of TMBETA proteins. This was also supported by clustering analysis showing clustering of disorder with periplasmic data set. It should be noted that, though disorder is more prevalent on periplasmic region, the proposed "functional fragments of disorder" is exhibited more by the extracellular region. The disordered state of the longer extracellular loops, which are continuously exposed to the environment, may allow more flexibility to capture specific ligands and they become ordered while ligand binding. The conformational changes occurring in the extracellular loop regions may also allow the pore opening for the entry of molecules.[44] A recent study has shown that in CorA closed channel X-ray structures, the periplasmic loops are largely disordered. Their structural data suggest that the periplasmic loops have prominent role in stabilization of open conformation of CorA channels.[45] Likewise, the more disordered residues observed in periplasmic regions in our study may be an attribute of such functions. The lack of right partners for binding may also be the reason for the inability for all the disordered regions to exhibit disorder-to-order transitions.[36] The small size of the data

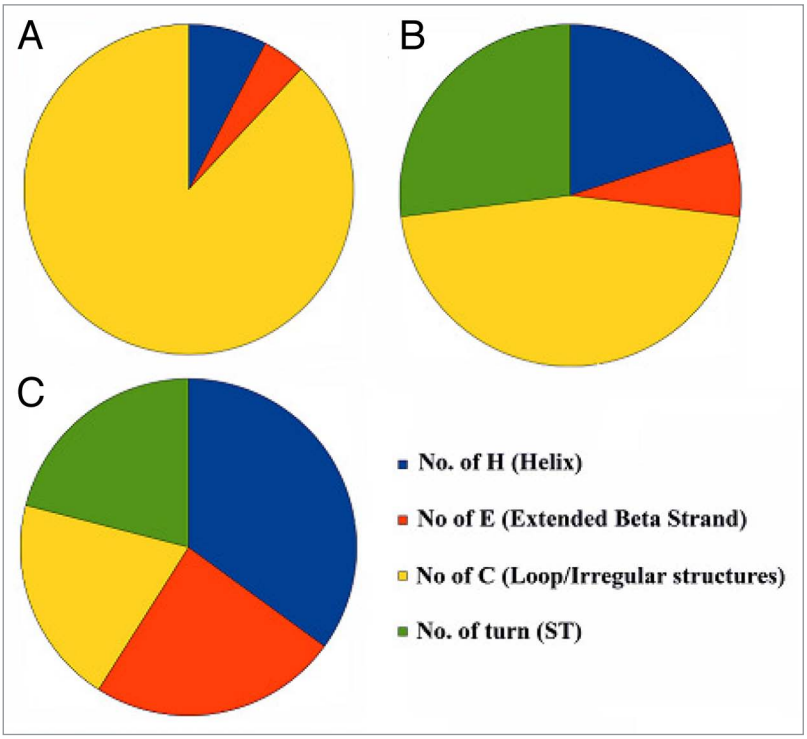

Figure 5. Secondary structure distribution of DP fragments in (**A**) outer membrane proteins, (**B**) non-membrane proteins and (**C**) in PDB.

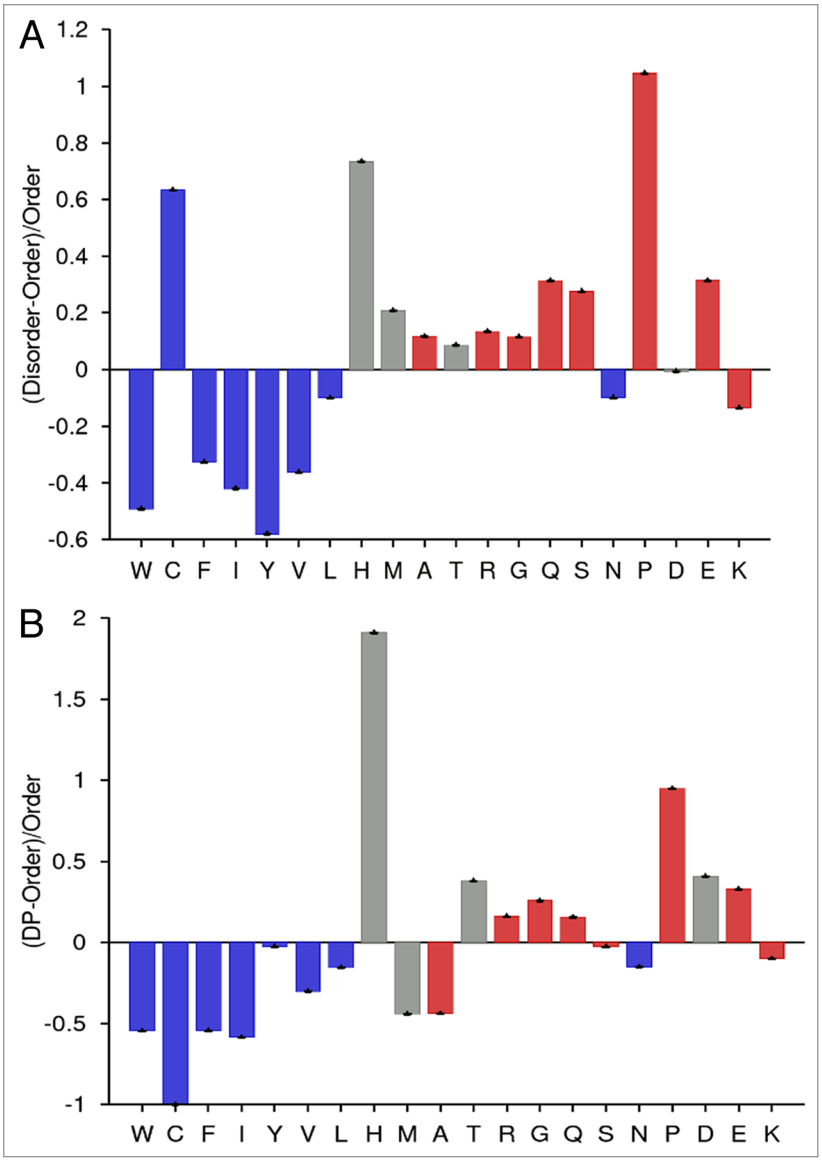

Figure 6. The fractional difference between the amino acid frequencies of (**A**) disordered data set and (**B**) DP data set to that of the reference ordered data set.

set available currently has made it difficult to do more detailed investigation regarding this aspect.

The current study demonstrates the importance of intrinsically disordered regions of TMBETA proteins, in making interactions with the external environment, facilitating binding of molecules from the environment for their smooth transport across the membrane and into the cell. The conformational dynamics of OpcA crystal structure when transplanted to the bilayer environment was found to be due to the presence of disordered residues in the external loops. These residues may be helping in the interactions of OpcA with different molecules like heparin, vitronectin, fibronectin, integrins, etc.[46,47] It is already shown that binding sites for heparin and integrins are found in disordered sequences of extracellular proteins.[48] The native disorder in these loop regions may also provide space to fit different sizes of the ligands, while forming the binding pocket. The Opc protein is already in use, as a surface exposed immunogenic adhesin of *N. meningitidis* present in an experimental outer membrane vesicle vaccine used for vaccination against serogroup B disease.[49] The transient, weak and specific interactions of the disorder-to-order transitions has opened new avenue to be explored via the drug discovery process. Either the drugs mimicking the critical regions of disordered partner compete with the binding site on structured partner or drugs directly targeting disordered regions can be developed.[50] As the OpcA is an important adhesin particularly in meningitis in humans caused by *N. meningitides*,[51] drugs which target OpcA by exploiting the disordered interface can be further investigated.

Disorder in globular proteins has been earlier shown to be strongly associated with signaling, by providing binding sites for other molecules.[52] In the present study, while looking at the disorder in highly structured TMBETA proteins which exhibits disorder-order transitions, we obtained the properties and functions of disorder that are retained for this type of protein. The distinct compositions of disordered residues in different regions of TMBETA point toward the need for developing different predictors for the prediction of disorder in these proteins. The proposed "functional fragments of disorder" were found to be enriched in residues, which are subsets of enriched residues in the disordered data set. The clustering of functional fragments of disorder with ordered data set shows their tendency to become ordered than to be in their native disordered state. The example of BtuB examined here illustrates the importance of intrinsic disorder in cobalamin transport across the outer membrane. The importance of intrinsic disorder shown in host-pathogen interactions of OpcA protein may open a new way to use the intrinsic disorder in membrane proteins to target candidate drugs.

Materials and Methods

Disordered data set. The 209 structures of TMBETA utilized in this study were retrieved from the PDB of transmembrane proteins (PDBTM)[53] as of November 2011. This data set was filtered to obtain the primary data set as following: after excluding structures other than X-ray, 194 PDBs were used and the structures with resolution worse than 3 Å were discarded. In the next step, CD-HIT[54] was used to cluster the remaining 158 PDBs at a 100% identity cut off. Totally, 116 clusters with same number of representative chains were obtained. From this, 2 chains with < 50 AA length were removed. Again, repeated chains were also removed in different clusters, which reduced the data set to 109 chains.

The software XML2PDB[55] was used to identify the disordered residues in PDB by searching for residues with missing coordinates. Out of the 109 PDBs, only 70 chains have shown to possess at least one disordered residue. As the cut off identity used is 100%, some chains in different clusters had similar disordered regions, identified using sequence similarity. Such chains were removed, to keep only 1 representative chain. Further, 4 chains with < 3 disordered residues were also removed from the set. The N-terminal His tags and N-Met, treatment of which is not very consistent in protein models, were removed. Thus the primary data set contained a total of 48 chains.

Adding more chains to the primary data set. Although the chains are 100% identical sequentially, there is a chance of structural dissimilarity among them. Therefore, as a further step to increase the number of chains, structural comparison was performed. Out of 115 clusters obtained using CD-HIT, 68 clusters have more than one identical chain. Disorder was identified in each chain in these 68 clusters. Eighteen clusters that did not show any disorder were excluded from data set preparation. Hence, 50 clusters were available in total. Out of these 50 clusters, 31 were removed because each of them had chains from the same structure. So, finally 19 clusters remained for structural comparison.

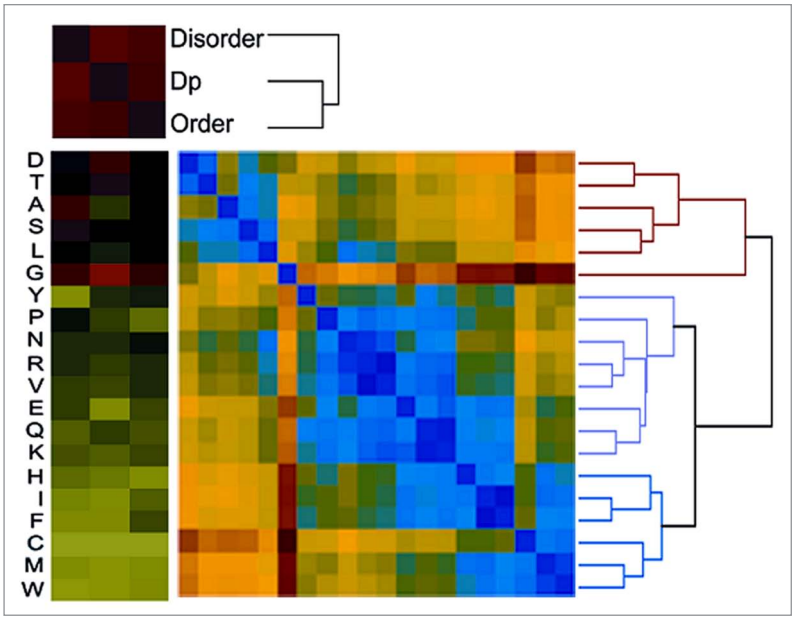

Figure 7. Clustering of DP fragments with order and disorder shown using GAP program. The topmost one is the column proximity matrix. Just below it is the raw data matrix and the right one is row proximity matrix. The raw data matrix together with two proximity matrices are projected through color spectrums to create matrix maps. The raw data matrix uses green-black-red color scheme to show low to high frequency of 20 amino acids in the three data sets used. The distance of the amino acids in the three data sets used are shown in row proximity matrix (color spectra - Rainbow 130).

In order to obtain the structural identity of these sequentially identical chains, pairwise structural comparisons were performed for these chains in each cluster using the SUPERPOSE server.[56] Out of these 19 clusters, 5 clusters were found to be structurally dissimilar with an RMSD > 3 Å.

Taking the 5 structurally different clusters, the mean agreement of disordered residues between the pairs of proteins in each cluster was calculated using the following equation:

$$Od(A, B) = 0.5(Ad\cap Bd/AD + Ad\cap Bd/BD)$$

Where, Od (A, B) is the overlap between disorder in chain A and chain B. AD and BD are the total number of disorder in A and B chain respectively and Ad∩Bd is the common agreement of disorder between A chain and B chain. "A" must have at least one disordered region while "B" can be fully ordered.[57]

By applying the above criteria, out of 5 clusters, 2 clusters were shown to have mean agreement of disorder < 70%. Between the structures, 2HDFA:2HDIA, the mean fraction was 0.62 and for 1NQEA:2GUFA, it was 0.595. Therefore, the 2 chains 2HDIA and 2GUFA are included in the data set. Thus after doing structural comparisons, the data set increased to 50 chains. The final disordered data set of TMBETA with 1459 residues was made by filtering out the dual personality fragments and ordered fragments from these 50 chains.

Ordered data set. There were 39 fully ordered chains in TMBETA in the initial data set of 109 PDBs. These chains (after removing His tags) were clustered at 30% identity using CD-HIT to get the final fully ordered data set of 16 chains with 4768 residues.

Table 1. Distribution of residues in extracellular (EC), periplasmic (IC), transmembrane (TM) regions in the transmembrane β data set

	EC	IC	TM	Total
DP residues	**138**	90	6	234
All residues	**8043**	6224	7046	21313
Fraction DP f_{DP}	**0.59**	0.38	0.03	1.0
Fraction all residues f_{all}	**0.38**	0.29	0.33	1.0
Odds ratio (f_{DP}/f_{all})	**1.56**	1.32	0.08	-
Logs-odd (bits)	**0.644**	0.397	–3.689	-

Data set of dual personality fragments. Out of 115 clusters in our primary data set of TMBETA, 68 clusters that have two or more identical sequences were used to find DP fragments in this study. The residues that are found to be present at least in one protein but absent in other proteins of the same cluster were recognized as DP fragments. Out of 68 clusters, 24 were shown to have DP fragments. The data set totally includes 61 DP fragments with 234 residues.

Composition analysis. The number of amino acids and their frequencies for each data set were calculated using the PIR composition calculation.[58] Composition Profiler[59] was used to find the statistically significant patterns of amino acid enrichment or depletion between the different sets of protein sequences. In this analysis, the fractional differences between the given protein set composition to that of the reference protein set was calculated

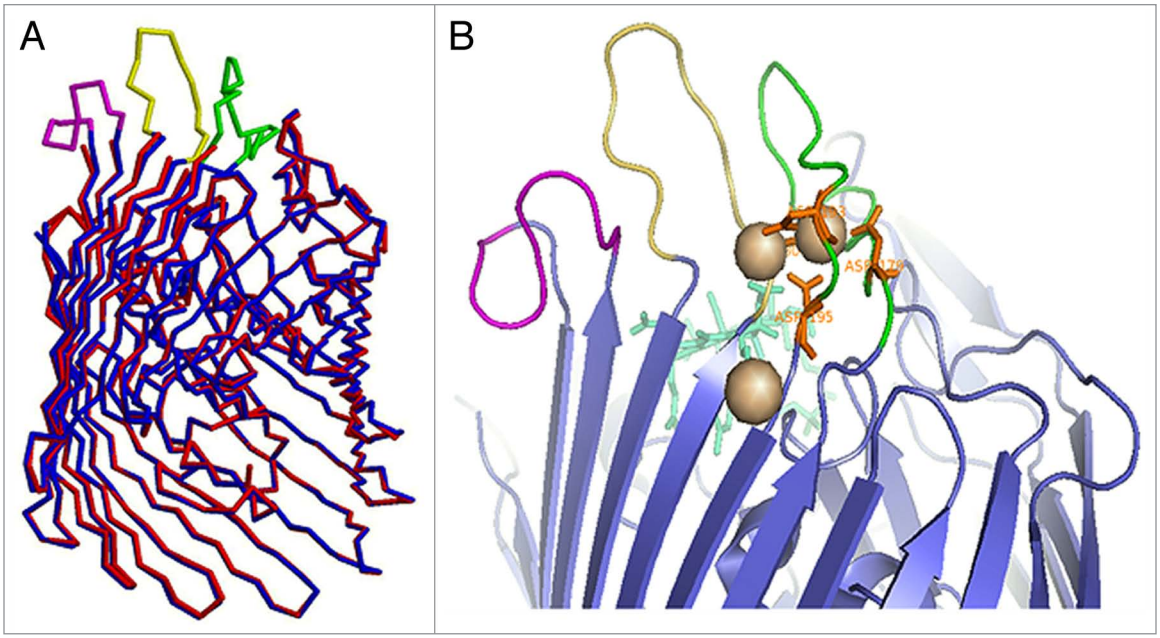

Figure 8. (**A**) The superimposed backbone structures of 3M8D (Blue–Ligand bound state) and 3M8B (Red–Apostate). 3M8D shows fragments 178–195 (green), 229–240 (yellow), 278–287 (magenta) in ordered state. These fragments are disordered and so cannot be seen in the crystal structure of spin-labeled BtuB V10R1 in the apostate (3M8B). (**B**) Figure shows the zoom-in of ligand bound state: B12 (cyan) and Ca^{2+} (spheres) bound to the four aspartate residues (orange) in 3M8D.

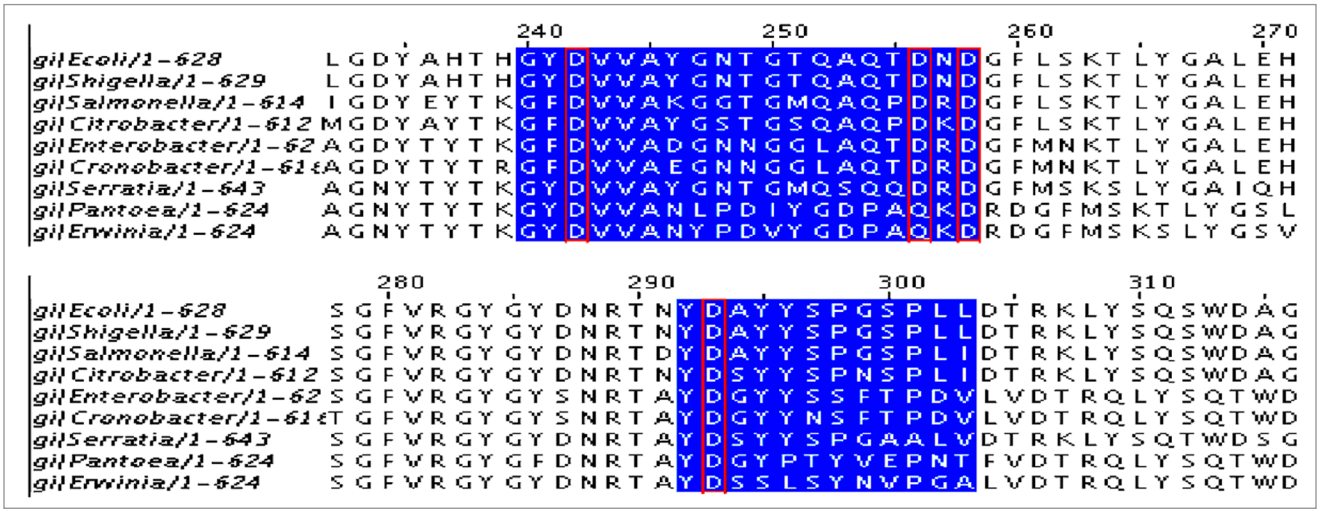

Figure 9. The orthologous proteins of BtuB from different organisms are aligned and the conserved DP fragments are shown in blue background. The highly conserved aspartate residues are shown in red boxes.

Table 2. Topology of DP fragments in OpcA

DP fragments	Topology
Q^{5}-T^{6}	N-term
K^{124}-T^{129}	Loop-3
G^{174}-K^{180}	Loop-4
E^{203}-S^{204}	Turn-4
V^{226}-S^{232}	Loop-5

and then the amino acids were arranged in the order of their flexibility. Ten thousand bootstrap iterations were performed with a significant value of 0.05 and the color scheme used to discriminate amino acids show their disorder propensity. Disorder-promoting residues are colored red, order-promoting residues are colored blue, and disorder-order neutral residues are colored gray. The positive and negative values indicate amino acids that are enriched and depleted respectively in the given data sets compared with the reference data set. Based on the P value table produced from composition profiler, the residues more enriched

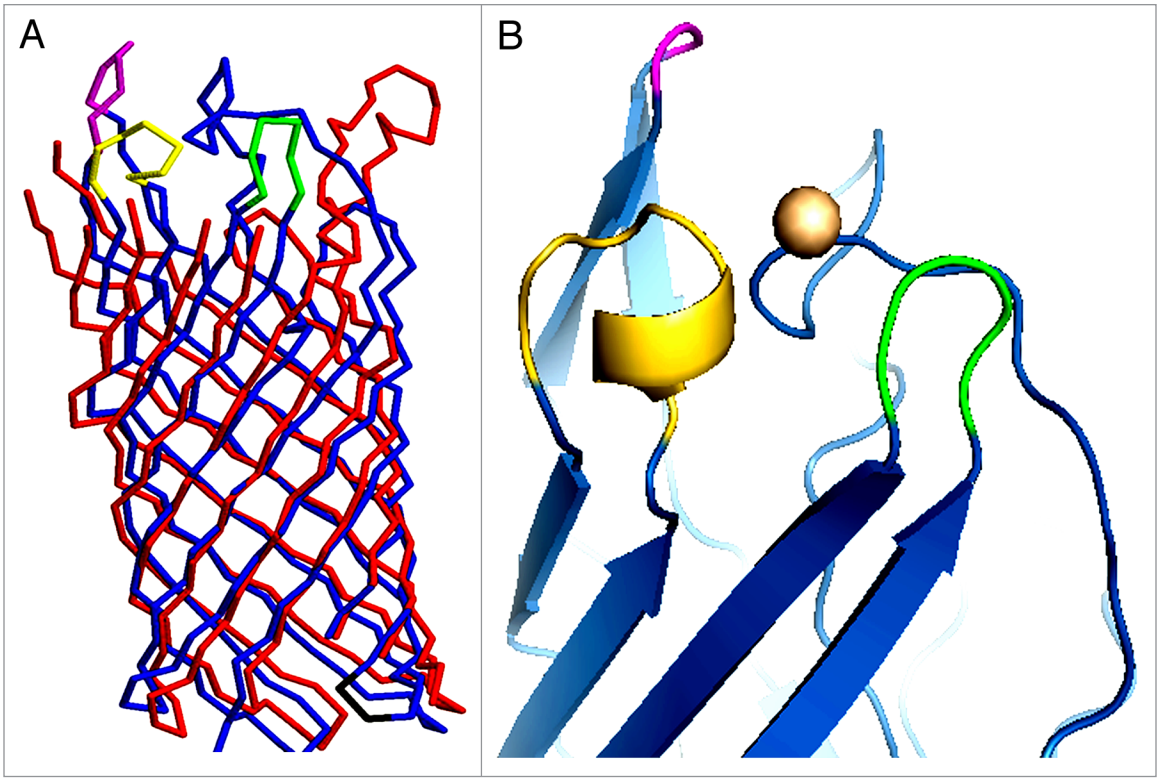

Figure 10. (**A**) The superimposed backbone structures of 1K24 (Blue–Ligand bound state) and 2VDF (Red–Apostate). The DP fragments 124–129 (green), 174–180 (yellow), 226–232 (magenta) correspond to L3, L4, L5 and the DP fragment 203–204 (black) correspond to T4 in 1K24. (**B**) Figure shows the zoom-in of the ligand bound state: ordered loops L3, L4 and L5 in the presence of zinc ion (sphere).

Table 3. Disordered residue classification

Always disordered in:	Always disordered in:			DP residues
1 structure	2 structures	3 structures	> 3 structures	234
1,003 residues	283 residues	139 residues	34 residues	

and depleted in the data sets analyzed were determined. Scansite search[38] was performed to analyze whether any DP fragment is acting as functional signaling motif. Scansite identifies short protein sequence motifs that are recognized by different signaling proteins.

Clustering of amino acids. Based on the distribution of each amino acid in structured, periplasmic and extracellular regions, their preferences were found by using Generalized Association Plot (GAP) program for clustering.[37] This clustering was also performed using disordered, ordered and DP fragments. The input data are the single amino acid distributions on different categories of fragments. For example, AA-order, AA-disorder, AA-DP and the pair-wised score of each amino acid were the euclidean distances of their 3-dimensional vectors defined by the single amino acid distributions. There are 3 major pieces of information contained in any multivariate data set with n subjects and p variables: (1) the linkage among n subject points in the p-dimensional space, (2) the linkage between p variable vectors in the n-dimensional space, and (3) the interaction linkage between the sets of subjects and variables.

Sequence analysis. The NCBI-microbial BLAST program[60] was utilized to find the orthologous proteins from different species. The proteins from selected genomes were used for multiple sequence alignment using ClustalW.[61] The Jalview sequence editor was used to identify the conserved residues from the alignment.[62]

Structural analysis. Superpose server,[56] was used to perform pairwise superimposing of structures. To view the superimposed structures to distinguish disordered and DP fragments, PYMOL molecular viewer was utilized.[63]

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We acknowledge the Department of Biotechnology, Government of India, for supporting Centre of Excellence in Bioinformatics, MKU.

References

1. Romero P, Obradovic Z, Kissinger CR, Villafranca JE, Garner E, Guilliot S, et al. Thousands of proteins likely to have long disordered regions. Pac Symp Biocomput 1998; 3:437-48; PMID:9697202
2. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 1999; 293:321-31; PMID:10550212; http://dx.doi.org/10.1006/jmbi.1999.3110
3. Brown CJ, Takayama S, Campen AM, Vise P, Marshall TW, Oldfield CJ, et al. Evolutionary rate heterogeneity in proteins with long disordered regions. J Mol Evol 2002; 55:104-10; PMID:12165847; http://dx.doi.org/10.1007/s00239-001-2309-6
4. Chen JW, Romero P, Uversky VN, Dunker AK. Conservation of intrinsic disorder in protein domains and families: II. functions of conserved disorder. J Proteome Res 2006; 5:888-98; PMID:16602696; http://dx.doi.org/10.1021/pr060049p
5. Toth-Petroczy MB, Simon I, Dunker AK, Uversky VN, Fuxreiter M. Assessing conservation of disordered regions in proteins. The Open Proteomics j 2008; 1:46-53; http://dx.doi.org/10.2174/1875039700801010046.
6. Bellay J, Han S, Michaut M, Kim T, Costanzo M, Andrews BJ, et al. Bringing order to protein disorder through comparative genomics and genetic interactions. Genome Biol 2011; 12:R14; PMID:21324131; http://dx.doi.org/10.1186/gb-2011-12-2-r14
7. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z. Intrinsic disorder and protein function. Biochemistry 2002; 41:6573-82; PMID:12022860; http://dx.doi.org/10.1021/bi012159+
8. Tompa P. Intrinsically unstructured proteins. Trends Biochem Sci 2002; 27:527-33; PMID:12368089; http://dx.doi.org/10.1016/S0968-0004(02)02169-2
9. Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Uversky VN, et al. Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. J Proteome Res 2007; 6:1882-98; PMID:17391014; http://dx.doi.org/10.1021/pr060392u
10. Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, et al. Analysis of molecular recognition features (MoRFs). J Mol Biol 2006; 362:1043-59; PMID:16935303; http://dx.doi.org/10.1016/j.jmb.2006.07.087
11. Gao J, Thelen JJ, Dunker AK, Xu D. Musite, a tool for global prediction of general and kinase-specific phosphorylation sites. Mol Cell Proteomics 2010; 9:2586-600; PMID:20702892; http://dx.doi.org/10.1074/mcp.M110.001388
12. Gao J, Dong X. Correlation between posttranslational modification and intrinsic disorder in proteins. Pac Symp Biocomput 2012; 17:94-103
13. Jones DT, Ward JJ. Prediction of disordered regions in proteins from position specific score matrices. Proteins 2003; 53(Suppl 6):573-8; PMID:14579348; http://dx.doi.org/10.1002/prot.10528
14. Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, Dunker AK. Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. BMC Genomics 2008; 9(Suppl 1):S1; PMID:18366598; http://dx.doi.org/10.1186/1471-2164-9-S1-S1
15. Hsu WL, Oldfield C, Meng J, Huang F, Xue B, Uversky VN, et al. Intrinsic protein disorder and protein-protein interactions. Pac Symp Biocomput 2012; 17:116-27; PMID:22174268
16. Hsu WL, Oldfield CJ, Xue B, Meng J, Huang F, Romero P, et al. Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding. Protein Sci 2013; 22:258-73; PMID:23233352; http://dx.doi.org/10.1002/pro.2207
17. Xue B, Williams RW, Oldfield CJ, Goh GK, Dunker AK, Uversky VN. Viral disorder or disordered viruses: do viral proteins possess unique features? Protein Pept Lett 2010; 17:932-51; PMID:20450483; http://dx.doi.org/10.2174/092986610791498984
18. Pancsa R, Tompa P. Structural disorder in eukaryotes. PLoS One 2012; 7:e34687; PMID:22496841; http://dx.doi.org/10.1371/journal.pone.0034687
19. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol 2004; 337:635-45; PMID:15019783; http://dx.doi.org/10.1016/j.jmb.2004.02.002
20. Xue B, Dunker AK, Uversky VN. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. J Biomol Struct Dyn 2012; 30:137-49; PMID:22702725; http://dx.doi.org/10.1080/07391102.2012.675145
21. Minezaki Y, Homma K, Nishikawa K. Intrinsically disordered regions of human plasma membrane proteins preferentially occur in the cytoplasmic segment. J Mol Biol 2007; 368:902-13; PMID:17368479; http://dx.doi.org/10.1016/j.jmb.2007.02.033
22. Tompa P, Dosztanyi Z, Simon I. Prevalent structural disorder in E. coli and S. cerevisiae proteomes. J Proteome Res 2006; 5:1996-2000; PMID:16889422; http://dx.doi.org/10.1021/pr0600881
23. Pavlovic-Lažetic GM, Mitic NS, Kovacevic JJ, Obradovic Z, Malkov SN, Beljanski MV. Bioinformatics analysis of disordered proteins in prokaryotes. BMC Bioinformatics 2011; 12:66; PMID:21366926; http://dx.doi.org/10.1186/1471-2105-12-66
24. Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol 2002; 323:573-84; PMID:12381310; http://dx.doi.org/10.1016/S0022-2836(02)00969-5
25. Schleiff E, Soll J. Membrane protein insertion: mixing eukaryotic and prokaryotic concepts. EMBO Rep 2005; 6:1023-7; PMID:16264426; http://dx.doi.org/10.1038/sj.embor.7400563
26. Rosenbusch JP. Characterization of the major envelope protein from Escherichia coli. Regular arrangement on the peptidoglycan and unusual dodecyl sulfate binding. J Biol Chem 1974; 249:8019-29; PMID:4609976
27. Cordwell SJ. Technologies for bacterial surface proteomics. Curr Opin Microbiol 2006; 9:320-9; PMID:16679049; http://dx.doi.org/10.1016/j.mib.2006.04.008
28. Bina JE, Provenzano D, Wang C, Bina XR, Mekalanos JJ. Characterization of the Vibrio cholerae vexAB and vexCD efflux systems. Arch Microbiol 2006; 186:171-81; PMID:16804679; http://dx.doi.org/10.1007/s00203-006-0133-5
29. Grandi G. Antibacterial vaccine design using genomics and proteomics. Trends Biotechnol 2001; 19:181-8; PMID:11301131; http://dx.doi.org/10.1016/S0167-7799(01)01600-6
30. Bernardini G, Braconi D, Martelli P, Santucci A. Postgenomics of Neisseria meningitidis for vaccines development. Expert Rev Proteomics 2007; 4:667-77; PMID:17941821; http://dx.doi.org/10.1586/14789450.4.5.667
31. Xue B, Li L, Meroueh SO, Uversky VN, Dunker AK. Analysis of structured and intrinsically disordered regions of transmembrane proteins. Mol Biosyst 2009; 5:1688-702; PMID:19585006; http://dx.doi.org/10.1039/b905913j
32. Hyre DE, Klevit RE. A disorder-to-order transition coupled to DNA binding in the essential zinc-finger DNA-binding domain of yeast ADR1. J Mol Biol 1998; 279:929-43; PMID:9642072; http://dx.doi.org/10.1006/jmbi.1998.1811
33. Fuxreiter M, Simon I, Friedrich P, Tompa P. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. J Mol Biol 2004; 338:1015-26; PMID:15111064; http://dx.doi.org/10.1016/j.jmb.2004.03.017
34. Lee SH, Kim DH, Han JJ, Cha EJ, Lim JE, Cho YJ, et al. Understanding pre-structured motifs (PreSMos) in intrinsically unfolded proteins. Curr Protein Pept Sci 2012; 13:34-54; PMID:22044148; http://dx.doi.org/10.2174/138920312799277974
35. Zhang Y, Stec B, Godzik A. Between order and disorder in protein structures: analysis of "dual personality" fragments in proteins. Structure 2007; 15:1141-7; PMID:17850753; http://dx.doi.org/10.1016/j.str.2007.07.012
36. Dunker AK. Another window into disordered protein function. Structure 2007; 15:1026-8; PMID:17850741; http://dx.doi.org/10.1016/j.str.2007.08.001
37. Wu HM, Tien YJ, Chen CH. GAP: a graphical environment for matrix visualization and cluster analysis. Comput Stat Data Anal 2010; 54:767-78; http://dx.doi.org/10.1016/j.csda.2008.09.029
38. Obenauer JC, Cantley LC, Yaffe MB. Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. Nucleic Acids Res 2003; 31:3635-41; PMID:12824383; http://dx.doi.org/10.1093/nar/gkg584
39. Kretsinger RH. Calcium-binding proteins. Annu Rev Biochem 1976; 45:239-66; PMID:134666; http://dx.doi.org/10.1146/annurev.bi.45.070176.001323
40. Liu T, Altman RB. Prediction of calcium-binding sites by combining loop-modeling with machine learning. BMC Struct Biol 2009; 9:72; PMID:20003365; http://dx.doi.org/10.1186/1472-6807-9-72
41. Luan B, Carr R, Caffrey M, Aksimentiev A. The effect of calcium on the conformation of cobalamin transporter BtuB. Proteins 2010; 78:1153-62; PMID:19927326; http://dx.doi.org/10.1002/prot.22635
42. Prince SM, Achtman M, Derrick JP. Crystal structure of the OpcA integral membrane adhesin from Neisseria meningitidis. Proc Natl Acad Sci U S A 2002; 99:3417-21; PMID:11891340; http://dx.doi.org/10.1073/pnas.062630899
43. Cherezov V, Liu W, Derrick JP, Luan B, Aksimentiev A, Katritch V, et al. In meso crystal structure and docking simulations suggest an alternative proteoglycan binding site in the OpcA outer membrane adhesin. Proteins 2008; 71:24-34; PMID:18076035; http://dx.doi.org/10.1002/prot.21841
44. Bond PJ, Derrick JP, Sansom MS. Membrane simulations of OpcA: gating in the loops? Biophys J 2007; 92:L23-5; PMID:17114231; http://dx.doi.org/10.1529/biophysj.106.097311
45. Palombo I, Daley DO, Rapp M. The periplasmic loop provides stability to the open state of the CorA magnesium channel. J Biol Chem 2012; 287:27547-55; PMID:22722933; http://dx.doi.org/10.1074/jbc.M112.371484
46. Moore J, Bailey SE, Benmechernene Z, Tzitzilonis C, Griffiths NJ, Virji M, et al. Recognition of saccharides by the OpcA, OpaD, and OpaB outer membrane proteins from Neisseria meningitidis. J Biol Chem 2005; 280:31489-97; PMID:16006553; http://dx.doi.org/10.1074/jbc.M506354200
47. Hubert K, Pawlik MC, Claus H, Jarva H, Meri S, Vogel U. Opc expression, LPS immunotype switch and pilin conversion contribute to serum resistance of unencapsulated meningococci. PLoS One 2012; 7:e45132; PMID:23028802; http://dx.doi.org/10.1371/journal.pone.0045132
48. Peysselon F, Xue B, Uversky VN, Ricard-Blum S. Intrinsic disorder of the extracellular matrix. Mol Biosyst 2011; 7:3353-65; PMID:22009114; http://dx.doi.org/10.1039/c1mb05316g

49. Keiser PB, Biggs-Cicatelli S, Moran EE, Schmiel DH, Pinto VB, Burden RE, et al. A phase 1 study of a meningococcal native outer membrane vesicle vaccine made from a group B strain with deleted lpxL1 and synX, over-expressed factor H binding protein, two PorAs and stabilized OpcA expression. Vaccine 2011; 29:1413-20; PMID:21199704; http://dx.doi.org/10.1016/j.vaccine.2010.12.039
50. Uversky VN, Dunker AK. Understanding protein nonfolding. Biochim Biophys Acta 2010; 1804:1231-64; PMID:20117254; http://dx.doi.org/10.1016/j.bbapap.2010.01.017
51. Unkmeir A, Latsch K, Dietrich G, Wintermeyer E, Schinke B, Schwender S, et al. Fibronectin mediates Opc-dependent internalization of Neisseria meningitidis in human brain microvascular endothelial cells. Mol Microbiol 2002; 46:933-46; PMID:12421301; http://dx.doi.org/10.1046/j.1365-2958.2002.03222.x
52. Tompa P. The interplay between structure and function in intrinsically unstructured proteins. FEBS Lett 2005; 579:3346-54; PMID:15943980; http://dx.doi.org/10.1016/j.febslet.2005.03.072
53. Tusnády GE, Dosztányi Z, Simon I. PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank. Nucleic Acids Res 2005; 33(Database issue):D275-8; PMID:15608195; http://dx.doi.org/10.1093/nar/gki002
54. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics 2010; 26:680-2; PMID:20053844; http://dx.doi.org/10.1093/bioinformatics/btq003
55. Dunbrack R. xml2pdb. http://dunbrack.fccc.edu/xml2pdb.php. (Accessed December 2, 2011)
56. Maiti R, Van Domselaar GH, Zhang H, Wishart DS. SuperPose: a simple server for sophisticated structural superposition. Nucleic Acids Res 2004; 32(Web Server issue):W590-4; PMID:15215457; http://dx.doi.org/10.1093/nar/gkh477
57. Mohan A, Uversky VN, Radivojac P. Influence of sequence changes and environment on intrinsically disordered proteins. PLoS Comput Biol 2009; 5:e1000497; PMID:19730682; http://dx.doi.org/10.1371/journal.pcbi.1000497
58. Wu CH, Yeh LS, Huang H, Arminski L, Castro-Alvear J, Chen Y, et al. The Protein Information Resource. Nucleic Acids Res 2003; 31:345-7; PMID:12520019; http://dx.doi.org/10.1093/nar/gkg040
59. Vacic V, Uversky VN, Dunker AK, Lonardi S. Composition Profiler: a tool for discovery and visualization of amino acid composition differences. BMC Bioinformatics 2007; 8:211; PMID:17578581; http://dx.doi.org/10.1186/1471-2105-8-211
60. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990; 215:403-10; PMID:2231712
61. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics 2007; 23:2947-8; PMID:17846036; http://dx.doi.org/10.1093/bioinformatics/btm404
62. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 2009; 25:1189-91; PMID:19151095; http://dx.doi.org/10.1093/bioinformatics/btp033
63. The PyMOL Molecular Graphics System. Version 1.5.0.4 Schrödinger, LLC. http://sourceforge.net/projects/pymol/ (Accessed May 8, 2012).