Comparative Analysis of TM and Cytoplasmic β-barrel Conformations Using Joint Descriptor

Jayaraman Thangappan1,2, Sangwook Wu1 & Sun-Gu Lee2

Macrosopic descriptors have become valuable as coarse-grained features of complex proteins and are complementary to microscopic descriptors. Proteins macrosopic geometric features provide effective clues in the quantification of distant similarity and close dissimilarity searches for structural comparisons. In this study, we performed a systematic comparison of β-barrels, one of the important classes of protein folds in various transmembrane (TM) proteins against cytoplasmic barrels to estimate the conformational features using a joint-based descriptor. The approach uses joint coordinates and dihedral angles (β and γ) based on the β-strand joints and loops to determine the arrangements and propensities at the local and global levels. We then confirmed that there is a clear preference in the overall β and γ distribution, arrangements of β-strands and loops, signature patterns, and the number of strand effects between TM and cytoplasmic β-barrel geometries. As a robust and simple approach, we determine that the joint-based descriptor could provide a reliable static structural comparison aimed at macroscopic level between complex protein conformations.

β-barrel proteins are among the most significant and abundant entities in both membrane and cytoplasmic proteins as well as helical domains. These β-barrel structures are composed of a repertoire of β-strand segments joined by loops. Each strand is hydrogen bonded to its neighbor in an antiparallel/parallel arrangement, thus forming β sheets. Such β sheets twist up together to form a closed or open cylinder-like structure. Although membrane and cytoplasmic β-barrel proteins differ sequentially in their amino acid propensities, both β-barrel proteins share similar structural folds. In addition, β-barrel proteins can carry out a variety of biological functions, which has gained them wide-spread attention. Most transmembrane β-barrel (β-barrelsTM) structures are closed and act as pores or channels representing uniquely shaped Outer Membrane Proteins (OMP)6,7, while the cytoplasmic β-barrel structures (β-barrelsacytoplasm) are mixed fractions of open and closed structures functioning as catalytic and ligand binding units with diverse topology8-10. Most OMPs are detected with an even number of strands ranging from about four to 2611-14. On the other hand, most β-barrelsacytoplasm proteins are diverse and range between four to 10 strands, with fewer numbers of proteins ranging up to 14 strands15. OMP β-barrels are monomeric with a single chain barrel, and barrels of cytoplasmic proteins often show complicated topologies in which more than one barrel contributes to a single chain16. Together, both β-barrelsTM and β-barrelsacytoplasm structures are significant for therapeutic targeting because they resemble the channeling system of helical proteins17-19.

While β-barrelsacytoplasm structures are clearly determined, β-barrelsTM structure determinations are still progressing under ever-increasing challenges in relation to TM protein purification and crystallization20-22. Only a few hundred unique β-barrelsTM structures have been identified, and these constitute an inconsequential part of the known structure deposited in the Protein Data Bank (PDB), succeeding the membrane helical protein genomes sequenced so far. Supplementary to experimental procedures, computational modeling has been accelerating our understanding of β-barrel proteins23-25. However, such analyses have not yet illuminated the nature of β-barrel structures due to the complexities of conformational distribution. Comparison of the intricacies of any distinctive features of β-barrelsTM and β-barrelsacytoplasm structures could be refined down by simple representations. Several descriptions of protein structures have been illustrated to show the relationship between TM and cytoplasmic β-barrel geometries.

1Department of Physics, Pukyong National University, Busan, 608-737, Republic of Korea. 2Department of Chemical Engineering, Pusan National University, Busan, 609-735, Republic of Korea. Correspondence and requests for materials should be addressed to S.W. (email: sangwoow@pknu.ac.kr) or S.-G.L. (email: sungulee@pusan.ac.kr)
between secondary structural elements (SSEs) and several essential parameters that govern the structure and constraints. Elucidation of geometric features using detailed atomic and Cα residue representations is a common approach to the analysis of helical and β-sheet packing characterizations. In addition, various levels of coarse-grained models have been suggested to reduce the complexity. However, a basic model that relates to the structural arrangement of both β-barrels and β-barrels structures at the macroscopic level is required to define the connective similarities and dissimilarities. Because β-strand structural arrangements play a critical role in the function and organization of β-barrel proteins, models that utilize the β-strand conformations would be more useful. Furthermore, such models would facilitate the understanding of how β-barrel structural organizations endure both membrane and water-soluble environments while utilizing their sheet arrangements, overall topology, and conformational heterogeneity. Therefore, methods that identify the macroscopic principles underlying β-barrel protein folding and design factors in the overall structure, function, dynamics, and interactions become promising.

In the present study, we attempted to perform a systematic comparison of β-barrels and β-barrels structures using the joint-based description approach. For comparison, structures of high-resolution, non-homologous β-barrels and β-barrels protein structures were obtained from the PDB and grouped based on their geometric analyses. We considered the understanding of both β-barrel types with the topological patterns that exhibit diverse structural and functional relationships. We believe that β-barrel structural organizations can be statistically quantified based on β-strands arrangements and conformational distributions. Thus, both β-barrels and β-barrels structures were analyzed in terms of two dihedral angles; β and γ. Our results demonstrate that TM β-strands are predominantly right-handed and anti-parallel as reported previously, and they are comparatively like the conformational flexibility of TM helices in nature. In contrast, structures of β-barrels in the cytoplasm are short, mixed and varied in their arrangements with one another. All the observed conformational distributions and adjacent frequencies were associated with the arrangement of β-strands and loops in the context of how β-barrel structures have adapted in both membrane and water-soluble environments respectively. Despite uniformity, we observed that there are some considerable differences between β-barrels and β-barrels that share an equal number of β-strands. The comparative analysis results of β-barrel proteins prove that the joint-based description approach is a powerful descriptor that can be used to boost the geometrical characterization of β-barrels at both local and global levels. Possibly, our joint-based descriptor approach will be an asset in β-barrels structural prediction, as well as in the in-silico design, modeling, and validation of topologies with the given β-strand and loop segments.

**Results**

**Macroscopic descriptor for β-barrel structures.** The present study deals in relation to the recently developed joint-based description of protein structures. To understand the macroscopic geometric features of both TM and cytoplasmic β-barrel proteins such as their conformational distribution and the arrangements of β-strands, we utilize the joint constraint principles that were previously prepared to represent TM helices. These β-barrel proteins always show consecutive elements of β-strands and loops linked one another as described in Fig. 1. Like TM helical proteins, a set of joints connecting the individual β-strands and loops were selected to describe a β-barrel geometry based on the joint-based description approach. Specifically, the Cα carbon coordinates of the first and last residues of each β-strand were considered as structural joining points as they represent the basic elements of protein structures. The spatial arrangement of such joint points was measured by the dihedral angles between them. For example, a protein composed of eight β-strands (S1, S2, S3, S4, S5, S6, S7, and S8) and seven loops (L1, L2, L3, L4, L5, L6, and L7), a group of 16 joints (P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P11, P12, P13, P14, P15, and P16) can be assigned (Fig. 1). As a result, dihedral angle involving first four joints P1, P2, P3, and P4 can be determined by measuring the angle between two planes made by P1, P2, P3, and P4, P5, P6, P7, and P8. Then, the second dihedral angle can be found by relating the structural points P7, P8, P9, and P10, P11, P12, and P13, P14, P15, P16, and P17, respectively. Thus, the β-strand conformations of both β-barrels and β-barrels structures can be represented by the set of joints (P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P11, P12, P13, P14, P15, P16, and P17) at the macroscopic level. Clockwise and counter-clockwise angle signs were specified based on a positive value (from 0 to 180 degrees) or a negative value (from −180 to 0 degrees) respectively (Fig. 1).

**Dataset collection of targets β-barrel proteins.** For β-barrels, 29 proteins with an even number from 4 to 26 β-strands and odd number 19 TM β-barrel proteins were collected as an initial dataset. For β-barrels, a total of 51 proteins from several superfamilies were chosen. The total number of both TM and cytoplasmic β-barrel proteins with various topologies, β-strands, and loop numbers were listed in Table 1 and Table 2 respectively. These datasets were directly obtained from Protein Data Bank (PDB) by grouping into β-barrels and β-barrels proteins. The dataset collection procedure is cited in the Methods section. Concisely in both cases, i) We selected 30% sequence criteria and filtered the high-resolution X-ray crystal structures from PDB, ii) We have separated the non-homologous monomeric chains with respect to different number of β-strand was grouped as final target proteins as shown in Fig. 2(a,b). In the end, the target dataset of 29 and 51 β-barrel structures was analyzed and all such analyzed β and γ angles are grouped and classified (Supporting Information SI Tables (1, 2, 3 and 4)).

**Comparison of β and γ distributions and overall arrangements of both β-barrels and β-barrels.** In this section, we compare both the β-γ plots for β-barrels and β-barrels at the macroscopic level in analogy with Ramachandran ϕ-ψ plot at the atomistic level to find out the distribution
density of dihedral angles. Our β-γ plot reflects the Ramachandran map of peptide distribution, such as favorable and unfavorable regions for β-strands and loops as given in Fig. 3(a,b). The overall distribution can possibly assist to understand the similarities and dissimilarities in their structural arrangement patterns. Overall β and γ distributions are restricted to certain regions for β-barrelTM indicating they show conformationally similar arrangements that are observed in Ω-λ plot for TM helical proteins. On the other hand, β and γ distributions for β-barrelscytoplasm are sparsely distributed over greater regions signifying dissimilar patterns in their arrangements than the membrane counterparts. Conventionally, the relative conformation of adjacent structural elements i.e. neighboring elements both at residues and secondary structures level have influence on each other and that can be identified by the distribution of their dihedral angles. Our new joint-derived dihedral angles such as β and γ distributions were related to the structural arrangement of β-strands for both β-barrelTM and β-barrelscytoplasm architecture just as previously demonstrated with TM helical proteins using Ω and λ dihedral angles. The dihedral angle between the joints is not only associated with the arrangements of the individual β-strands and loops

Table 1. Selected non-homologous β-barrelsTM protein structures (29), their PDB IDs, and the total number of β types and γ type dihedral angles used in this study. All protein names and their classifications used in this work are described in the Supplementary Information Table S9.
but also involves in the dependency of adjacent secondary structural element orientations. If β-strands and loops are simplified as shown in Fig. 1, β-strands arrangements can be explained by β type dihedral angles between the ith β-strand (S_i) and its neighboring i + 1th β-strand (S_{i+1}). Similarly, the type γ can represent the arrangement of the loop segments between the ith loop (L_i) and its neighboring i + 1th loop (L_{i+1}). In addition, type γ dihedral angle also provides the relative arrangement between the ith β-strand (S_i) and i + 2th β-strand (S_{i+2}). Figure 4 show some specific relation between the dihedral angles and β-strands (or loops) arrangements. When the dihedral angle β_i is close to 0 degrees, β-strand S_i is antiparallel with the adjacent β-strand S_{i+1} (Fig. 4(a)). In a similar way, when the dihedral angle β_i is close to ±180 degrees, β-strand S_i is parallel with the adjacent β-strand S_{i+1} (Fig. 4(b)). When the dihedral angle γ_i is close to 0 degrees, loop L_i is antiparallel with the adjacent loop L_{i+1}, and β-strand S_i and β-strand S_{i+2} are on the same side with respect to β-strand S_{i+1} (Fig. 4(c)). Meanwhile, when the dihedral angle γ_i is close to ±180 degrees, loop L_i is parallel with the adjacent loop L_{i+1}, and β-strand S_i and β-strand S_{i+2} are on the opposite side respect to β-strand S_{i+1} (Fig. 4(d)). Here, we attempt to compare both β-barrelsTM and β-barrels cytoplasm geometric characteristics through consecutive β-strands and loops arrangements by investigating the distribution of β and γ dihedral angles.

The overall distribution plots for all β and γ angles that belong to both β-barrelsTM (29) and β-barrels cytoplasm (51) non-homologous protein structures was shown in Fig. 5. In both cases, it was evident that the β type dihedral angles were mostly confined to the range of −30° to +30°. Figure 5(a) shows more than 80% of β distribution between 0° to +30° for β-barrelsTM, where the frequency in a clockwise (positive angles) orientation is evidently dominant than that of anti-clockwise (negative angles). We observed the most favored dihedral angle distribution regions between +10° and +20° corresponding to their restricted distribution space for TM β dihedral angles. Meanwhile, β type distributions for β-barrels cytoplasm were observed in an almost entire range of −150° to +150° with a strong preference towards 0° to +30° as given in Fig. 5(b). In contrast, type γ dihedral angles

| Group | PDB IDs | # of Structures | # of β types | # of γ types |
|-------|---------|----------------|-------------|-------------|
| 4N    | 2B97A & 1G3PA | 2 | 5 | 4 |
| 5N    | 1NB9A, 1WHIA, 1PV4A, 1GC9A, 1HK9A, 1R6JA, 1O6AA, & 1G31A | 8 | 32 | 24 |
| 6N    | 2JDD, 1WIXA, 1AQJA, 1DFUP, 1S89A, 1EFTA, 2D9RA, 2JMLA, 1BCOA, 1TS8A, 2B1NA & 2FILA | 12 | 60 | 48 |
| 7N    | 1WPOA, 1JK9A, 1ORUA, 1JF9A, 1RQP8, 1OP7A, 1H0WA, 2GUJA, & 1QZ9A | 9 | 54 | 45 |
| 8N    | 1EAR, 1UE0A, 1A1XA, 1NQCA, 1OEWA, 1XE1A, 1T2WA, 1PQHA, 2CPLA, 1Y12A, 2Q03A, 2F9HA, 1BEBA, 1GQ8A, & 1F3UB | 15 | 105 | 90 |
| 10N   | 3P12A | 1 | 9 | 8 |
| 11N   | 2FR2A | 1 | 10 | 9 |
| 12N   | 4FGA | 1 | 11 | 10 |
| 13N   | 1NLSA | 1 | 12 | 11 |
| 14N   | 1H4GA | 1 | 13 | 12 |
| Total numbers | 51 | 312 | 261 |

Table 2. Selected non-homologous β-barrels cytoplasm protein structures (51), their PDB IDs, and the total number of β types and γ type dihedral angles used in this study. All protein names and their classifications used in this work are described in the Supplementary Information Table S10.
were distributed in the entire possible region between $-180^\circ$ to $+180^\circ$ for both $\beta$-barrels$^{\text{TM}}$ and $\beta$-barrels$^{\text{cytoplasm}}$ proteins. However, TM $\gamma$ distribution shows dominant peaks between $+90^\circ$ to $+180^\circ$ region (Fig. 5(c)). Possibly, $\gamma$ preference of TM structures observed to be very dominance between $+100^\circ$ to $+150^\circ$ considering the overall distribution region. While type $\gamma$ dihedral angles for $\beta$-barrels$^{\text{cytoplasm}}$ are distributed widely with equal preference (Fig. 5(d)). Interestingly, the frequencies of TM $\gamma$ type dihedral angles in clockwise orientations are apparent than counter-clockwise, suggesting a preferable arrangement of adjacent $\beta$-strands tend to be right-handed orientation. Together, TM $\beta$-strands topological arrangements differ from cytoplasmic $\beta$-strands in overall $\beta$-barrel architecture, i.e. both $\beta$ and $\gamma$ dihedral angle distributions vary with one another. Membrane $\beta$-strands incline to pack within smaller conformational space, whereas cytoplasmic $\beta$-strand arrangements were relatively different. The propensities for antiparallel and right-handed orientation properties of $\beta$-strands for TM and soluble proteins are reported in several proteins packing studies$^{47,48}$. Certainly, both $\beta$-barrels$^{\text{TM}}$ and $\beta$-barrels$^{\text{cytoplasm}}$ share related conformational distribution space with significant variations observed in the cytoplasmic structures.
β-barrels were observed with (β-dihedral angle clusters such as mixed arrangements in the depicted in Fig. 4. It should also be noted that cytoplasmic γ) Histogram showing the distribution of dihedral angles for β-barrelscytoplasm proteins. (Figure 5. The overall distribution of β and γ dihedral angles. (a) Histogram showing the distribution of β dihedral angles for β-barrelsTM proteins. (b) Histogram showing the distribution of β dihedral angles for β-barrelscytoplasm proteins. (c) Histogram showing the distribution of γ dihedral angles for β-barrelsTM proteins. (d) Histogram showing the distribution of γ dihedral angles for β-barrelscytoplasm proteins.

In terms of arrangements, Fig. 5(a,b) show β type dihedral angles indicated the preference in the narrow range as a main accessible region that two neighboring β-strand (S_i and S_{i+1}) most likely to be arranged in antiparallel as depicted in Fig. 4. It should also be noted that cytoplasmic β-strand arrangements show antiparallel, parallel and mixed arrangements in the β-barrel proteins. Despite the sparse distribution in the all possible range of −180° to +180°, it is observed that TM γ-type dihedral angles were predominantly distributed between 30° to +180° region according to Fig. 4(c). However, the overall orientations of L_i and L_{i+1} indicate that β-barrelsTM neighboring loops are appeared to exist in both anti-parallel and parallel arrangements. On the other hand, cytoplasmic γ-type dihedral angles do not have a strong preference towards any region supporting the all possible antiparallel, parallel and mixed arrangements. Thus, cytoplasmic β-strand S_{i+2} are freely arranged between the same side and opposite side to β-strand S_i.

Comparison of local patterns. A comparison of local patterns of TM β-barrels with cytoplasmic β-barrel geometries was attempted by analyzing the nearest neighbor frequencies of β and γ-type dihedral angles. The measurement of dihedral angle β enables the prediction of the arrangement of neighboring β-strands (S_i and S_{i+1}) and the relative positions of two β-strands S_i and S_{i+1} can be determined by measuring dihedral angle γ. These indicate that the measurements of the continuous dihedral angles can allow us to predict how the β-strands in β-barrel proteins are continuously arranged and to which extent they are twisted and varied. For instance, the information of β_i and β_{i+1} can determine the arrangement of S_i, S_{i+1}, and S_{i+2}, and the information of γ_i and γ_{i+1}, may allow the prediction of the relative positions of S_{i+2} and S_{i+3} to S_i and S_{i+1}. Here, to study the β-strands arrangements in both β-barrelsTM and β-barrelscytoplasm systems, we examined the local patterns of continuous dihedral angle clusters such as \( \beta_i - \beta_{i+1}, \gamma_i - \gamma_{i+1}, \beta_i - \beta_{i+1} - \beta_{i+2} \) and \( \gamma_i - \gamma_{i+1} - \gamma_{i+2} \). Most of the β-strand structures in β-barrelsTM were observed with (+, +) and (+, +, +) patterns for \( \beta_i - \beta_{i+1}, \gamma_i - \gamma_{i+1} \) and the same trend have been traced with signature patterns of β-barrelscytoplasm as well. Interestingly, (+, +) signature pattern means a strong bias toward right-handed topology.

Dihedral angles of β and γ-type were categorized into two groups: clockwise - positive value (from 0 to +180 degrees) and counter-clockwise - negative value (from −180 to 0 degrees). Then, the dihedral angles (β and γ) for both β-barrelsTM (29) and β-barrelscytoplasm (51) structures were interpreted as the combination of signatures (positive and negative sign) (SI Tables 5, 6, 7 and 8)). Firstly, dyad signature (i.e. Two adjacent/consecutive units) distribution patterns of β_i - β_{i+1} and γ_i - γ_{i+1} for both datasets was examined and that indicated a strong tendency towards (+, +) pattern (Fig. 6). For a β_i - β_{i+1} cluster, the frequency of (+, +) pattern was obviously higher than others for both β-barrelsTM and β-barrelscytoplasm, respectively (Fig. 6(a,b)). Also, the frequency of (+, +) pattern was higher than others for the γ_i - γ_{i+1} cluster (Fig. 6(c)). Only cytoplasmic γ_i - γ_{i+1} has shown a moderately different trend in all their groups (Fig. 6(d)). For example, the pattern (−, −) showing considerable frequency, whereas all other dyad clusters have it as the least common. Secondly, we examined the triad sequential signature patterns (i.e. three adjacent/consecutive units) of β type and γ type dihedral angles for both β-barrelsTM and β-barrelscytoplasm structures. We observed β_i - β_{i+1} - β_{i+2} and γ_i - γ_{i+1} - γ_{i+2} clusters predominant towards (+, +, +) patterns among the 8 possible combinations (Fig. 7(a–c)). And patterns such as (−, −, +), (+, +, −),
Figure 6. Frequencies of the dyad patterns for consecutive $\beta$ or $\gamma$ type dihedral angles when $\beta$ or $\gamma$ type angles are categorized as $(+)$ and $(-)$. The bar diagrams show the observed numbers of two consecutive $\beta$ type angles, $\beta_i - \beta_{i+1}$, for $\beta$-barrel$^{TM}$ proteins. (b) $\beta$-barrel$^{cytoplasm}$ proteins showing four different patterns of two consecutive $\beta$ type angles, $\beta_i - \beta_{i+1}$ (c) Four different patterns of two consecutive $\gamma$ type angles, $\gamma_i - \gamma_{i+1}$, for $\beta$-barrel$^{TM}$ proteins. (d) $\beta$-barrel$^{cytoplasm}$ proteins showing four different patterns of two consecutive $\gamma$ type angles, $\gamma_i - \gamma_{i+1}$.

and $(+, -, +)$ were also observed noticeably. However, cytoplasmic $\gamma_i - \gamma_{i+1} - \gamma_{i+2}$ (Fig. 7(d)) has been identified with relatively different inclination in all their signature patterns. Finally, the comparisons of $\gamma$ type dihedral angles of TM with cytoplasmic proteins were performed. Both $\gamma$-type distributions were found in the entire range from $-180^\circ$ to $+180^\circ$. Subsequently, the dihedral angle space was divided into four quadrants as (I) $^A$: $0^\circ$ to $+90^\circ$, (II) $^B$: $-90^\circ$ to $0^\circ$, (III) $^C$: $+90^\circ$ to $+180^\circ$, and (IV) $^D$: $-180^\circ$ to $-90^\circ$ for further adjacent frequency distribution analysis. In the quadrant, the pattern of $\gamma_i - \gamma_{i+1}$ cluster was studied more in detail. Among all possible 16 combinations of $\gamma_i - \gamma_{i+1}$ cluster, the most abundant dihedral angle distribution was observed in the range of $(+90^\circ$ to $+180^\circ$, $+90^\circ$ to $+180^\circ$) for $\beta$-barrel$^{TM}$ (Fig. 8(a)). There is no such clear preference for $\beta$-barrel$^{cytoplasm}$. However, there are few positive (AA), (BB), (AB) and (BA) regions with relatively better propensities were observed commonly in both the cases. The observed $\beta$-strands and loop angles are intensely positive, corresponding to a right-handed twist orientation. Altogether, type $\beta_i - \beta_{i+1}$ and $\gamma_i - \gamma_{i+1}$ signature analysis has revealed that the positive sign as a clearly dominant region for adjacent $\beta$-strands which can be correlated to the right-handed twist.

The twisted property of $\beta$-strands can be explained by both the conformation of individual $\beta$-strands and the relative orientations of adjacent $\beta$-strands using $\beta$ and $\gamma$ type dihedral angles. Figure 9 shows a conventional image of $\beta$-barrel proteins with several $\beta$-strands that have previously used to portray adjacent TM helical arrangements. Figure 9(a,b) show the schematic arrangement of three consecutive $\beta$-strands, i.e. $S_i, S_{i+1}$, and $S_{i+2}$, depending on the pattern of $\beta_i - \beta_{i+1}$ cluster, observed in the front and side view respectively. As recorded in Fig. 4, four different $(+, +), (+, -), (-, +)$ and $(-, -)$ patterns determine four different types of arrangement between $S_i$ and $S_{i+2}$ in parallel (Fig. 9(b)). The dominance of $(+, +)$ and $(+, +, +)$ patterns in $\beta_i - \beta_{i+1}$ and $\beta_i - \beta_{i+1} - \beta_{i+2}$ cluster indicates that $\beta$-barrel favor twisted pattern. Figure 9(c) shows a schematic picture of $\beta$-barrel proteins with several $\beta$-strands in top view. It shows how the $\beta$-strands were extended depending on the pattern of $\gamma_i - \gamma_{i+1}$ cluster. The $(-, -)$ and $(+, +)$ patterns suggest that $\beta$-strands are set in one direction with a distorted pattern in such a way it forms a twisted topology. The $(+, -)$ and $(-, +)$ patterns, however, show that $\beta$-strands are arranged such that $\beta$-strands are tightly packed. The clear preference of $(+, +)$ and $(+, +, +)$ pattern for $\gamma_i - \gamma_{i+1}$ and $\gamma_i - \gamma_{i+1} - \gamma_{i+2}$ clusters suggests that $\beta$-strands in the membrane proteins favor the arrangements with a twisted pattern in the right-hand direction. Similarly, the dominant dihedral angular distribution of ($90^\circ$ to $180^\circ$, $180^\circ$ to $180^\circ$) in the 16 possible patterns of $\gamma_i - \gamma_{i+1}$ cluster (Fig. 8) implies that there is also some angle preference in the twisted type arrangement of TM $\beta$-strands than cytoplasmic ones. These dihedral angle preferences and the right-handed properties of the $\beta$-strands are strongly correlated with the overall $\beta$-barrel architectures. The specific conformations of $\beta_i - \beta_{i+1}$ and $\gamma_i - \gamma_{i+1}$ cluster confirms that adjacent positive angles may provide continued stability, thus $\beta$-strands twisting possibly control the conformational space.

The right-handed orientation and antiparallel arrangement of $\beta$-barrel$^{TM}$ can be related to the distribution of $\beta$ type and $\gamma$ type dihedral angles found between $+10^\circ$ to $+20^\circ$ and $+100^\circ$ to $+150^\circ$ respectively. Few TM $\gamma$ type dihedral angles show a trend of relatively small dihedral value for short intracellular loops and larger
dihedral value for long extracellular loops in their distributions locally, whereas no such bias has observed with β-barrel proteins. In general, conformational energetics strongly favors the antiparallel β-strand arrangements which have an advantage over the stability in terms of the packing interaction. It is also known that the antiparallel barrel folds are particularly more stable when the strand orders are a consecutively right-handed conformation. The combined orientations of individual strands and adjacent strands adopt an overall closed barrel or partially opened or distorted barrels. Structurally, the right-handed connectivity is common for all β-strands embedded in the finite hydrophobic membranes that prefer the antiparallel arrangements. Meanwhile, no specific trends have been found with β-barrel proteins in terms of their γ-type. Both TM and cytoplasmic adjacent loop orientations were observed as they can be found opposite side with no direct interaction between them. Though loops are arranged in a random orientation that can lead to maximization of entropy, they adopt

Figure 7. Frequencies of the triad patterns for consecutive β or γ type dihedral angles when β or γ type angles are categorized as (+) and (−). The bar diagrams show the observed numbers of (a) eight different patterns of three consecutive β type angles, $\beta_i - \beta_{i+1} - \beta_{i+2}$, for β-barrels proteins. (b) β-barrel proteins eight different patterns of three consecutive β type angles, $\beta_i - \beta_{i+1} - \beta_{i+2}$. (c) Eight different patterns of three consecutive γ type angles, $\gamma_i - \gamma_{i+1} - \gamma_{i+2}$, for β-barrels proteins. (d) β-barrel proteins eight different patterns of three consecutive γ type angles, $\gamma_i - \gamma_{i+1} - \gamma_{i+2}$.

Figure 8. Frequencies of the dyad patterns for consecutive γ type dihedral angles when the γ type dihedral angle was split into four regions. The bar diagram shows the observed numbers of the 16 different patterns of two consecutive γ type angles, $\gamma_i - \gamma_{i+1}$. Here, all λ type angles were split into four regions, i.e. $+A(0° to 90°)$, $+B(90° to 180°)$, $-A(-90° to 0°)$, and $-B(-90° to -180°)$ (a) Out of 16 patterns from the combinations of two consecutive γ type angles, $\gamma_i - \gamma_{i+1}$, [BB] pattern were abundant for β-barrels proteins (29) and (b) for β-barrel proteins, there is no obvious preference in the patterns in their 51 non-homologous dataset.
few recurrent conformations (adjacent loops orientations can be specific for protein families). Besides, we subsequently quantified the respective dihedral angle distributions in effect with the number and position of β-strands for both β-barrelsTM and β-barrelscytoplasm proteins.

Comparison of β-strand numbers and positions effect in both β-barrels. We also examined how the similarities and dissimilarities for both β-barrelsTM and β-barrelscytoplasm in terms of the number and position of β-strands. As the numbers of β-strands vary, the arrangements of β-strands and loops might be changed due to change of interaction energy between β-strands and entropy of the flexible loop region. The distributions of dihedral angles for β and γ types were plotted according to their relative positions in β-strands for TM fold (Fig. 10(a,c)) and cytoplasmic fold (Fig. 10(b,d)) respectively. It seems β-barrelsTM proteins are relatively larger than β-barrelscytoplasm proteins, and so the numbers of β-strands are observed less in cytoplasmic folds. According to the number of β-strands, the β and γ type dihedral distribution were plotted for TM configurations (Fig. 11(a,c)) and cytoplasmic configurations (Fig. 11(b,d)) respectively. For β-barrelsTM proteins with 4, 10, 12, 14 and 19 β-strands obviously suggest that γ type distribution is also restricted to specific accessible regions and on the other hand, rather sparse distribution noted with 16, 18, 22 and 26 β-strands (Fig. 11(c)). For β-barrelscytoplasm proteins with 5, 6, 7 and 8 β-strands were found to be the major groups determining their distributions (Fig. 11(b)). In both cases, the distribution of β and γ type dihedral angles was quite different compared to those of overall distributions as shown in Fig. 5. These results suggest that the dihedral angles are significantly affected by the relative position of β-strands and loops as the number of the β-strands fluctuate in both groups. Further, the patterns of the distribution of β−βi+1 and γi−γi+1 cluster were analyzed according to three different numbers of groups for both cases.

Membrane β-barrel structures with 4–10 TM β-strands, with 12–16 TM β-strands, and with 18–26 TM β-strands were considered exclusively for size effect and grouped into smaller (structures with 4–10 TM β-strands), medium (structures with 12–16 TM β-strands) and larger membrane (structures with 18–26 TM β-strands) β-barrel proteins (Fig. 12(a,b)). In contrast, cytoplasmic β-barrel structures were grouped into 4–6N β-strands, 7–8N β-strands, and 10–14N β-strands according to their size grouping, where N denotes the number of β-strands (Fig. 13(a,d)). As shown in Fig. 12(a), (+, +) pattern for β−βi+1 cluster show up more frequently for the proteins with 4–10 TM β-strands. The pattern remained same as the TM increases from 4–10 TM β-strands to 12–16 TM β-strands and for 18–26 TM β-strands. However, it is found that (−, +), (+, −) and (−, −) patterns also shows up notably with 18–26 TM β-strands group compared to other groups. Dihedral angles of β17, β20 and β23 are sparsely distributed in the negative angles (−60° to 0°) belongs to β-barrel proteins with 18–24 TM β-strands. The (+, +) pattern becomes more dominant with all group proteins sharing β-strands. Especially, pattern (+, +) shows up more frequently for the proteins with 4–10 TM β-strands for the γi−γi+1 cluster. As the number of TM β-strands increase, it becomes more prevail in 12–16 TM β-strands and 18–26 TM β-strands groups (Fig. 12(b)). Noticeably, the (+, +) pattern is abundant for all the TM β-strands groups. Though (+, +) pattern is dominant, (−, −), (+, −) and (−, −) patterns are also observed considerably for β−βi+1 and γi−γi+1

Figure 9. Relation to the dihedral angle patterns and β-strands arrangements or extensions. (a) Front view of the linearly ordered β-strands in β-barrel proteins with β-strands and loops. (b) Side view of the arrangement of three consecutive β-strands S1−S1+1 = S1+2 depending on the four different patterns of two consecutive β type angles, β−βi+1. (c) Top view configuration of four consecutive helices, S1−S1+1−S1+2−S1+3, for the four different patterns of two consecutive γ type angles, γi−γi+1.

Figure 10. Dependencies of dihedral angles for β- and γ-type of TM β-strands and overall distributions of TM β-strands. As the number of β- and γ-type of TM β-strands increases, the dihedral angles 0° to 60° (−, −) pattern is more abundant, (−, +), (+, −) and (−, −) pattern are more frequently observed for the TM β-strands groups with 12–16 and 18–26 TM β-strands. However, for the TM β-strands groups with 4–10 TM β-strands, (+, +) and (−, −) pattern is more predominant. In addition, (-, +) pattern are more prevalent for 12–16 TM β-strands groups. Other patterns are observed for the TM β-strands groups with 4–10 TM β-strands. As the number of TM β-strands increase, the pattern (+, +) becomes more prevalent in 12–16 TM β-strands and 18–26 TM β-strands groups (Fig. 11(b)). Noticeably, the (+, +) pattern abundance is abundant for all the TM β-strands groups. Though (+, +) pattern is dominant, (−, −), (+, −) and (−, −) patterns are also observed considerably for β−βi+1 and γi−γi+1.

Figure 11. Distribution of dihedral angles for β- and γ-type of TM β-strands and overall distributions of TM β-strands. As the number of β- and γ-type of TM β-strands increases, the dihedral angles 0° to 60° (−, −) pattern is more abundant, (−, +), (+, −) and (−, −) pattern are more frequently observed for the TM β-strands groups with 12–16 and 18–26 TM β-strands. However, for the TM β-strands groups with 4–10 TM β-strands, (+, +) and (−, −) pattern is more predominant. In addition, (-, +) pattern are more prevalent for 12–16 TM β-strands groups. Other patterns are observed for the TM β-strands groups with 4–10 TM β-strands. As the number of TM β-strands increase, the pattern (+, +) becomes more prevalent in 12–16 TM β-strands and 18–26 TM β-strands groups (Fig. 11(b)). Noticeably, the (+, +) pattern abundance is abundant for all the TM β-strands groups. Though (+, +) pattern is dominant, (−, −), (+, −) and (−, −) patterns are also observed considerably for β−βi+1 and γi−γi+1.
Figure 10. Dihedral angle distributions in the β-barrels\textsuperscript{TM} and β-barrels\textsuperscript{cytoplasm} proteins depending on their configurations. (a) Scatter plot of the β type dihedral angles of β-barrels\textsuperscript{TM} proteins depending on their configurations. (b) Scatter plot of β type dihedral angles of β-barrels\textsuperscript{cytoplasm} proteins depending on their configurations. (c) Scatter plot of the γ type dihedral angles of β-barrels\textsuperscript{TM} proteins depending on their configurations. (d) Scatter plot of γ type dihedral angles of β-barrels\textsuperscript{cytoplasm} proteins depending on their configurations. For (a–d), all the \(i\)\textsuperscript{th} dihedral angles (\(\beta_i\) or \(\gamma_i\) values) of both of β-barrels\textsuperscript{TM} (29) and β-barrels\textsuperscript{cytoplasm} (51) non-homologous β-barrel proteins were collected and plotted against \(\beta_i\) or \(\gamma_i\) in the x-axis.

Figure 11. Dihedral angle distributions in the β-barrels\textsuperscript{TM} and β-barrels\textsuperscript{cytoplasm} proteins depending on the β-strands numbers. (a) Scatter and line plot of the β type dihedral angles of β-barrels\textsuperscript{TM} proteins depending on the β-strands numbers. (b) Scatter and line plot of β type dihedral angles of β-barrels\textsuperscript{cytoplasm} proteins depending on the β-strands numbers. (c) Scatter and line plot of the γ type dihedral angles of β-barrels\textsuperscript{TM} proteins depending on the β-strands numbers. (d) Scatter and line plot of γ type dihedral angles of β-barrels\textsuperscript{cytoplasm} proteins depending on the β-strands numbers. For (a–d), all the \(i\)\textsuperscript{th} dihedral angles (\(\beta_i\) or \(\gamma_i\) values) of both β-barrels\textsuperscript{TM} (29) and β-barrels\textsuperscript{cytoplasm} (51) non-homologous β-barrel proteins were given in y-axis and plotted against β-strands numbers in the x-axis.
Figure 12. Frequencies of the patterns for two consecutive β or γ type dihedral angles depending on the β-barrelsTM proteins with a different number of β-strands. All the 29 non-homologous β-barrelsTM proteins were categorized into three groups, i.e., proteins with 4 to 10 TM β-strands, 12 to 16 TM β-strands, and 18 to 26 TM β-strands. The bar diagrams show the observed numbers of (a) four different patterns of two consecutive β type angles, $\beta_i - \beta_{i+1}$, in each group, and (b) four different patterns of two consecutive γ type angles, $\gamma_i - \gamma_{i+1}$, in each group.

clusters of the proteins with larger TM structures (18–26 TM β-strands), respectively. On the other hand, cytoplasmic β-barrel proteins show slightly different trends. Although cytoplasmic $\beta_i - \beta_{i+1}$ also follows the same (+, +) pattern as the preferred pattern in all respective groups as observed with TM proteins (Fig. 13(a)) and the patterns trend for $\gamma_i - \gamma_{i+1}$ cluster are different. As shown in Fig. 13(b), for $\gamma_i - \gamma_{i+1}$ cluster, (+, −) pattern is the least common pattern of the small proteins with 4–6N β-strands. However, the pattern shifts as the β-strands (N) increases from 4–6N β-strands to 7–8N β-strands, where (+, −) pattern become dominant, yet other patterns such as (+, +), (−, +) and (−, −) are also observed substantially. For 10–14N β-strands, all (+, +), (−, −), (−, +) and (−, −) patterns remains equally dominant. Despite the overall uniformity in their structural arrangements, these results imply that β-strands arrangement of both β-barrel structures prefers to be positioned in a specific pattern for the efficient packing of β-strands as the number of β-strands changes. Like the overall distribution, the dihedral angles (β and γ) and signature preferences for various β-strands numbers and position are rather different between β-barrel proteins in the membrane and cytoplasmic origin.

Discussion

The present study provides a systematic comparison of macroscopic features of β-barrelsTM and β-barrelscytoplasm proteins, providing a more simplistic way to understand the structural organization of complex proteins. Furthermore, our studies help to differentiate the specificity through various β-strands arrangements by analyzing the overall conformational space, signature patterns, positions and numbers of β-strands. For the comparative analysis of TM β-barrels with cytoplasmic structures, the joint-derived dihedral angles of β-strands were used. The β-γ dihedral angles are macroscopic descriptions of protein structures based on joint-based approach. In both cases, specific allowed and disallowed conformational accessible spaces were found. Here, we used non-homologous datasets of β-barrelsTM and β-barrelscytoplasm proteins for comparison and the approach can be extended to compare the conformational heterogeneities and geometric features of other polymers. It is well known that the adjacent β-strands in the membrane were mostly arranged in antiparallel due to a confinement inside membrane favoring stability and conformational energetics. The overall dihedral angle distribution analyses for the β and γ-type suggest that the neighboring two β-strands would be antiparallel orientation as shown in Fig. 3(a) for TM proteins. On the other hand, cytoplasmic β-strands with greater conformational flexibility can possibly have all the antiparallel, parallel and mixed arrangements. In addition, the analyses of adjacent dihedral signature pattern analysis, such as $\beta_i - \beta_{i+1}$ and $\gamma_i - \gamma_{i+1}$ intensely supports the TM β-strands favored to be right-handed orientations with (+, +) pattern being clearly dominant. For cytoplasmic proteins, the same (+, +) signature patterns are clearly abundant preferring the right-handed orientations. Commonly, the (+, +) patterns were dominant over other signatures for β-strands arrangements. Though γ-type overall distribution is different in both the cases, they have common features of (+, +) patterns for loop-loop arrangements. Based on the results, we envisage that right-handed orientation is profoundly relevant to the stereochemistry of interaction...
between β-strands at the atomic level. Presumably, the successive combination of (+, +) pattern for \( \beta_i - \beta_{i+1} \) (orientations of β-strands) and \( \gamma_i - \gamma_{i+1} \) (orientations of loops) has an advantage for TM β-strands packing. With relatively greater conformational space, cytoplasmic proteins tend to have slightly loosely packed β-strands. Together, these results prove the packing similarities and dissimilarities between β-barrels\(^{TM} \) and β-barrels\(^{cytoplasm} \) proteins.

The difference in the overall dihedral angle distributions according to the number of β-strands for both the β and γ-type reveals additional specifics. We support the fact that the β-strands positioning of β-barrel proteins could be significantly restricted more inside membrane lipid bilayer than the cytoplasmic environment as the number of TM β-strands changes. Our results could aid to comprehend the differences in their overall structure as it is reflected in the variability of the dihedral angle distributions of β and γ-type at the macroscopic level. This infers the local arrangement of β-strands and loops can be affected and varied by the number of β-strands increases. The \( \gamma_i - \gamma_{i+1} \) cluster analysis also shows a clear preference of a “right-handed” conformation as the number of β-strands fluctuates, implying a right-handed orientation is optimized form required for the efficient TM β-strand packing geometry. In contrast, (+, −) pattern appear marginally higher than (+, +) pattern of the cytoplasmic proteins with 4–6N β-strands. It was clearly observed that the arrangements of β-strand orientations are diverse depending on the β-strand positions in cytoplasmic proteins. Taken together, our results indicate that the β-barrel structure in the lipid membrane environment is more significantly controlled than the structures of cytoplasmic β-strand structures at local and global arrangements.

All such macroscopic geometric analyses provide significant clues for better understanding to improve in-silico protein structure and function prediction, modeling, and validation. The relationship between the dihedral angles and β-strand arrangements provide more focus on β-barrels engineering and designing principles. In general, any rule that is related to such structural property supports to compare and classify the protein functions. Here, the antiparallel arrangements and right-handed orientations are common for most of the β-strand structures, while the mixed loop arrangement is also essential for β-barrel structural formation. Perhaps, non-adjacent and various combinations of β and γ-type dihedral angles would be distinguished to locate interaction and non-interaction part of the β-strands. In future, the combination of macroscopic joint description of protein structures as a coarse-grain level with all-atom or CoS residue level can help us to improve our understanding towards large structural rearrangements and dynamic behavioral studies of complex structures.

**Methods**

**Datasets as target β-barrel proteins used in this study.** At first, we selected and filtered all membrane proteins from the Protein Data Bank by using the key word “Membrane proteins” and then, with the help of selection mode, β-barrel proteins X-ray crystal structures were separated. It was found that 837 structures belong to TM β-barrel proteins. The dataset of 592 β-barrel structures having sequence identity less than 90%, were identified to be unique proteins containing both homologous and non-homologous protein chains. From the above
Joint-based structural joint points and \( \beta \) and \( \gamma \)-type dihedral measurements. The first and last residue Cα atoms of each TM and cytoplasmic secondary structural (SS) segment were projected as joint structural points for the dihedral calculations. Structural joint selections and \( \beta \)-strands boundaries were prepared based on the annotations given in OPM and STRIDE server respectively. The previously developed in-house program was used for the joint-derived dihedral angle measurements. It should be noted that the total number of \( \beta \) and \( \gamma \)-type dihedral angles for a protein structure entirely depends on the total number of TM \( \beta \)-strands and loops present in the structure i.e. the total number of SS segments and joint coordinates. While forming all linkage of the joint points, a macroscopic description of the overall protein structure was portrayed.

Statistical analyses dihedral angle patterns. To find out the preferred relative orientations, the conformational analysis was performed based on the dihedral angles obtained and the signature patterns dominance was revealed among various combinations of consecutive dihedral angles. For the selected 29 and 51 structures for both TM and cytoplasmic cases presented by the \( \beta_3^{-\gamma_0} \beta_3^{-\gamma_1} \beta_3^{-\gamma_1} \gamma_0 \) dihedral angle set, as summarized in Supplementary Information Tables (S1, S2, S3 and S4); N stands for \( \beta \) strand numbers in the protein structures. The calculated dihedral angles were converted to positive (\( \text{ve} \)) and negative (\( \text{ne} \)) signatures to represent the conformations, as given in Supplementary Information Tables (S5, S6, S7 and S8). A consecutive \( \beta \)-pattern was selected for each fold as \( \beta_3^{-\gamma_0} \beta_3^{-\gamma_1} \). Grouped \( \beta_3^{-\gamma_0} \beta_3^{-\gamma_1} \) should be a consecutive, adjacent set, and no fixed order, whereas non-consecutive \( \beta_3^{-\gamma_0} \beta_3^{-\gamma_1} \) were not considered. For example, the \( \beta \)-pattern signatures were selected from \( \beta_1^{-\gamma_0} \beta_2^{-\gamma_0} \beta_3^{-\gamma_0} \gamma_0 \) to \( \beta_3^{-\gamma_0} \gamma_0 \) as any consecutive \( \beta_3^{-\gamma_0} \). To make more defined distribution patterns, the consecutive \( \beta_3^{-\gamma_0} \beta_3^{-\gamma_1} \) and \( \beta_3^{-\gamma_0} \gamma_0 \) were also tested.

References

1. Fairman, J. W., Noinaj, N. & Buchanan, S. K. The structural biology of beta-barrel membrane proteins: a summary of recent reports. Curr Opin Struct Biol 21, 523–531, https://doi.org/10.1016/j.sbi.2011.05.005 (2011).
2. Iacoboni, L., Martelli, P. L., Fariselli, P., De Pinto, V. & Casadio, R. Prediction of the transmembrane regions of beta-barrel membrane proteins with a neural network-based predictor. Protein Sci 10, 779–787, https://doi.org/10.1101/ps.37201 (2001).
3. Galdiero, S., Galdiero, M. & Pedone, C. Beta-Barrel membrane bacterial proteins: structure, function, assembly and interaction with lipids. Curr Protein Pept Sci 8, 63–82 (2007).
4. Tamm, L. K., Arora, A. & Kleinschmidt, J. H. Structure and assembly of beta-barrel membrane proteins. J Biol Chem 276, 32399–32402, https://doi.org/10.1074/jbc.R100021200 (2001).
5. Bishop, R. E. Structural biology of membrane-intrinsic beta-barrel enzymes: sentinels of the bacterial outer membrane. Biochim Biophys Acta 1778, 1881–1896, https://doi.org/10.1016/j.bbamem.2007.07.021 (2008).
6. Kim, K. H., Aulakh, S. & Paetzel, M. The bacterial outer membrane beta-barrel assembly machinery. Protein Sci 21, 731–788, https://doi.org/10.1002/pro.2069 (2012).
7. Tamm, L. K., Hong, H. & Liang, B. Folding and assembly of beta-barrel membrane proteins. Biochim Biophys Acta 1666, 250–263, https://doi.org/10.1016/j.bbamem.2004.06.011 (2004).
8. Stepunenko, O. V., Stepunenko, O. V., Kuznetsova, I. M., Verkhusha, V. V. & Turoverov, K. K. Beta-barrel scaffold of fluorescent proteins: folding, stability and role in chromophore formation. Int Rev Cell Mol Biol 302, 221–278, https://doi.org/10.1016/B978-0-12-407699-0-00004-2 (2013).
9. Flower, D. R. The up-and-down beta-barrel proteins: three of a kind. FASEB J 9, 566–567 (1995).
10. Castillo, R. M. et al. A six-stranded double-psi beta barrel is shared by several protein superfamilies. Struct Fold Des 7, 227–236, https://doi.org/10.1016/S0969-2126(99)80028-8 (1999).
11. Rollauer, S. E., Sooreshjami, M. A., Noinaj, N. & Buchanan, S. K. Outer membrane protein biogenesis in Gram-negative bacteria. Philos T R Soc B 370, https://doi.org/10.1098/rstb.2015.0023 (2015).
12. Noinaj, N., Gumbart, J. C. & Buchanan, S. K. The beta-barrel assembly machinery in motion. Nat Rev Microbiol 15, https://doi.org/10.1038/nrmicro.2016.191 (2017).
13. Tarigov, K. D., Elofsson, A. & Bagos, P. G. PRED-TMBB2: improved topology prediction and detection of beta-barrel outer membrane proteins. Bioinformatics 32, 665–671, https://doi.org/10.1093/bioinformatics/btw444 (2016).
14. Gromiha, M. M. & Selvaraj, S. Protein secondary structure prediction in different structural classes. Protein Eng 11, 249–251, https://doi.org/10.1093/protein/11.4.249 (1998).
15. Lasters, I., Wodak, S. J., Alard, P. & Vancutsem, E. Structural Principles of Parallel Beta-Barrels in Proteins. P Natl Acad Sci USA 85, 3338–3342, https://doi.org/10.1073/pnas.85.10.3338 (1988).
16. Cheng, H. et al. ECOD: An Evolutionary Classification of Protein Domains. PloS Comput Biol 10, https://doi.org/10.1371/journal.pcbi.1003926 (2014).
17. Witther, D. M., Papic, D., Bos, M. P., Tommassen, J. & Rapaport, D. Signals in bacterial beta-barrel proteins are functional in eukaryotic cells for targeting to and assembly in mitochondria. P Natl Acad Sci USA 106, 2531–2536, https://doi.org/10.1073/pnas.0807830106 (2009).
18. Bakalov, J., Buchanan, S. K. & Nothnagel, N. Structural snapshots of the beta-barrel assembly machinery. *Fubs* 284, 1778–1786, https://doi.org/10.1111/febs.13960 (2017).

19. Arinaminpathy, Y., Khurana, E., Engelmann, D. M. & Gerstein, M. Computational analysis of membrane proteins: the largest class of drug targets. *Drug Discov Today* 14, 1130–1135, https://doi.org/10.1016/j.drudis.2009.08.006 (2009).

20. Tulumello, D. V. & Deber, C. M. Efficiency of detergents at maintaining membrane protein structures in their biologically relevant forms. *Biochim Biophys Acta* 1818, 1351–1358, https://doi.org/10.1016/j.bbamem.2012.01.013 (2012).

21. Carpenter, E. P., Beis, K., Cameron, A. D. & Iwata, S. Overcoming the challenges of membrane protein crystallography. *Curr Opin Struct Biol* 18, 581–586, https://doi.org/10.1016/j.sbi.2008.07.001 (2008).

22. Moraes, I., Evans, G., Sanchez-Weatherby, J., Newstead, S. & Stewart, P. D. S. Membrane protein structure determination The next generation. *BBA-Biomembranes* 1838, 76–87, https://doi.org/10.1016/j.bbamem.2013.07.010 (2014).

23. Venko, N., Boy Choudhury, A. & Novic, M. Computational Approaches for Revealing the Structure of Membrane Transporters: Case Study on Bilitranslocase. *Comput Struct Biotechnol* 15, 223–242, https://doi.org/10.1016/j.csbet.2017.01.008 (2017).

24. Gessmann, D. et al. Outer membrane beta-barrel protein folding is physically controlled by periplasmic lipid head groups and BamP. *PNAS* 111, 5878–5883, https://doi.org/10.1073/pnas.1224731111 (2014).

25. Lin, M. S., Gessmann, D., Naveed, H. & Liang, J. Outer Membrane Protein Folding and Topology from a Computational Transfer Free Energy Scale. *J Am Chem Soc* 138, 2592–2601, https://doi.org/10.1021/jacs.5b10307 (2016).

26. Tran Vdu, T. Chassagnon, P., Chassagnon, S. & Steyaert, J. M. A graph theoretic approach for classification and structure prediction of transmembrane beta-barrel proteins. *BMC Genomics* 13(Suppl 2), S5, https://doi.org/10.1186/1471-2164-13-S2-55 (2012).

27. Murzin, A. G., Lesk, A. M. & Chothia, C. Principles determining the structure of beta-sheet barrels in proteins. II. The observed structures. *J Mol Biol* 236, 1382–1400 (1994).

28. Murzin, A. G., Lesk, A. M. & Chothia, C. Principles determining the structure of beta-sheet barrels in proteins. I. A theoretical analysis. *J Mol Biol* 236, 1369–1381 (1994).

29. Chothia, C. & Janin, J. Orthogonal packing of beta-sheeted proteins. *Biochemistry* 21, 3955–3965 (1982).

30. King, R. D., Clark, D. A., Shirazi, J. & Sternberg, M. J. E. On the Use of Machine Learning to Identify Topological Rules in the Orthogonally Packed Beta-Sheets in Proteins. *FEBS Lett* 12, 1295–1303, https://doi.org/10.1016/S0014-5793(94)81095-8 (1994).

31. Hayat, S., Sander, C., Marks, D. S. & Elofsson, A. All-atom 3D structure prediction of transmembrane beta-barrel proteins from sequences. *PNAS* 112, 5413–5418, https://doi.org/10.1073/pnas.1419596112 (2015).

32. Yang, A. S. & Honig, B. Free-Energy Determinants of Secondary Structure Formation. 2. Antiparallel Beta-Sheets. *Proc Natl Acad Sci USA* 98, 13521–13525, https://doi.org/10.1073/pnas.98.24.13521 (2001).

33. Wang, G. & Dunbrack, R. L. Jr. PISCES: a protein sequence culling server. *Bioinformatics* 17, 1326–1327, https://doi.org/10.1093/bioinformatics/17.12.1326 (2001).

34. Orengo, C. A. PDB: the protein data bank. *Nucleic Acids Res* 36, W500–502, https://doi.org/10.1093/nar/gkq429 (2004).

35. Yang, A. S. & Honig, B. Free-Energy Determinants of Secondary Structure Formation. 2. Antiparallel Beta-Sheets. *Journal of Molecular Biology* 252, 366–376, https://doi.org/10.1006/jmbi.1995.0303 (1995).

36. Wang, G. & Dunbrack, R. L. Jr. PISCES: recent improvements to a PDB sequence culling server. *Nucleic Acids Res* 33, W94–98, https://doi.org/10.1093/nar/gkj402 (2005).

37. Wang, G. & Dunbrack, R. L. Jr. PISCES: a protein sequence culling server. *Bioinformatics* 19, 1589–1591 (2003).

38. Lomize, M. A., Lomize, A. L., Pogozheva, I. D. & Mosberg, H. I. OPM: Orientations of proteins in membranes database. *Bioinformatics* 22, 623–625, https://doi.org/10.1093/bioinformatics/btl023 (2006).

39. Turna, G. E., Doutzandy, L. & Simon, I. PDB: TM selection and membrane localization of transmembrane proteins in the protein data bank. *Nucleic Acids Research* 33, D275–D278, https://doi.org/10.1093/nar/gkj002 (2005).

40. Orengo, C. A. et al. CATH—a hierarchical classification of protein domain structures. *Structure* 5, 1093–1108 (1997).

41. Hubbard, T. J., Murzin, A. G., Brenner, S. E. Chothia, C. SCOP: a structural classification of proteins database. *Current Opinion in Structural Biology* 25, 236–239 (1997).

42. Fox, N. K., Brenner, S. E. & Chandonia, J. M. SCOP: Structural Classification of Proteins—extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Res* 42, D304–309, https://doi.org/10.1093/nar/gkt1240 (2014).

43. Heinig, M. & Frishman, D. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res* 32, W500–502, https://doi.org/10.1093/nar/gkq429 (2004).
Acknowledgements
This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2015R1D1A1A01061125). This research was also supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) (2018R1A2B6001670).

Author Contributions
J.T. and S.G.L. designed research; J.T. performed research; J.T., S.W., and S.G.L. analyzed data; J.T., S.W., and S.G.L. wrote the paper. All authors reviewed the manuscript.

Additional Information
Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-32136-4.

Competing Interests: The authors declare no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018