Comparative analysis of $^{13}$C chemical shifts of β-sheet amyloid proteins and outer membrane proteins

Noah H. Somberg$^1$ · Martin D. Gelenter$^1$ · Mei Hong$^1$

Received: 8 January 2021 / Accepted: 24 March 2021 / Published online: 12 April 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract
Cross-β amyloid fibrils and membrane-bound β-barrels are two important classes of β-sheet proteins. To investigate whether there are systematic differences in the backbone and sidechain conformations of these two families of proteins, here we analyze the $^{13}$C chemical shifts of 17 amyloid proteins and 7 β-barrel membrane proteins whose high-resolution structures have been determined by NMR. These 24 proteins contain 373 β-sheet residues in amyloid fibrils and 521 β-sheet residues in β-barrel membrane proteins. The $^{13}$C chemical shifts are shown in 2D $^{13}$C–$^{13}$C correlation maps, and the amino acid residues are categorized by two criteria: (1) whether they occur in β-strand segments or in loops and turns; (2) whether they are water-exposed or dry, facing other residues or lipids. We also examine the abundance of each amino acid in amyloid proteins and β-barrels and compare the sidechain rotameric populations. The $^{13}$C chemical shifts indicate that hydrophobic methyl-rich residues and aromatic residues exhibit larger static sidechain conformational disorder in amyloid fibrils than in β-barrels. In comparison, hydroxyl- and amide-containing polar residues have more ordered sidechains and more ordered backbones in amyloid fibrils than in β-barrels. These trends can be explained by steric zipper interactions between β-sheet planes in cross-β fibrils, and by the interactions of β-barrel residues with lipid and water in the membrane. These conformational trends should be useful for structural analysis of amyloid fibrils and β-barrels based principally on NMR chemical shifts.

Keywords
Chemical shifts · Amyloid proteins · Beta-barrel membrane proteins · Conformational distribution

Introduction
Chemical shifts report the local electronic environment of nuclear spins and are thus sensitive to the conformation and electrostatic interaction of functional groups in molecules. As a result, the different backbone conformations and sidechain structures of amino acids in proteins cause characteristic chemical shift differences. Protein $^{13}$C chemical shifts are sensitive to the backbone (ϕ, ψ) torsion angles (Wishart et al. 1991), sidechain χ₁ and χ₂ torsion angles, as well as weak but functionally important interactions such as hydrogen bonding and aromatic interactions (Vranken and Rieping 2009). The empirical relationship between $^{13}$C chemical shifts and protein (ϕ, ψ) angles is well established for globular proteins (Wishart and Sykes 1994; Wishart et al. 1991, 1992). By correlating NMR chemical shifts with high-resolution structures solved using distance-restrained NMR data and crystal structures (Shen and Bax 2013; Shen et al. 2009; Spera and Bax 1991), databases such as TALOS-N (Shen and Bax 2013) can predict (ϕ, ψ) and χ₁ angles based on measured Cα, Cβ and CO chemical shifts. While TALOS-N predicts protein torsion angles, the ROSETTA Monte Carlo program has been successfully used to predict the structures of small globular proteins and amyloid proteins based on chemical shifts (Sgourakis et al. 2015; Shen et al. 2008; Skora and Zweckstetter 2012). While structure determination from chemical shifts is the principal goal of NMR spectroscopists, the reverse task of accurately predicting chemical shifts from known structures is also beneficial. This would allow a comparison of protein structures solved using X-ray crystallography and cryoEM with the structures of proteins whose NMR chemical shifts are available. Reliable predictions of chemical shifts can also simplify resonance assignment and facilitate studies of protein dynamics. At present, chemical shifts can be predicted from structures using SHIFTX2 (Han et al. 2011) and SPARTA+ (Shen and
an interesting comparison with amyloid proteins for understanding the intrinsic conformational preferences and \(^{13} \text{C}\) chemical shifts of \(\beta\)-sheet proteins.

In this work, we analyze the \(^{13} \text{C}\) chemical shifts of 17 high-resolution cross-\(\beta\) amyloid protein structures and 7 \(\beta\)-barrel membrane protein structures solved by solid-state and solution NMR, in order to deduce the conformational differences between these two families of proteins. More \(\beta\)-barrel membrane protein structures are available in the literature (Dutta et al. 2017; Horst et al. 2014). However, since a \(\beta\)-barrel protein is typically much larger than an amyloid protein and hence contributes more residues to the dataset, we analyzed a subset of \(\beta\)-barrel structures such that the two datasets have similar numbers of chemical shifts. Two OmpG structures are used: one structure was solved in the detergent octyl-\(\beta\)-glucopyranoside using solution NMR, with only backbone chemical shifts available (Liang and Tamm 2007), while the other structure was solved in \textit{Escherichia coli} lipid extracts using solid-state NMR (Retel et al. 2017) and has both sidechain and backbone chemical shifts. In the amyloid dataset, multiple A\(\beta\)40 and A\(\beta\)42 structures are used; these A\(\beta\) structures are polymorphic and distinct, thus have different chemical shifts. The \(^{13} \text{C}\) chemical shifts of these proteins are presented in 2D \(^{13} \text{C}–^{13} \text{C}\) correlation maps, and the mean and standard deviations of the chemical shifts of each carbon are calculated. We also present the abundance of each amino acid in amyloid proteins, \(\beta\)-barrels and all proteins, and the sidechain \(\chi_1\) and \(\chi_2\) angle distributions. This analysis allows us to identify several conformational trends that differ between amyloid fibrils and \(\beta\)-barrel membrane proteins for various amino acids.

### Methods

The 17 amyloid proteins and 7 \(\beta\)-barrels whose \(^{13} \text{C}\) chemical shifts are analyzed here are listed in Table 1. The chemical shifts were obtained from Biological Magnetic Resonance Databank (BMRB) or from the original publications where no BMRB entries are available (Dregni et al. 2019; van der Wel et al. 2007). DSS was used as the chemical shift reference. TMS-referenced chemical shifts were converted to the DSS scale by adding 2.00 ppm to the values (Morcombe and Zilm 2003; Wishart et al. 1995). No change was made to TSP-referenced chemical shifts.

To correlate \(^{13} \text{C}\) chemical shifts with structures, we categorized residues according to (1) whether they come from an amyloid protein or a \(\beta\)-barrel protein and (2) whether they lie on a \(\beta\)-strand or in a loop or turn. \(\beta\)-sheet residues in \(\beta\)-barrels are further distinguished by whether they face the lipids or water-filled pore. \(\beta\)-sheet residues in amyloid proteins are further distinguished by whether they face the dry interior or solvent. Non \(\beta\)-strand residues in amyloid proteins
are also distinguished by whether they are dry or solvent-exposed. In β-barrels, non β-sheet residues often cannot be clearly distinguished between water-exposed and lipid-exposed positions, thus we did not make further distinctions for loop, turn, or helical residues in β-barrels. In total, six structural categories were created for each amino acid: water-exposed β-sheet residues in β-barrels, lipid-exposed β-sheet residues in β-barrels, dry β-sheet residues in fibrils, water-exposed β-sheet residues in fibrils, dry loop or turn residues in fibrils, and water-exposed loop or turn residues in fibrils (Table 2). We compiled and analyzed the chemical shifts of the six categories, excluding the non-β-sheet residues in β-barrels. The average and standard deviation of β-sheet residues in fibrils and β-barrels are tabulated in Table 3.

Conformational information, including \((\phi, \psi)\) and \((\chi_1, \chi_2)\) torsion angles and the β-sheet designations, was obtained from the Protein Data Bank (PDB). For structures without PDB entries, information such as β-sheet designation and sidechain structure were obtained from the original publication. For PDB entries that include multiple copies of the same β-strand, we extracted a single monomer from the center of the deposited structure to avoid potential torsion angle distortions due to edge defects. For protein structures that have multiple conformations, we analyzed the lowest energy conformation.

### Table 1: Protein structures whose chemical shifts are analyzed in this study

| Proteins              | Reference                  | pH  | T (K) | BMRB # | PDB #  | CS reference | Strand | No. β-sheet residues | Total no. residues |
|-----------------------|----------------------------|-----|-------|--------|---------|--------------|---------|----------------------|-------------------|
| Aβ40                  | Lu et al. (2013)           | 7.4 | 273   | 19009  | 2M4J    | DSS          | D       | 22                   | 38                |
| Aβ40                  | Bertini et al. (2011)      | 8.5 | 283   | 34454  | 6TI5    | TSP          | C       | 22                   | 30                |
| Aβ40                  | Paravastu et al. (2008)    | 7.4 | 300   | 18131  | 2LMQ    | TMS          | C       | 12                   | 32                |
| Osaka Aβ40            | Schütz et al. (2015)       | 7   | 273   | 25289  | 2MVX    | DSS          | D       | 19                   | 37                |
| Aβ42                  | Colvin et al. (2016)       | 8   | 277   | 30121  | 5KK3    | DSS          | D       | 13                   | 30                |
| Aβ4β                  | Wälti et al. (2016)        | 7.4 | 273   | 26692  | 2NAO    | DSS          | E       | 18                   | 40                |
| Aβ42                  | Gremer et al. (2017)       | 2   | 278   | 27212  | 5QOV    | DSS          | C       | 30                   | 42                |
| α-synuclein           | Tuttle et al. (2016)       | 7   | 273   | 25518  | 2N0A    | DSS          | E       | 48                   | 138               |
| HET-s                 | Wasm et al. (2008)         | 7.5 | 278   | 11028  | 2RN7    | DSS          | C       | 28                   | 77                |
| Necrosome             | Mompeán et al. (2018)      | 6.5 | 278   | 30273  | 5V7Z    | DSS          | D, E    | 20                   | 33                |
| FUS-LC                | Murray et al. (2017)       | 7.4 | 298   | 30304  | 5W3N    | DSS          | E       | 20                   | 59                |
| Tau                   | Dregni et al. (2019)       | 7.4 | 260–293 | –   | –     | TMS          | N/A     | 30                   | 30                |
| Glucagon              | Gelerter et al. (2019)     | 2.1 | 293   | 30572  | 6NZN    | TMS          | E, K    | 51                   | 54                |
| GNNQQNYa              | Sawaya et al. (2007), van der Wel et al. (2007) | 2–3 | 277 | – | – | 2OMM | DSS | A 7 | 7 |
| TTR                   | Fitzpatrick et al. (2013)  | 2   | 298   | 19062  | 2MSN    | DSS          | D       | 11                   | 11                |
| Zn-binding fibril     | Lee et al. (2017)          | 8   | 298   | 30227  | 5UGK    | TMS          | D       | 5                    | 5                 |
| MAX1                  | Nagy-Smith et al. (2015)   | 9   | 273   | 25558  | 2N1E    | DSS          | C       | 17                   | 20                |
| VDAC-1                | Hiller et al. (2008)       | 7.0 | 303   | 16381  | 2K4T    | DSS          | N/A     | 169                  | 283               |
| OmpA BBP              | Johansson et al. (2007)    | 6   | 303   | 15045  | 2MJM    | DSS          | N/A     | 84                   | 156               |
| OmpG                  | Liang and Tamm (2007)      | 6.3 | 313   | 15426  | 2QYQ    | DSS          | N/A     | 178                  | 275               |
| OmpG                  | Retel et al. (2017)        | 6.8 | 280   | 34088  | 5MWV    | DSS          | N/A     | 121                  | 279               |
| OmpX                  | Hagn et al. (2013)         | 6.5 | 318   | 18796  | 2M06    | DSS          | N/A     | 100                  | 148               |
| OprH                  | Edrington et al. (2011)    | 6.1 | 273   | 17842  | 2LHF    | DSS          | N/A     | 79                   | 178               |
| YadA                  | Shahid et al. (2012)       | 7   | 275   | 18108  | 2LME    | DSS          | B       | 43                   | 105               |

| Total                 |                           | 373 | 683   |        |        |              |         |                      |                   |

| a | An interior β-strand in the deposited amyloid protein structure that is selected to extract torsion angles |
| b | The number of residues in the β-sheet conformation |
| c | The total number of residues in one β-strand for amyloid fibrils and the total number of residues in the β-barrel structures |
| d | Chemical shifts were extracted directly from the publication. No BMRB or PDB entry exists for this publication, and no secondary structural data was used |
| e | Secondary structural data was extracted from PDB entry 2OMM (Sawaya et al. 2007). Chemical shifts were extracted directly from (van der Wel et al. 2007) No BMRB entry exists for this publication |
All torsion angles were analyzed in PyMol. Sidechain torsion angles were obtained using a home-written PyMol script that measures each angle from the PDB structure. For structures with multiple identical subunits, we chose a single subunit from the interior of the structure. For the \( \chi_1 \) angle around the C\( \alpha \)-C\( \beta \) bond, 0\( ^\circ \) is defined as when the C\( \beta \)-C\( \gamma \) bond is aligned with the N-C\( \alpha \) bond. For the \( \chi_2 \) angle around the C\( \beta \)-C\( \gamma \) bond, 0\( ^\circ \) is defined as when the C\( \alpha \)-C\( \beta \) bond is aligned with the C\( \gamma \)-C\( \delta \) bond. (Lovell et al. 2000).

Home-written MATLAB scripts were used to extract the chemical shifts of each amino acid residue and to associate the chemical shifts with the structural categories. This dataset was used to construct the 2D \( ^{13}\text{C} \)–\( ^{13}\text{C} \) chemical shift maps for each amino acid type. A given \( ^{13}\text{C} \) chemical shift was correlated to all other chemical shifts in the same residue, thus the constructed 2D correlation map represents all possible combination of \( ^{13}\text{C} \)–\( ^{13}\text{C} \) cross peaks within a residue.

For a given cross peak such as C\( \alpha \)–C\( \beta \) in one of the six structural categories, the mean of the chemical shifts for \( \beta \)-sheet residues was computed for both the direct and indirect dimensions. An ellipse was then plotted with its center at the position of the mean in each dimension. The standard deviation of the \( \beta \)-sheet chemical shifts is calculated for each dimension and represented as half the semi-major and semi-minor axes lengths of the ellipse. In other words, the ellipse represents twice the standard deviation of the chemical shift in each dimension. These ellipses are plotted separately for fibril and barrel \( \beta \)-sheet residues, and separately for each type of cross peaks for a given amino acid. The ellipse positions and sizes are calculated using a MATLAB script that computes the mean and 2\( \sigma \) confidence interval for each cross peak. The chemical shift limits for each type of cross peaks are defined manually. When the boundary of a cross peak crosses the diagonal, we only calculated the chemical shifts of peaks on one side of the diagonal.

### Results and discussion

We first examined the abundance of each amino acid in the \( \beta \)-sheet segments of amyloid fibrils, \( \beta \)-barrels, and all proteins in the UniProtKB/Swiss-Prot databank (Consortium 2018). The percentages are relative to the total number of residues in each of the three protein categories. Figure 1 shows that Val is highly enriched in amyloid proteins, accounting for more than 1/6 of all residues. This is consistent with the known ability of the two methyl groups of the Val sidechain to engage in steric zipper interactions (Nelson et al. 2005). Ile and Gln are also enriched in amyloid proteins relative to their abundance in all proteins. In comparison, several amino acids such as Arg and Leu are

| Residues | All proteins | \( \beta \)-barrel | \( \beta \)-sheets | Cross-\( \beta \) fibrils |
|----------|--------------|-------------------|------------------|-------------------------|
|          | Total | Barrel | Fibril | Lipid facing | Water-facing | \( \beta \)-sheet, dry | \( \beta \)-sheet, wet | Loop/turn, dry | Loop/turn, wet |
| ALA      | 147   | 99    | 48    | 34          | 29           | 15              | 5               | 9              | 19          |
| ARG      | 80    | 66    | 14    | 3           | 31           | 4               | 6               | 1              | 3           |
| ASN      | 126   | 98    | 28    | 7           | 30           | 8               | 5               | 8              | 7           |
| ASP      | 133   | 101   | 32    | 6           | 17           | 6               | 6               | 4              | 16          |
| CYS      | 3     | 2     | 1     | 1           | 1            | 1               | 0               | 0              | 0           |
| GLN      | 93    | 53    | 40    | 17          | 17           | 13              | 11              | 9              | 7           |
| GLU      | 123   | 82    | 41    | 5           | 32           | 3               | 11              | 3              | 24          |
| GLY      | 250   | 163   | 87    | 24          | 74           | 11              | 11              | 11             | 54          |
| HIS      | 49    | 22    | 27    | 3           | 4            | 6               | 7               | 2              | 12          |
| ILE      | 80    | 45    | 35    | 28          | 3            | 22              | 3               | 3              | 7           |
| LEU      | 140   | 110   | 30    | 58          | 10           | 17              | 4               | 4              | 5           |
| LYS      | 112   | 66    | 46    | 7           | 24           | 2               | 24              | 1              | 19          |
| MET      | 35    | 22    | 13    | 5           | 8            | 3               | 4               | 2              | 4           |
| PHE      | 96    | 71    | 25    | 40          | 9            | 8               | 7               | 6              | 4           |
| PRO      | 46    | 37    | 9     | 10          | 1            | 2               | 0               | 0              | 7           |
| SER      | 146   | 92    | 54    | 7           | 35           | 12              | 11              | 15             | 16          |
| THR      | 124   | 94    | 30    | 16          | 32           | 9               | 9               | 7              | 5           |
| TRP      | 38    | 35    | 3     | 17          | 3            | 0               | 2               | 0              | 1           |
| TYR      | 126   | 96    | 30    | 51          | 26           | 6               | 8               | 3              | 13          |
| VAL      | 159   | 70    | 89    | 39          | 10           | 37              | 23              | 17             | 12          |
| Total    | 2106  | 1424  | 682   | 378         | 396          | 185             | 157             | 105            | 235         |
### Table 3

Mean and standard deviation of $^{13}$C chemical shifts for β-sheet residues in cross-β fibrils and β-barrels. Trp and Cys are omitted due to insufficient statistics.

| Residue | Atom | Average chemical shift of β-sheet residues | BMRB average CS (Ulrich et al. 2008) | Count |
|---------|------|-----------------------------------------|--------------------------------------|-------|
|         |      | Fibrils | Barrels | ∆CS | Fibrils | Barrels |
| Ala     | C    | 175.6 ± 1.7 | 175.5 ± 0.9 | 0.0 | 177.8 | 18 | 46 |
|         | CA   | 50.5 ± 0.4 | 50.6 ± 1.0 | −0.2 | 53.2 | 18 | 59 |
|         | CB   | 21.9 ± 2.2 | 22.8 ± 1.4 | −0.9 | 19.0 | 18 | 56 |
| Arg     | C    | 174.6 ± 0.9 | 174.2 ± 0.9 | 0.4 | 176.5 | 8 | 28 |
|         | CA   | 54.8 ± 0.3 | 54.5 ± 0.9 | 0.3 | 56.8 | 8 | 30 |
|         | CB   | 33.9 ± 0.9 | 33.8 ± 2.2 | 0.1 | 30.6 | 8 | 30 |
|         | CD   | 43.5 ± 1.0 | 44.3 ± 0.4 | −0.8 | 43.2 | 7 | 8 |
|         | CG   | 27.4 ± 0.7 | 27.4 ± 0.9 | 0.1 | 27.2 | 7 | 8 |
|         | CZ   | 159.7 ± 0.4 | 159.6 ± 0.2 | 0.1 | 159.9 | 4 | 9 |
| Asn     | C    | 174.4 ± 1.6 | 173.6 ± 1.2 | 0.8 | 175.3 | 11 | 25 |
|         | CA   | 53.1 ± 1.0 | 52.4 ± 1.1 | 0.8 | 53.5 | 12 | 35 |
|         | CB   | 41.1 ± 2.9 | 41.8 ± 2.0 | −0.8 | 38.7 | 11 | 35 |
|         | CG   | 176.3 ± 1.1 | 177.7 ± 0.7 | −1.5 | 176.7 | 10 | 4 |
| Asp     | C    | 173.8 ± 0.5 | 174.4 ± 1.3 | −0.6 | 176.4 | 9 | 13 |
|         | CA   | 52.7 ± 0.8 | 53.3 ± 1.0 | −0.5 | 54.7 | 9 | 20 |
|         | CB   | 41.6 ± 2.7 | 43.3 ± 1.6 | −1.7 | 40.9 | 9 | 21 |
|         | CG   | 178.7 ± 1.7 | 180.3 ± 0.7 | −1.7 | 178.9 | 9 | 2 |
| Gln     | C    | 174.1 ± 1.0 | 174.3 ± 1.3 | −0.2 | 176.4 | 18 | 24 |
|         | CA   | 54.3 ± 0.9 | 54.1 ± 1.1 | 0.2 | 56.6 | 19 | 31 |
|         | CB   | 32.9 ± 1.1 | 31.7 ± 2.4 | 1.2 | 29.2 | 19 | 30 |
|         | CD   | 178.6 ± 1.6 | 179.9 ± 0.3 | −1.3 | 179.7 | 16 | 2 |
|         | CG   | 33.8 ± 1.5 | 35.0 ± 1.7 | −1.2 | 33.8 | 14 | 3 |
| Glu     | C    | 174.0 ± 1.1 | 174.6 ± 1.2 | −0.6 | 176.9 | 10 | 22 |
|         | CA   | 54.6 ± 1.7 | 54.9 ± 1.2 | −0.3 | 57.3 | 12 | 31 |
|         | CB   | 32.6 ± 1.3 | 32.8 ± 2.2 | −0.3 | 30.0 | 12 | 27 |
|         | CD   | 182.5 ± 0.9 | 183.2 ± 0.8 | −0.7 | 182.3 | 8 | 3 |
|         | CG   | 36.5 ± 1.6 | 37.4 ± 0.5 | −0.9 | 36.1 | 9 | 4 |
| Gly     | C    | 171.5 ± 1.6 | 171.2 ± 1.2 | 0.2 | 173.9 | 18 | 72 |
|         | CA   | 45.9 ± 2.0 | 45.3 ± 1.0 | 0.7 | 45.4 | 19 | 89 |
| His     | C    | 173.4 ± 0.7 | 173.6 ± 1.6 | −0.2 | 175.3 | 12 | 5 |
|         | CA   | 52.0 ± 1.4 | 55.0 ± 0.7 | −3.1 | 56.5 | 11 | 6 |
|         | CB   | 32.8 ± 1.5 | 32.0 ± 1.7 | 0.9 | 30.3 | 11 | 6 |
|         | CD2  | 117.4 ± 3.4 | 121.3 ± 0.7 | −3.9 | 120.3 | 10 | 3 |
|         | CE1  | 139.8 ± 5.0 | 136.8 ± 1.3 | 2.9 | 137.6 | 7 | 2 |
|         | CG   | 134.6 ± 2.8 | 133.2 ± 3.6 | 1.4 | 132.2 | 10 | 3 |
| Ile     | C    | 174.5 ± 1.1 | 173.8 ± 1.4 | 0.7 | 176.0 | 23 | 19 |
|         | CA   | 59.6 ± 1.1 | 59.6 ± 1.3 | 0.0 | 61.7 | 25 | 29 |
|         | CB   | 40.8 ± 1.8 | 41.0 ± 1.8 | −0.3 | 38.6 | 24 | 27 |
|         | CD1  | 14.1 ± 1.0 | 14.5 ± 0.8 | −0.3 | 13.4 | 21 | 14 |
|         | CG1  | 27.4 ± 0.6 | 28.3 ± 1.1 | −1.0 | 27.7 | 20 | 6 |
|         | CG2  | 17.8 ± 0.8 | 18.7 ± 0.6 | −0.9 | 17.5 | 22 | 6 |
| Leu     | C    | 174.9 ± 1.5 | 175.0 ± 1.3 | −0.1 | 177.1 | 17 | 41 |
|         | CA   | 54.3 ± 1.3 | 53.5 ± 0.9 | 0.8 | 55.7 | 18 | 66 |
|         | CB   | 44.5 ± 1.6 | 45.7 ± 1.8 | −1.1 | 42.2 | 17 | 61 |
|         | CD1  | 25.1 ± 1.8 | 25.1 ± 1.0 | 0.1 | 24.6 | 16 | 33 |
|         | CD2  | 24.9 ± 2.6 | 25.4 ± 1.0 | −0.5 | 24.1 | 14 | 26 |
|         | CG   | 28.5 ± 1.4 | 28.1 ± 0.7 | 0.4 | 26.8 | 15 | 12 |
Table 3 (continued)

| Residue | Atom | Average chemical shift of β-sheet residues | BMRB average CS (Ulrich et al. 2008) | Count |
|---------|------|------------------------------------------|--------------------------------------|-------|
|         |      | Fibrils | Barrels | ΔCS | Fibrils | Barrels |
| Lys     | C    | 174.0 ± 1.2 | 174.8 ± 1.1 | − 0.8 | 176.7 | 13      | 15      |
|         | CA   | 55.4 ± 1.5  | 54.8 ± 1.1  | 0.6  | 57.0  | 13      | 30      |
|         | CB   | 36.2 ± 1.8  | 35.4 ± 1.8  | 0.8  | 32.8  | 12      | 26      |
|         | CD   | 30.0 ± 0.8  | 30.2 ± 1.5  | − 0.1 | 29.0  | 8       | 4       |
|         | CE   | 42.1 ± 0.4  | 42.1 ± 0.6  | − 0.1 | 41.9  | 7       | 4       |
|         | CG   | 25.7 ± 0.4  | 24.2 ± 1.9  | 1.6  | 24.9  | 10      | 4       |
| Met     | C    | 173.9 ± 0.8 | 174.2 ± 0.8 | − 0.3 | 176.3 | 6       | 8       |
|         | CA   | 54.2 ± 1.4  | 54.0 ± 0.7  | 0.2  | 56.2  | 6       | 13      |
|         | CB   | 36.9 ± 1.1  | 36.4 ± 1.3  | 0.5  | 32.9  | 6       | 12      |
|         | CE   | 17.5 ± 0.6  | 16.8 ± 1.7  | 0.7  | 17.1  | 5       | 4       |
|         | CG   | 32.1 ± 0.6  | 32.6 ± 0.8  | − 0.5 | 32.0  | 5       | 4       |
| Phe     | C    | 173.7 ± 1.1 | 173.9 ± 1.6 | − 0.2 | 175.5 | 13      | 28      |
|         | CA   | 55.4 ± 2.3  | 56.6 ± 1.9  | − 1.2 | 58.1  | 13      | 44      |
|         | CB   | 42.8 ± 1.7  | 42.2 ± 1.7  | 0.7  | 39.9  | 13      | 40      |
|         | CD1  | 130.8 ± 1.1 | 132.2 ± 0.3 | − 1.4 | 131.6 | 6       | 2       |
|         | CD2  | 131.6 ± 0.9 | 132.2 ± 0.3 | − 0.6 | 131.6 | 6       | 2       |
|         | CE1  | 131.4 ± 0.6 | 130.2 ± 1.7 | 1.1  | 130.7 | 4       | 3       |
|         | CE2  | 131.6 ± 0.8 | 130.2 ± 1.7 | 1.4  | 130.7 | 2       | 3       |
|         | CG   | 139.2 ± 2.0 | 139.5 ± 0.5 | − 0.4 | 138.3 | 9       | 9       |
|         | CZ   | 129.1 ± 1.1 | 129.1       | 0.0  | 129.2 | 4       | 1       |
| Pro     | C    | 175.7 ± 1.3 | 175.8 ± 1.0 | − 0.1 | 176.8 | 2       | 9       |
|         | CA   | 62.5 ± 0.2  | 62.6 ± 0.7  | − 0.2 | 63.4  | 2       | 10      |
|         | CB   | 32.5 ± 0.2  | 33.5 ± 0.7  | − 1.0 | 31.8  | 2       | 10      |
|         | CD   | 48.4 ± 1.7  | 50.7 ± 0.9  | − 2.3 | 50.3  | 2       | 4       |
|         | CG   | 28.3 ± 0.4  | 28.0 ± 0.6  | 0.2  | 27.2  | 2       | 4       |
| Ser     | C    | 173.6 ± 2.4 | 172.9 ± 0.9 | 0.8  | 174.6 | 17      | 27      |
|         | CA   | 56.8 ± 1.3  | 57.0 ± 0.8  | − 0.3 | 58.7  | 15      | 41      |
|         | CB   | 65.6 ± 1.4  | 65.7 ± 1.3  | − 0.2 | 63.8  | 16      | 40      |
| Thr     | C    | 173.5 ± 1.3 | 172.5 ± 1.5 | 1.0  | 174.6 | 15      | 26      |
|         | CA   | 61.0 ± 1.0  | 60.3 ± 1.3  | 0.7  | 62.2  | 16      | 45      |
|         | CB   | 71.2 ± 1.2  | 70.9 ± 1.1  | 0.3  | 69.7  | 15      | 42      |
|         | CG2  | 21.6 ± 1.1  | 21.0 ± 2.0  | 0.6  | 21.5  | 14      | 7       |
| Tyr     | C    | 175.5 ± 3.2 | 173.6 ± 1.3 | 1.9  | 175.5 | 10      | 52      |
|         | CA   | 57.1 ± 1.6  | 56.5 ± 1.2  | 0.6  | 58.2  | 10      | 71      |
|         | CB   | 40.7 ± 2.1  | 41.8 ± 1.2  | − 1.1 | 39.3  | 10      | 71      |
|         | CD1  | 133.1 ± 0.8 | 132.7 ± 0.5 | 0.3  | 132.7 | 7       | 5       |
|         | CD2  | 133.2 ± 0.5 | 132.7 ± 0.5 | 0.5  | 132.7 | 7       | 5       |
|         | CE1  | 118.1 ± 0.5 | 117.9 ± 0.4 | 0.2  | 117.9 | 9       | 6       |
|         | CE2  | 117.9 ± 0.4 | 117.9 ± 0.4 | − 0.1 | 117.9 | 6       | 6       |
|         | CG   | 128.8 ± 2.0 | 130.3 ± 1.6 | − 1.4 | 129.7 | 9       | 12      |
|         | CZ   | 157.5 ± 1.1 | 158.8 ± 0.6 | − 1.3 | 156.7 | 8       | 4       |
| Val     | C    | 174.4 ± 1.6 | 173.8 ± 1.2 | 0.6  | 175.7 | 43      | 38      |
|         | CA   | 60.5 ± 1.1  | 60.2 ± 1.6  | 0.4  | 62.6  | 45      | 47      |
|         | CB   | 34.9 ± 1.4  | 34.9 ± 1.3  | 0.0  | 32.7  | 44      | 46      |
|         | CG1  | 21.4 ± 0.9  | 21.2 ± 1.2  | 0.1  | 21.5  | 40      | 15      |
|         | CG2  | 20.8 ± 1.3  | 21.6 ± 0.5  | − 0.8 | 21.3  | 34      | 11      |
depleted in cross-β fibrils, suggesting steric and electrostatic effects. In β-barrel membrane proteins, Val is less enriched relative to their abundance in all proteins, while Gly, Tyr and Phe are over-represented relative to their abundance in all proteins. These statistics suggest that the structural flexibility of Gly and the aromatic interactions of Tyr and Phe with lipids may be important for stabilizing β-barrels in lipid bilayers.

We analyzed 17 amyloid fibrils and 7 β-barrel structures for which high-resolution structures and chemical shift data are available (Table 1). These amyloid proteins and β-barrels contain 683 and 985 residues, respectively, among which 373 (55%) residues in fibrils and 521 (53%) residues in β-barrels are located in β-strands. The remaining residues lie in loops, turns, or short helices. All residues are used for parsing the (φ, ψ) torsion angles and (χ₁, χ₂) rotameric angles, but only residues with reported chemical shifts can be used for constructing the 2D 13C–13C correlation maps. To illustrate the structural categories analyzed here, Fig. 2 depicts the structures of three amyloid proteins, Aβ42, Osaka Aβ40, and glucagon; and two β-barrels, VDAC-1 and OmpG. Glucagon is an example of a long and straight antiparallel hydrogen-bonded β-strand with alternating dry steric-zipper residues and water-exposed residues (Fig. 2a). Aβ42 and Osaka Aβ40 form parallel-in-register β-sheets where the two protofibrils contain β-strands interspersed by disordered turns (Fig. 2b, c). VDAC-1 and OmpG are two β-barrel proteins containing 19 and 14 β-strands, respectively. Each β-strand has a pore-facing side and a lipid-facing side (Fig. 2e), which are preferentially enriched in polar and hydrophobic residues, respectively. Table 2 lists the number of each amino acid in the chemical shift dataset, broken down according to the six structural categories. The percentages of residues in each structural environment for Val, Leu, Gly, Ala, Gln, and Arg are also shown in Fig. 3.

We present the 13C chemical shifts in 2D 13C–13C correlation maps for thirteen amino acids. These amino acids are chosen for their high abundance in these β-sheet proteins, with either at least 50 occurrences in the combined dataset or with a high prevalence among either amyloid fibrils or β-barrel β-sheets. In our analysis, we consider chemical shift differences of 0.5 ppm or larger to be significant based on the typical 13C linewidths of solid-state proteins. Figure 4 shows the 2D 13C–13C correlation map of Val. Val exhibits a narrower Cα chemical shift distribution but a larger Cγ chemical shift distribution in amyloid fibrils compared to β-barrels (Fig. 4b). The Cα chemical shift standard deviation (σCα) is 1.1 ppm in fibrils and increases to 1.6 ppm in β-barrels (Table 3), suggesting that the extended hydrogen-bonding in cross-β fibrils narrows the Val backbone conformational distribution compared to β-barrel Val residues. In contrast, the Val Cγ2 chemical shift distribution is much wider (1.3 ppm) in cross-β fibrils than in β-barrels (0.5 ppm). The former is mostly contributed by dry β-sheet residues, suggesting that the participation of Val in steric zippers in amyloid fibrils increases the static conformational disorder of the sidechain.

Figure 5 displays the 2D 13C–13C chemical shift correlation maps of the two other methyl-rich hydrophobic residues, Leu and Ile. In contrast to Val, Leu shows much larger Cα, Cγ and Cδ chemical shift dispersions in amyloid fibrils than in β-barrels. The difference is mainly manifested by dry fibril residues and lipid-facing residues in β-barrels. This observation suggests that water exposure in either protein leads to similar averaged sidechain conformations, whereas the dry fibril interior, including steric zippers, creates larger static conformational disorder compared to lipid-facing residues. For Ile, the Cγ1 methyl chemical shift is more narrowly distributed in fibrils (0.6 ppm) than in β-barrels (1.1 ppm) (Fig. 5b), and the Cγ1 and Cγ2 chemical shifts are 0.9–1.0 ppm shifted upfield in amyloid fibrils than in...
β-barrels (Table 3). Thus, sidechain conformational differences exist between the two types of proteins, which may be caused by sidechain packing in cross-β amyloid fibrils versus protein-lipid interactions in β-barrels.

The $^{13}$C chemical shift distributions of the small Gly and Ala residues also differ between amyloid fibrils and β-barrels (Fig. 6), with fibrils exhibiting larger chemical shift dispersion than β-barrels. Gly residues in both the β-strand and turn regions of fibrils contribute to the Cα and CO chemical shift dispersion, whereas the β-barrel Gly Cα and CO chemical shifts are tightly clustered, especially for water-facing Gly residues (Fig. 6a). These trends indicate that the amphipathic β-strands in β-barrels, sandwiched by lipids on one side and a water-filled pore on the other, constrain the backbone conformation of Gly more than the cross-β fibril. Indeed, Gly in amyloid fibrils predominantly appear in flexible loop regions (Fig. 3c), whereas in β-barrels more than half of the Gly residues are located in β-sheet segments.
This could be due to the unusually long β-strands found in β-barrels, which cause backbone conformational strains that is alleviated by the flexible Gly residues. Ala also exhibits larger chemical shift dispersions for all carbons in fibrils than in membrane-bound β-barrels (Fig. 6b). For Ala Cβ, even when chemical shifts from turn residues are excluded, the majority of the β-sheet residues in fibrils display a larger chemical shift dispersion than in β-barrels.

Compared to the small Gly and Ala, the bulky aromatic Phe exhibits distinct chemical shifts between the cross-β fibrils and membrane-bound β-barrels. Among the sidechain carbons, Phe Cγ chemical shift is significantly more distributed in fibrils than in barrels, with a standard deviation of 2.0 ppm in fibrils and 0.5 ppm in β-barrels (Table 3). These differences indicate that the Phe sidechain conformation, dictated by the χ1 torsion angle, has a larger static disorder in amyloid fibrils, which is likely caused by aromatic stacking in the dry steric zipper interface. In comparison, Phe sidechains in membrane-bound β-barrels may undergo significant conformational motion, thus giving narrowly clustered aromatic 13C chemical shifts.

Figure 7 compares the 13C chemical shift correlation maps of six polar residues, including Glu, Gln, Asn, Ser, Thr and Tyr. Glu 13C chemical shifts do not display significant differences between fibrils and barrels. In contrast, the Gln Cβ chemical shifts are much more narrowly distributed in amyloid fibrils than in barrels, as shown by the Cβ–Cγ and Cα–Cγ correlation peaks: the σCβ value is 1.1 ppm for fibril Gln residues and increases to 2.4 ppm for β-barrel Gln residues (Table 3, Fig. 7b). Moreover, the mean Cβ, Cγ and Cδ chemical shifts deviate by ~1.2 ppm between fibril and barrel Gln residues. Gln residues play a key role
in steric zippers due to the hydrogen-bonding ability of the amide sidechain, which stabilizes the β-strand interface. Canonical steric zippers such as GNNQQNY in the yeast prior protein Sup35, and VQIINK and VQIVYK in the tau protein, all contain Gln residues (Nelson et al. 2005; Sawaya et al. 2007). Thus, the narrow Cβ chemical shift distribution of Gln in fibrils suggests preferential rotameric conformation and/or hydrogen-bonding of the Gln sidechain. This sidechain order differs from the sidechain disorder of Val, although both amino acids are common in steric zippers of amyloid proteins. We examined the (χ₁, χ₂) torsion angle distributions of the amino acids in cross-β fibrils and β-barrels (Fig. 9) and did not find the Gln and Val rotamer distributions to be narrower in fibrils than in β-barrels. Thus, we attribute the narrow chemical shift distribution of Gln residues in amyloid fibrils to sidechain hydrogen bonding in steric zippers, while the larger Val chemical shift dispersion is attributed to the χ₁ torsional angle distribution. Interestingly, Asn, which also possesses an amide sidechain but is one CH₂ group shorter than Gln, and which also occurs frequently in steric zippers, exhibits a larger Cβ chemical shift distribution in fibrils (σCβ = 2.9 ppm) than in β-barrels (σCβ = 2.0 ppm) (Fig. 7c, Table 3). These chemical shift dispersions are mainly contributed by water-exposed Asn residues, suggesting that the shorter sidechain endows Asn with larger conformational freedom compared to Gln residues.

The hydroxy-bearing Ser and Thr residues exhibit different Cα chemical shift trends: the Ser Cα chemical shift is more distributed in amyloid fibrils than β-barrel membrane proteins, as seen in the Cβ–Cα correlation peaks (Fig. 7d), whereas Thr Cα and Cγ chemical shifts are more narrowly clustered in amyloid fibrils than in β-barrels (Fig. 7e). For the hydroxy-bearing aromatic Tyr, most 13C chemical shifts are more distributed in amyloid fibrils than in β-barrels (Fig. 7f). Interestingly, the average Tyr Cβ chemical shifts differ noticeably between fibrils and barrels: the former is 1.1 ppm smaller than the latter (Table 3, Fig. 7f). The Tyr χ₁ angles are similarly distributed in fibrils and barrels: both proteins exhibit a preference for the trans (180°) and -60° states over the +60° state (Fig. 8). Thus, at present we attribute this Cβ chemical shift difference to small backbone conformational differences between amyloid fibril Tyr residues and β-barrel Tyr residues.

The 2D 13C–13C correlation map of the cationic Arg shows an interesting trend where most carbons except for Cδ...
and Cζ have narrower chemical shift distributions in amyloid fibrils than in β-barrels (Fig. 9). We attribute the backbone conformational order of Arg in cross-β fibrils to the constraints of extensive β-strand hydrogen bonds, while the relative disorder of Arg in membrane proteins is attributed to the high energetic cost of inserting Arg into the hydrophobic portion of lipid bilayers (Moon and Fleming 2011). In comparison, the Arg guanidinium group can form bidentate complexes with lipid phosphate groups, stabilized by electrostatic attraction and hydrogen bonding. This salt bridge interaction is well documented for Arg-rich antimicrobial peptides based on distance measurements between Arg Cζ and lipid 31P (Su et al. 2009; Tang et al. 2007). This salt bridge interaction should narrow the conformational distribution of the end of the Arg sidechain in the lipid membrane, thus explaining the narrow Cδ and Cζ chemical shift distribution in β-barrels.

We summarize the 13C chemical shifts and the ensuing conformational trends of these 13 amino acids in β-sheet proteins as follows. Specifically, we focus on the static conformational disorder, which is reflected by chemical shift distributions. First, bulky methyl-rich hydrophobic residues (Val, Leu and Ile) exhibit more ordered backbone but more disordered sidechain conformations in cross-β fibrils than in β-barrels. Second, the small Ala and Gly are more disordered in fibrils than in barrels. Third, aromatic Phe and Tyr residues have more disordered backbone and sidechain conformations in amyloid fibrils than in β-barrels. Fourth, the polar
Ser, Thr and Arg are more structurally ordered in fibrils than in barrels. Finally, Gln and Asn exhibit opposite sidechain conformational trends: Gln is more ordered in fibrils than β-barrels, whereas Asn is more disordered in amyloid fibrils.

One of the clearest chemical shift differences between amyloid fibrils and β-barrels is found for Val and Leu methyl carbons (Figs. 4, 5): fibrous Val and Leu residues, particularly those located at the dry β-strand interface, display much larger methyl $^{13}$C chemical shift distributions than β-barrel Val and Leu residues. Recent studies of several amyloid proteins, including glucagon and the tau protein (Dregni et al. 2019; Gelenter et al. 2019), reported a splitting in the Val methyl groups that are involved in steric zippers, which is absent from water-exposed Val residues. This observation indicates that the solvent-exposed Val sidechain undergoes fast rotameric jumps around the Cα–Cβ bond (i.e. the $\chi_1$ angle), thus averaging the $\gamma_1$ and $\gamma_2$ chemical shifts. In comparison, Val sidechains at the dry steric zipper interfaces...
are conformationally locked, thus leading to resolved Cγ1 and Cγ2 chemical shifts. This effect may also exist for Leu to account for its larger Cδ chemical shift distribution in fibrils than in β-barrels (Fig. 5a). Taken together, Val and Leu sidechains are conformationally more dynamic when they are exposed to either water or lipids as compared to when they reside at β-strand sidechain interfaces.

The conclusion that sidechains involved in steric zippers are more rigid and conformationally distributed than solvent-exposed sidechains is consistent with two studies of amyloid protein dynamics. For HET-s, backbone order parameters were measured using $^1$H–$^{15}$N and $^1$H–$^{13}$Cα REDOR experiments (Smith et al. 2016). The $^1$H–$^{13}$Cα REDOR data show order parameters of 0.8–0.9 for most
solvent-exposed β-strand residues but higher order parameters of 0.9–1.0 for most residues at the dry β-strand interface. This result is consistent with the chemical shift distribution found here. In comparison, the backbone N–H order parameters of β-sheet residues are similarly high, above 0.9. This can be attributed to the fact that the N–H dipolar couplings are dominated by hydrogen bonding along the fibril axis. Another study reported 1H–13Cα order parameters of Aβ40 fibrils (Scheidt et al. 2012) but did not detect a difference between dry and solvent-exposed residues. We tentatively attribute this finding to the packing of multiple protofilaments within the mature fibrils, which may partly immobilize the solvent-exposed β-sheet residues.

Conclusion

This survey of the chemical shifts of more than 2100 residues in amyloid proteins and β-barrel membrane proteins reveal several interesting conformational trends. We found that methyl-rich non-polar residues, polar residues containing sidechain amide and hydroxyl groups, aromatic residues, and small residues, exhibit chemical shift distributions and hence static conformational distributions that are distinct between amyloid fibrils and β-barrels. These chemical shift differences can be explained by sidechain hydrogen-bonding among Gln and Asn residues, van der Waals interactions between methyl-rich Val, Leu and Ile residues in cross-β fibrils, and water or lipid exposure in both types of proteins.
These chemical shift trends should be useful for guiding structural analysis of amino acid residues in these β-sheet proteins based predominantly on NMR chemical shifts.

Acknowledgements This work is supported by NIH Grant AG059661 to M.H. and an NIH Ruth L. Kirschstein Individual National Research Service Award (1F31AI133989) to M.D.G.

Data availability The complete chemical shift datasets analyzed in the current study are available from Mei Hong at meihong@mit.edu upon request.

References

Andreas LB et al (2016) Structure of fully protonated proteins by proton-detected magic-angle spinning NMR. Proc Natl Acad Sci USA 113:9187–9192
Bertini I, Gonnelli L, Luchinat C, Mao J, Nesi A (2011) A new structural model of Aβ fibrils. J Am Chem Soc 133:16013–16022
Colvin MT et al (2016) Atomic resolution structure of monomorphic Aβ42 amyloid fibrils. J Am Chem Soc 138:9663–9674
Consortium TU (2018) UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47:D506–D515
Dregni AJ et al (2019) In vitro 0N4R tau fibrils contain a monomorphic β-sheet core enclosed by dynamically heterogeneous fuzzy coat segments. Proc Natl Acad Sci 116:16357
Dutta SK, Yao Y, Marassi FM (2017) Structural insights into the Yersinia pestis outer membrane protein Omp in lipid bilayers. J Phys Chem B 121:7561–7570
Edrington TC, Kintz E, Goldberg JB, Tamm LK (2011) Structural basis for the interaction of lipopolysaccharide with outer membrane protein H (OprH) from Pseudomonas aeruginosa. J Biol Chem 286:39211–39223
Fitzpatrick AWP et al (2013) Atomic structure and hierarchical assembly of a cross-β amyloid fibril. Proc Natl Acad Sci 110:5468–5473
Gelenter MD et al (2019) The peptide hormone glucagon forms amyloid fibrils with two coexisting β-strand conformations. Nat Struct Mol Biol 26:592–598
Gremer L et al (2017) Fibril structure of amyloid-β(1–42) by cryo-electron microscopy. Science 358:116–119
Hagn F, Etzkorn M, Raschle T, Wagner G (2013) Optimized phospholipid bilayer nanodiscs facilitate high-resolution structure determination of membrane proteins. J Am Chem Soc 135:1919–1925
Han B, Liu Y, Ginzinger SW, Wishart DS (2011) SHIFTX2: significantly improved protein chemical shift prediction. J Biomol NMR 50:43
Hiller S, Garces RG, Malia TJ, Orekhov VY, Colombini M, Wagner G (2008) Solution structure of the integral human membrane protein VDAC-1 in detergent micelles. Science 321:1206–1210
Horst R, Stanczak P, Wüthrich K (2014) NMR polypeptide backbone conformation of the E. coli outer membrane protein W. Structure 22:1204–1209
Johansson MU, Alioth S, Hu K, Walser R, Koebnik R, Pervushin K (2007) A minimal transmembrane beta-barrel platform protein studied by nuclear magnetic resonance. Biochemistry 46:1128–1140
Laskowski RA, Rullmann JA, MacArthur MW, Thornton J (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26:283–291
Laskowski RA, Rumlmann JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8:477–486
Lee M et al (2017) Zinc-binding structure of a catalytic amyloid from solid-state NMR. Proc Natl Acad Sci 114:6191–6196
Liang B, Tamm LK (2007) Structure of outer membrane protein G by solution NMR spectroscopy. Proc Natl Acad Sci 104:16140–16145
