

SCIENTIFIC REPORTS



OPEN

Uncovering the Networks of Topological Neighborhoods in β -Strand and Amyloid β -Sheet Structures

Luhan Zhai, Yuko Otani  & Tomohiko Ohwada 

Although multiple hydrophobic, aromatic π - π , and electrostatic interactions are proposed to be involved in amyloid fibril formation, the precise interactions within amyloid structures remain poorly understood. Here, we carried out detailed quantum theory of atoms-in-molecules (QTAIM) analysis to examine the hydrophobic core of amyloid parallel and antiparallel β -sheet structures, and found the presence of multiple inter-strand and intra-strand topological neighborhoods, represented by networks of through-space bond paths. Similar bond paths from side chain to side chain and from side chain to main chain were found in a single β -strand and in di- and tripeptides. Some of these bond-path networks were enhanced upon β -sheet formation. Overall, our results indicate that the cumulative network of weak interactions, including various types of hydrogen bonding (X-H \cdots Y; X, Y = H, C, O, N, S), as well as *non-H-non-H* bond paths, is characteristic of amyloid β -sheet structure. The present study postulated that the presence of multiple through-space bond-paths, which are local and directional, can coincide with the attractive proximity effect in forming peptide assemblies. This is consistent with a new view of the van der Waals (vdW) interactions, one of the origins of hydrophobic interaction, which is updating to be a directional intermolecular force.

In addition to amide hydrogen bonding, multiple interactions such as hydrophobic, aromatic π - π , and electrostatic interactions are proposed to be involved in forming peptide assemblies¹. In Alzheimer's disease (AD) and related diseases, amyloid- β peptide (A β) forms oligomers and amyloid fibrils consisting of β -sheet structures², which are the main component of the characteristic neuritic plaques³. The amyloid fibers are heterogeneous aggregates of highly ordered and stable proteins, which share common structural and staining characteristics, but appear to have little sequence homology¹. Much attention has been focused on understanding what drives particular peptide sequences to aggregate and how the amyloid peptide self-assembles. While hydrogen bonding is crucial, uncovering specific interactions within the amyloid structure is expected to throw light on the mechanisms of fibril formation and fibril stability. Among the A β isoforms, A β 40 and A β 42 are the most abundant⁴. A β 40 contains many hydrophobic (Val, Ile, etc.) and aromatic (Phe and Tyr) amino acids, and it has been suggested that multiple hydrophobic, aromatic, and electrostatic interactions are involved in amyloid fibril formation³, and that the self-assembly of amyloid peptides is mainly governed by non-covalent interactions, including hydrogen bonds, coulombic interactions and hydrophobic effects⁵⁻⁹. Aromatic π - π interactions were also proposed to contribute significantly to amyloid aggregation¹⁰⁻¹⁵, but other studies have failed to identify specific interactions involving π -electrons or aromatic character as forces that stabilize the whole fibril^{16,17}. Therefore, the actual mechanism of amyloid aggregation is still controversial. However, it has been established that the hydrophobic aromatic core of A β 40 (17-20) (Leu¹⁷-Val¹⁸-Phe¹⁹-Phe²⁰) is an important target for pharmaceutical inhibition of A β neurotoxicity in Alzheimer's disease¹⁸.

Amino acids bearing a branched side chain, such as Val, Ile, Thr, and Cys, and aromatic amino acids, such as Tyr, Trp and Phe, have high propensities to form β -strand structure, whereas Ala, Gly and Pro have poor propensities¹⁹⁻²². The former amino acids account for nearly 50% of the amino acids in A β 40 and A β 42. However,

Laboratory of Organic and Medicinal Chemistry, Graduate School of Pharmaceutical Sciences, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan. Correspondence and requests for materials should be addressed to T.O. (email: ohwada@mol.f.u-tokyo.ac.jp)

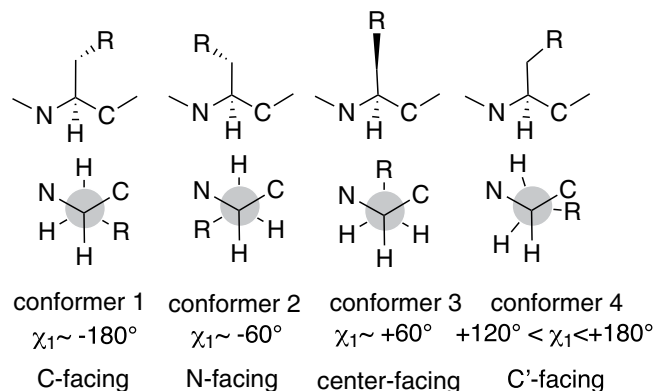


Figure 1. Configurations of the side chain.

these propensities are statistical in nature, and more rational explanations for the differences in intrinsic β -strand-forming propensities of these amino acids remain to be explored²³.

Bader's quantum theory of atoms in molecules (QTAIM)²⁴ is the topological analysis of the molecular electron density based on zero-flux surfaces. This partitioning of electron density by means of the QTAIM is well defined and can be applied to electron-density distributions obtained from both experiment and theory. A bond path²⁵ is a single line linking the two electron attracters (usually nucleus), minimum electron density linking the nuclei of two chemically bonded atoms^{24,26}. There is a minimum electron-density along the bond path, a bond critical point. The entire bond path runs from the first nucleus over the bond critical point to the second one. A bond path of strong interaction such as in a covalent bond almost always corresponds to a chemical bond. On the other hands, it has been repeatedly defined that bond paths are not chemical bonds, particularly in the cases of weak interactions (see also Methods: What is the bond path?)^{27–31}. While QTAIM has been applied to many types of compounds³², including peptides, detecting covalent bonding^{33–35}, and more recently to α -helix structures of peptides, which reasonably identified amide hydrogen-amide carbonyl oxygen hydrogen bonds³⁶, there have been few studies of β -strand and β -sheet structures relevant to amyloid- β peptide (A β). In this study, we employed QTAIM analysis to examine the hydrophobic and aromatic amino acids of A β -peptides, and found multiple inter-strand and intra-strand bond-path networks between specified atom pairs in the hydrophobic and aromatic amino acids of A β -peptides. The former inter-strand bond-path networks, representing topological neighborhoods, are characteristic of β -strand/sheet structures, and are never found in α -helical structure³⁶. We propose that intrinsic ability to form multiple through-space bond paths is relevant to β -strand/sheet structures that may promote inter-strand association of A β peptides. Indeed, some of these bond-path networks were found to be enhanced upon β -sheet formation.

Methods

Conformations of peptides. The initial conformations of various dipeptides (Ac-X-NHMe, X = amino acid), tripeptides, a single β -strand, parallel and antiparallel β -sheets of A β 40 were obtained from conformation search or a protein crystallographic database.

For dipeptides, the initial PPII (polyproline II helix) and β structures of the dipeptide were obtained by means of Monte Carlo torsional sampling conformational search with OPLS3 force field in the Maestro program (Schrodinger, Inc., USA.)³⁷, with a 10 kcal/mol energy window in water and in chloroform. Ten amino acid dipeptides (Ac-X-NHMe: X = Ala, Pro, Leu, Val, Ile, Thr, Cys, Phe, Tyr, and Trp) were calculated. The initial conformers (10–20 conformers for PPII and β structures, respectively), obtained by the conformation search were subjected to geometry full-optimizations and frequency calculations at the M06-2X/6-31 + G(d) level, using the Gaussian 09 and Gaussian 16 suites of programs³⁸. The M06-2X/6-31 + G(d) level method was used previously and this calculation method was demonstrated to correctly predict the backbone conformational populations of proline dipeptide and its derivatives in water and in chloroform³⁹. Therefore we adopted this method in this work. Solvent effects were evaluated by optimizing each conformer using an implicit solvent model, SMD solvation model⁴⁰, in water and in chloroform. The lack of negative frequencies confirmed that all conformers truly represent energy minima. The structures obtained by the conformation search and those obtained by the following DFT optimization were essentially consistent. The DFT-optimized structures were compared and could be converged into 2–4 conformers. These conformers are based on the different rotations of the side chain, distinguished by the dihedral angle χ_1 ($\angle \text{N-C}\alpha\text{-C}\beta\text{-X}\gamma$; X = heteroatom (not H)) (see Fig. 1).

The initial structures of tri-peptides were taken from the crystallographic data in the PDB database. Capping of the N-terminal (with acetyl) and C-terminal (with NHMe) was automatically executed in the Maestro software, which can mimic the peptide main chain. The tri-peptides examined in this work were Ac-Val-Val-Val-NH₂ (VVV from the β -strand sequence V⁶⁶-V⁶⁸ of PDB entry 1EST)⁴¹ Ac-Leu-Val-Phe-NH₂ (LVF from the β -strand sequence L¹⁰¹-F¹⁰³ of PDB entry 1AXC)⁴², Ac-Ile-Thr-Tyr-NH₂ (ITY from the β -strand sequence I²⁰⁵-Y²⁰⁷ of PDB entry 2PEC)⁴³, and Ac-Leu-Val-Ile-NH₂ (LVI from the β -strand sequence L⁵²-I⁵⁴ of PDB entry 2PEC)⁴³, which were compared with the PPII structures found in Ac-Gly-Pro-Ala-NH₂ (GPA from the PPII sequence G²⁴-A²⁶ of PDB entry 451C)⁴⁴ and Ac-Gly-Pro-Ser-NH₂ (GPS from the PPII sequence G¹²³-S¹²⁵ of PDB entry 3FZU)⁴⁵. The selection of the β -strand sequences, VVV, LVF, ITY, LVI, was based on their availability in the X-ray crystal

structure database. PDB entry 1EST (porcine pancreatic elastase)⁴¹ and 2PEC (tri β -strand coiled folds of pectate lyases) are classified as “all beta proteins”⁴³. PDB entry 1AXC (C-terminal region of p21^{WAF1/CIP1} complexed with Human PCNA) is classified as “alpha and beta proteins”⁴². The PPII sequences, **GPA**, and **GPS** were used for comparison with the β -strand sequences. PDB entry 451C (Cytochrome C_{551} from *Pseudomonas aeruginosa*) is classified as “all alpha proteins”⁴⁴. PDB entry 3FZU (IgG1) is classified as “mainly beta”⁴⁵. All the crystal structures were relaxed by energy minimization with force field OPLS3 in Maestro, followed by DFT structure optimization calculations at the M06-2X/6-31 + G(d) level with the SMD solvation model in water and in chloroform⁴⁰. These solvents were selected on the basis of the different dielectric constants, which would influence the folding properties of peptides. The structures obtained by energy minimization with force field or those obtained by the DFT optimization were almost identical with the initial crystallographic structures (except **ITY** in water; the structure **ITY** changed).

The single strand structure was taken from F⁶⁴-V⁶⁸ of PDB entry 1EST⁴¹. Parallel β -sheet and anti-parallel β -sheet structures were extracted from A β 40 (PDB 2LMP: K¹⁶-A²¹)⁴⁶ and A β 40 (PDB 2LNQ_A: K¹⁶-N²², 2LNQ_B: K¹⁶-Q²²)⁴⁷, respectively. For the single strand and β -sheet structures, we directly used the structures extracted from the protein crystal structures for the QTAIM analysis.

QTAIM calculations. Bond-path analysis was performed at the Slater-type triple-zeta-polarization (TZP) level with ADF (SCM, Netherlands)⁴⁸. A polarization function is added for H through Ar and for Ga through Kr. The detail of optimization of Slater-type basis sets was reported previously⁴⁹. The QTAIM analysis was applied on the basis of the DFT-optimized energy-minimum structures and PDB database structures.

What is the bond path? The concepts of bond paths and bond critical points have been criticized^{50–54} and a rebuttal published²⁷. There have been also arguments of the presence/absence and the interpretation of H—H bond path in different systems (see Results and Discussion)^{28,55–57}. A bond path is often misidentified with a chemical bond, but bond paths have been regarded as indicative of *bonded interactions*, not chemical bonds, which are claimed to encompass all kinds of interactions²⁸. Usually covalent bonding corresponds well to the bond path. One of the different interpretations of a bond path of *weak interactions* is that “simply allowing two atoms to approach each other should often cause electron density to flow to the interatomic space, depending on the balance between nucleus-electron attraction and electron-electron repulsion, both coulomb and exchange”^{54,55}. QTAIM atoms are not simple spheres thus their neighborhood can be influenced by their complex topology. This can lead to a bond path with a bond critical point, even in the purely classical case in which exchange is not considered^{29–31}. This analysis suggests that the occurrence of a bond critical point should depend on the interatomic distance^{53,54}. However, in the present cases, the H-H bond path we detected in the Leu dipeptides did not always show the closest set of the two atoms: for example, the distance between the two atoms (H-H(N)) connected by a bond path in the PPII-2 structure in water is 2.347 Å while the distance between the corresponding same atom pair, H_a-H_b(N) in the PPII-2 structure in water is 2.233 Å, much shorter than the former (see Fig. 2(a)). However this is no bond path between the relevant H_a-H_b(N) atom pair. This indicated that *geometrical proximity* is not only a factor for the presence of a bond path. Therefore, bond paths, including the present H—H interactions observed in the Leu dipeptide and other dipeptides studied (vide infra) are consistent with the *topological proximity* of these relevant atoms and also indicative of weak interactions at least in the present peptide system.

Results and Discussion

Conformational preferences and bond paths of dipeptide. It has been reported that amino acids that favor β -strands in proteins are side-chain branched such as in Leu, Val, Ile, Thr, and Cys and aromatic such as in Phe, Tyr and Trp²². Thus, we focused on amino acids having a branched side chain or aromatic side chain, such as Leu, Val, Ile, Thr, Cys, Phe, Tyr and Trp. We showed the data of Lue, Phe and Trp dipeptides in the main text and showed the other data of Val, Ile, Thr, Cys and Tyr in Supplementary Data (see Supplementary Figs 3–7). The results of other dipeptides of Ala and Pro were also shown in Supplementary Data (see Supplementary Figs 1 and 2).

First, we studied through-space weak interactions in the β -strand conformations of the ten short amino acid dipeptides (Ac-X-NHMe, X = Ala, Pro, Leu, Val, Ile, Thr, Cys, Phe, Tyr, and Trp) in terms of bond paths. In order to examine whether the interaction is β -strand-specific or not, we compared the β -strand structure with the PPII structure of the same molecule. Several accessible conformers are based on the different rotations of the side chain, distinguished by the dihedral angle χ_1 (°) (Fig. 1, see also Supplementary Table 1): conformer 1: C-terminal-facing, dihedral angle $\chi_1 \sim -180^\circ$; conformer 2: N-terminal-facing, dihedral angle $\chi_1 \sim -60^\circ$; conformer 3: center-facing, dihedral angle $\chi_1 \sim 60^\circ$; and conformer 4: C'-terminal-facing, dihedral angle $120^\circ < \chi_1 < 180^\circ$. Both of conformer 1 (C-terminal-facing) and conformer 4 (C'-terminal-facing) are classified into C-terminal facing.

Supplementary Table 1 shows calculated main chain torsion angles (Φ and Ψ) of the local energy minimum structure of ten dipeptides obtained in water and in chloroform solvent environments after optimization by the DFT method. Energy differences of each conformer were also compiled in Supplementary Table 1. The energy difference is insignificant (in most cases, within 1.0–0.5 kcal/mol) and these energies are essentially based on enthalpy not Gibbs free energy. These conformers are in equilibrating.

Leu dipeptide. In the case of Leu, two PPII structures (conformer 1, PPII-1(C-facing) and conformer 2, PPII-2(N-facing)) with different directions of the side chain are identified in water (Fig. 2(a)). While the conformer 2 of PPII-2(N-terminal-facing) structure contained no through-space bond path, in the conformer 1 of PPII-1 structure (C-facing), there is a through-space bond path between methyl-H of the side chain and the main chain N-H (Fig. 2(a)). Consequently the cyclic structure was formed, which make a ring critical point (green

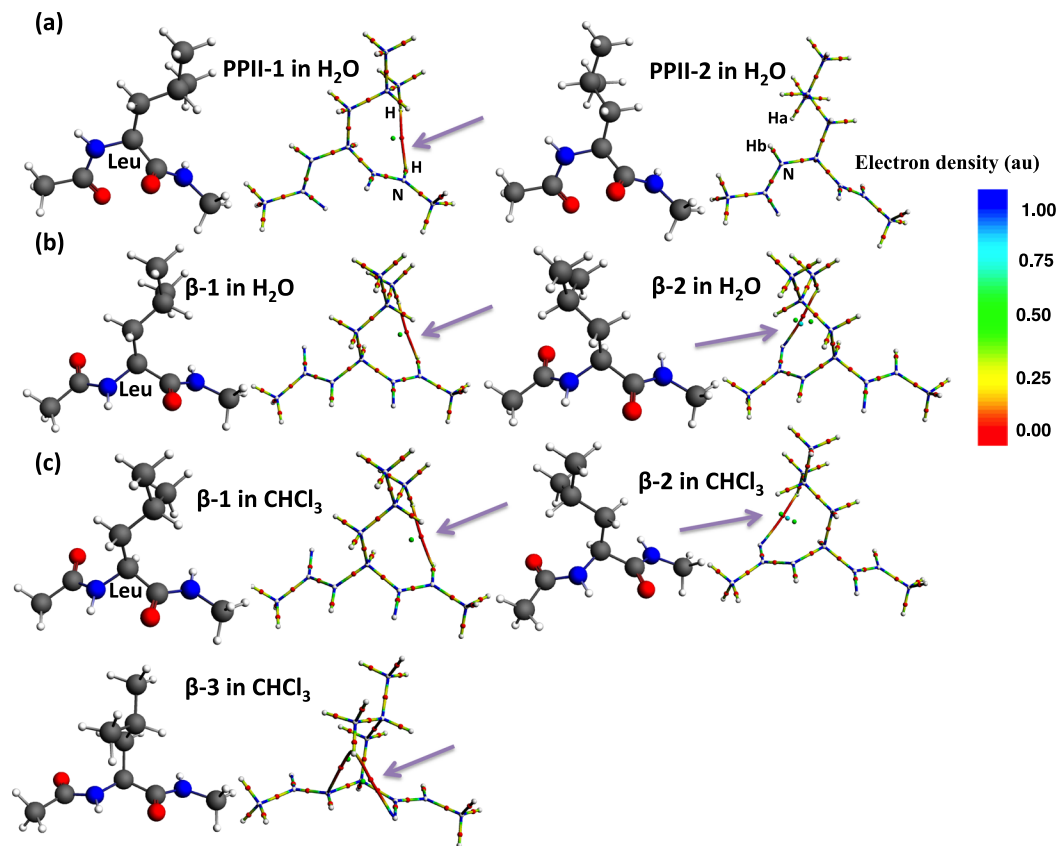


Figure 2. Leu dipeptide (Ac-Leu-NHMe) in water and in chloroform.

small ball) (Fig. 2(a)). Two β-strand structures (conformer 1, β-1(C-facing) and conformer 2, β-2(N-facing)) with different directions of the side chain are also obtained in water, and three β-strand structures (conformer 1, β-1(C-facing); conformer 2, β-2(N-facing); and conformer 3, β-3(center-facing)) are obtained in chloroform (Fig. 2(b,c)). In C-facing β-1 structures detected in water and in chloroform (Fig. 2(b,c)), the side chain is rotated to the C-terminal side of Leu and there is a bond path between methyl-H of the side chain and the main chain (N)-H in both solvents. As for the N-facing β-2 structures obtained in water and in chloroform (Fig. 2(b,c)), when the side chain is rotated to the N-terminal side, a bond path between methyl-H of the side chain and the acetyl carbonyl oxygen (O) is found. Furthermore, in center-facing β-3 in chloroform (Fig. 2(c)), when the side chain is rotated to the middle position, two bond paths are generated: the methyl-H of the side-chain and the carbonyl oxygen atom (O) of the C-terminal, and the same methyl-H and N atom, forming a cyclic bonding.

In the ball-and-stick model, the colors have the following meaning: red: oxygen; blue: nitrogen; dark grey: carbon; white: hydrogen. In the molecular graph (QTAIM) the colors have the following meaning: a color-gradation line: accumulation of electron density (bond path for covalent bonding); red line: through-space weak bond path (with a purple arrow); red small ball: bond critical point, green small ball: ring critical point or cage critical point.

Phe dipeptide. For the aromatic amino acid Phe dipeptide, three PPII structures with the side chain (PPII-1, PPII-2 and PPII-3) directed to the C-terminal, N-terminal and middle are obtained in water (Fig. 3(a)). Two β-strand structures (β-3 and β-4 in water, β-1 and β-3 in chloroform) with the different side chain rotation are obtained in water and chloroform, respectively (Fig. 3(b,c)). No through-space bond path is detected in any of these structures.

Trp dipeptide. On the other hands, for another aromatic amino acid dipeptide, Trp, three PPII structures (PPII-2, PPII-3 and PPII-4) with different side chain directions are obtained in water (Fig. 4(a)). Only the C'-facing PPII-4 structure in water contains two bond paths. One is between indole ring-H of the side chain and the main chain carbonyl oxygen (O), and the other is a C-H—π bond path between the indole ring and H of C-terminal methyl (Fig. 4(a)). Four kinds of β-strand structures (β-1, β-2, β-3 and β-4) with different side chain directions are obtained in water (Fig. 4(b)). In center-facing β-3 structure in water with the side chain rotated to the middle (Fig. 4(b)), there is a bond path between the indole ring-H and the N-terminal acetyl carbonyl carbon (C). In C'-facing β-4 structure in water, there are two bond paths between the indole ring-H with the N-terminal acetyl carbonyl oxygen and C-H—π bond path between the indole ring and H of C-terminal methyl. However, no bond path is found in C-facing β-1 structure or N-facing β-2 structure in water. One center-facing PPII-3 structure is obtained in chloroform and no bond path is found (Fig. 4(c)). Two β-strand structures (β-1 and β-3) appear in chloroform (Fig. 4(d)). In C-facing β-1 structure in chloroform, when the side chain is rotated to the C-terminal side, there are two bond paths. One is an intramolecular hydrogen bonding. The other is N-H—π

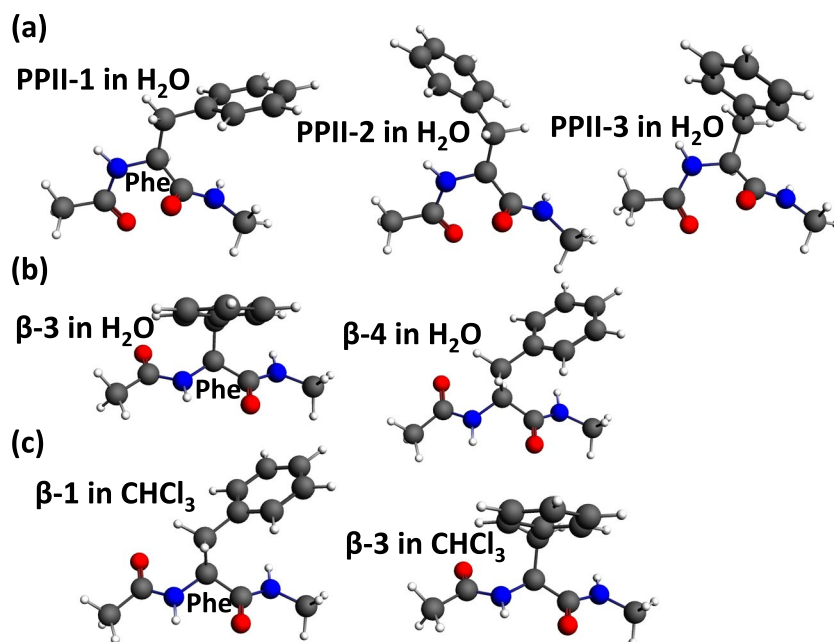


Figure 3. Aromatic dipeptides, Phe dipeptide in water and in chloroform.

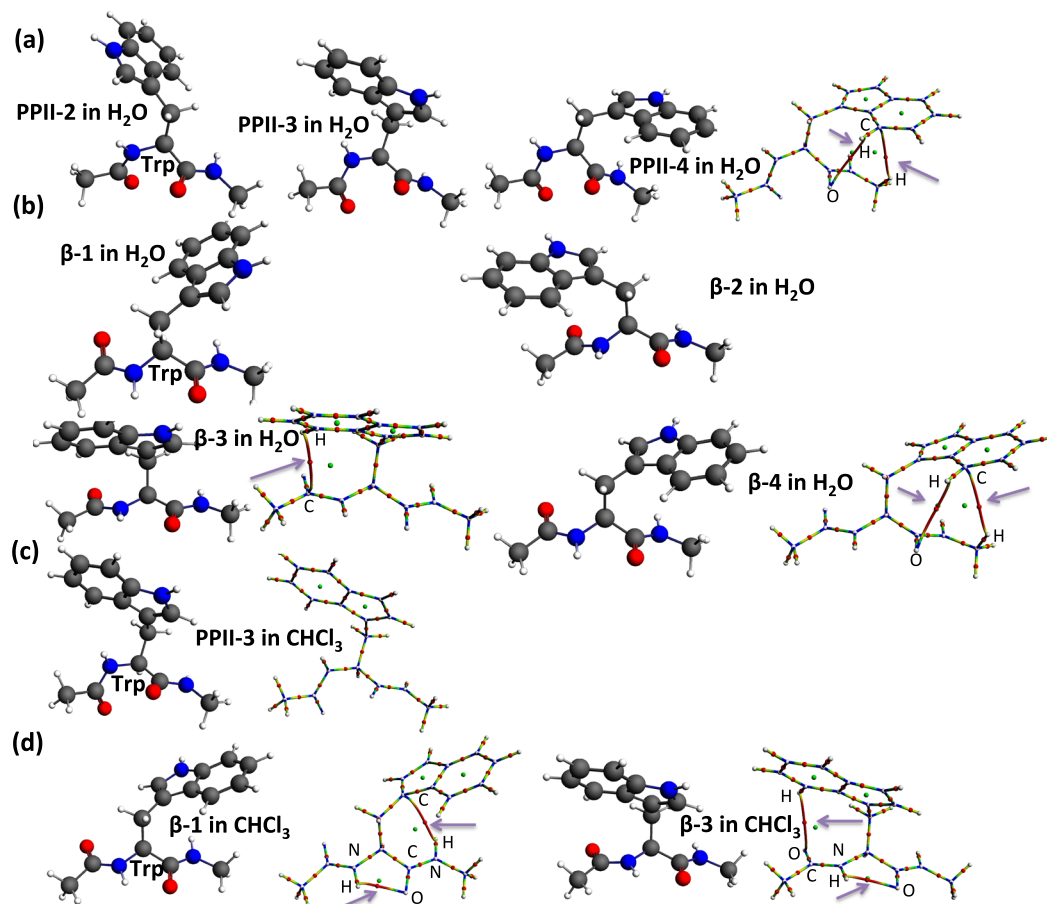


Figure 4. Aromatic dipeptides, Trp dipeptide in water and in chloroform.

interaction between the indole ring and the C-terminal H(-N) (Fig. 4(d)). In center-facing β -3 structure in chloroform, the side chain is rotated to the middle and there are two bond paths. One is an intramolecular hydrogen bond between carbonyl-O and H(-N). The other is between indole ring-H and acetyl carbonyl oxygen (O).

Entry	β -Strand of dipeptide	Atom pair	Distance (Å)	Electron density at the bcp (a.u.) ^a
1	Leu β -1 in H ₂ O	C-H...H-N	2.365	0.0056
2	Leu β -2 in H ₂ O	C-H...O=C	2.747	0.0065
3	Leu β -1 in CHCl ₃	C-H...H-N	2.329	0.0059
4	Leu β -2 in CHCl ₃	C-H...O=C	2.595	0.0083
5	Val β -3 in H ₂ O	C-H...N-C	2.761	0.0089
6	Val β -2 in CHCl ₃	C-C...O=C	3.279	0.0064
7	Val β -3 in CHCl ₃	C-H...N-C	2.713	0.0096
8	Ile β -1 in H ₂ O	C-H...H-N	2.347	0.0054
9	Ile β -1 in H ₂ O	C-H...N	2.769	0.0085
10	Ile β -3 in H ₂ O	C-H...O=C	2.827	0.0067
11	Ile β -1 in CHCl ₃	C-H...N	2.698	0.0096
12	Ile β -2 in CHCl ₃	C-H...O=C	2.778	0.0071
13	Thr β -1 in H ₂ O	N-H...O-C	2.432	0.0124
14	Thr β -2 in H ₂ O	O-H...O=C	1.845	0.0338
15	Thr β -1 in CHCl ₃	N-H...O-C	2.120	0.0206
16	Thr β -2 in CHCl ₃	O-H...O=C	1.837	0.0344
17	Thr β -3 in CHCl ₃	C-C...O=C	3.344	0.0076
18	Cys β -1 in CHCl ₃	N-H...S-C	2.709	0.0129
19	Trp β -3 in H ₂ O	C _{sp²} -H...C-C	2.732	0.0066
20	Trp β -4 in H ₂ O	C _{sp²} -H...C=O	2.956	0.0046
21	Trp β -4 in H ₂ O	C-H...C (sp ₂ , π)	2.850	0.0059
22	Trp β -1 in CHCl ₃	N-H...C (sp ₂ , π)	2.826	0.0106
23	Trp β -1 in CHCl ₃	N-H...O=C	2.245	0.0266
24	Trp β -3 in CHCl ₃	C _{sp²} -H...C=O	3.037	0.0045
25	Trp β -3 in CHCl ₃	N-H...O=C	2.203	0.0234
26	Hydrogen bond ^b	N-H...O (=C)	2.243	0.014
27	Hydrogen bond ^b	C-H...O (=C)	2.729	0.006

Table 1. Atom pair, bond length, and electron density of the bond paths in the β -strand structures of the selected dipeptides. ^abcp = bond critical point. ^bIn α -helix of peptides³⁶.

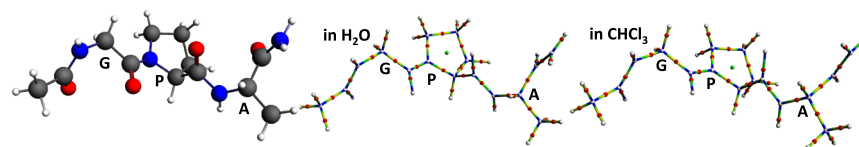
The data on atom pairs, bond lengths, and electron density of through-space bond paths found in the β -strand structures of the dipeptides (Leu, Val, Ile, Thr, Cys, Trp) are summarized in Table 1. We observed multi-center bonds which can be formed by small electron sharing between a single atom and many other atoms, or electron sharing between groups of atoms. It was proposed that multi-center bonding is not termed as bonded/non-bonded, rather neighboring/non-neighboring²⁹.

There are eight types of atom pairs involved in the bond paths observed in the dipeptides, that is, H...H, H...N, H...C, H...O, H...S, C...O and C-H... π -C and N-H... π -C pairs. The values of the electron densities at the bond critical points fall in the range of 0.0045–0.0344 (au). Among these bond paths, the C_{sp²}-H...O=C bond path (Trp β -3 in CHCl₃, entry 24 in Table 1) is the smallest and O-H...O=C bond path (Thr β -2 in CHCl₃, entry 16 in Table 1) is the largest in terms of accumulation of electrons at the bond critical points. In the case of an X-H bond path, for a similar distance of the atom pair, atoms (X) with strong negativity always form a relatively stronger bond path based on electron density. For example, compared with H...H and C...H bond paths, H...O, H...S and H...N bond paths are always of greater strength in terms of the electron density at the bond critical point.

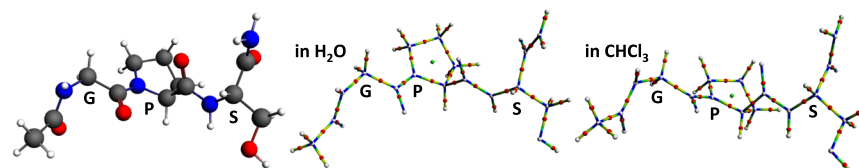
Based on the bond path study of amino acid dipeptides (Leu, Val, Ile, Thr, Ala, Pro, Phe, Tyr, Trp), bond paths can be observed in both PPII and β -strand structures. But, compared with PPII structures, bond paths are more extensively found in β -strand structures of the dipeptides. Because the bond path is located between the side chain and the main chain, we assume that the presence of this kind of side chain-main chain linking, that is, topological neighborhood, is a characteristic feature of β -strand structure, not of PPII structure. Amino acids bearing branched side chains often favor β -strand structure, which is consistent with the present observation that the side chain can form bond paths on both the N- and C-terminal of the main chain. In the present conformational search study, it proved very difficult to find PPII structure in chloroform, but we could obtain PPII structure in water. Therefore, in a water environment, both PPII and β -strand structure can be stabilized, while only β -strand structure can be stabilized in chloroform.

Conformational preferences and bond paths of tri-peptides. We next applied bond-path calculation to six selected tri-peptide sequences. All the initial structures are crystal structures isolated from protein crystals. Two are β -strand-disfavoring sequences (GPA from the PPII sequence G²⁴-A²⁶ of PDB entry 451C⁴⁴, GPS from the PPII sequence G¹²³-S¹²⁵ of PDB entry 3FZU)⁴⁵ and the other four sequences (LVF from the β -strand sequence L¹⁰¹-F¹⁰³ of PDB entry 1AXC⁴², ITY from the β -strand sequence I²⁰⁵-Y²⁰⁷ of PDB entry 2PEC⁴³, VVV from the β -strand sequence V⁶⁶-V⁶⁸ of PDB entry 1EST⁴¹, LVI from the β -strand sequence L⁵²-I⁵⁴ of PDB entry 2PEC)⁴³ are taken from the crystal structures of β -strand parts or β -hairpins of proteins (see Methods: Conformation of

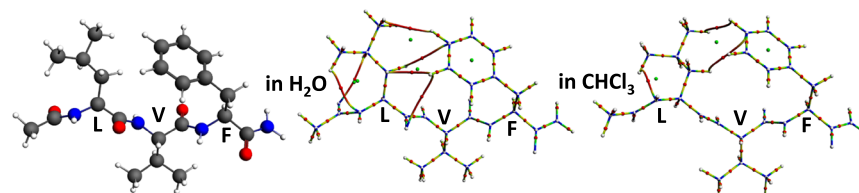
(a) GPA



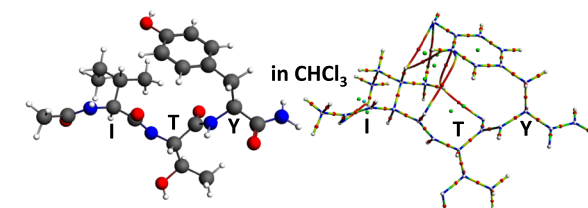
(b) GPS



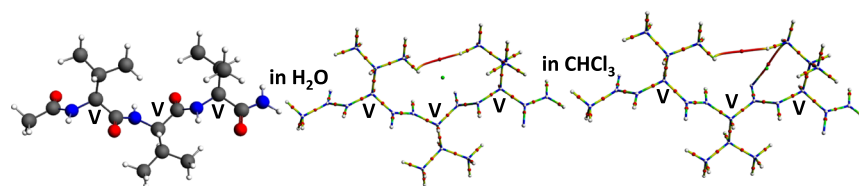
(c) LVF



(d) ITY



(e) VVV



(f) LVI

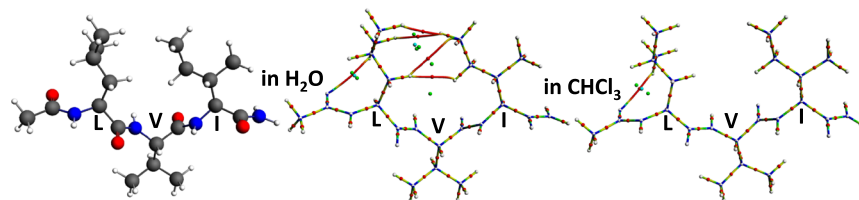


Figure 5. Bond paths in tri-peptides.

peptides). All structures were optimized using the M06-2X/6-31 + G(d) method with the SMD solvation model in water and in chloroform⁴⁰. DFT-optimized structures and original crystal structures showed no significant changes of the structures in terms of main chain and side chain angles (See Supplementary Information, except ITY in water: the structure was changed). Selected main chain torsion angles are shown in Supplementary Table 2.

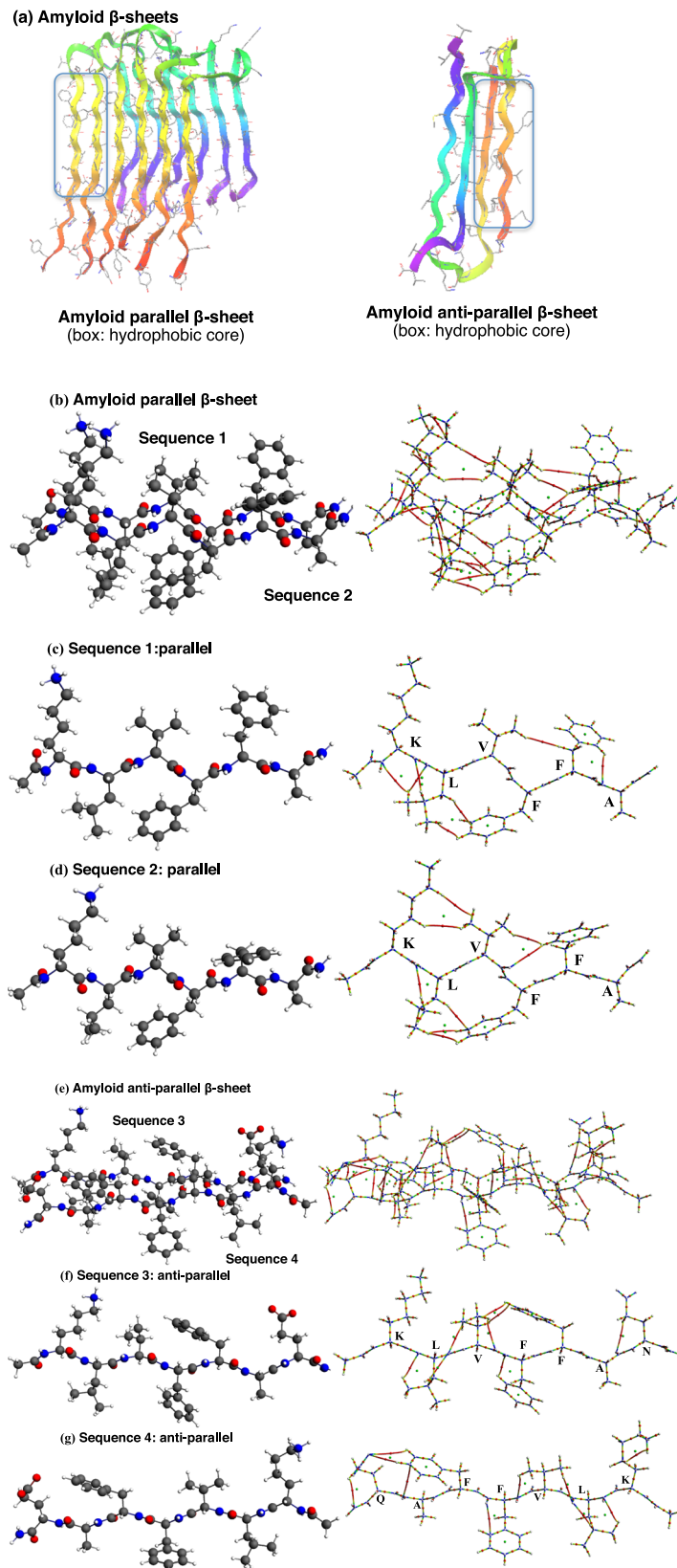


Figure 6. Molecular structures and molecular graphs. **(a)** Ribbon representations of amyloid parallel β -sheets and anti-parallel β -sheets. **(b)** Fragment of amyloid parallel β -sheets **(c,d)** a composition of parallel β -sheets, identical amino acid sequences of sequence 1 **(c)** and sequence 2 **(d)**: Ac-Lys¹⁶-Leu¹⁷-Val¹⁸-Phe¹⁹-Phe²⁰-Ala²¹-NH₂. **(e)** Fragment of amyloid anti-parallel β -sheet. **(f)** Sequence 3: Ac-Lys¹⁶-Leu¹⁷-Val¹⁸-Phe¹⁹-Phe²⁰-Ala²¹-Asn²²-NH₂; **(g)** Sequence 4: NH₂-Glu²²-Ala²¹-Phe²⁰-Phe¹⁹-Val¹⁸-Leu¹⁷-Lys¹⁶-Ac.

Bond type	n(BCP)	$\Sigma\rho(r_c)$	Average $\rho(r_c)$	Range of $\rho(r_c)$
Inter-strand interaction of Amyloid parallel β-sheet				
N-H...O=C	6	0.1314	0.0219	0.0170–0.0284
C-H...O=C	4	0.0454	0.0114	0.0041–0.0154
C=O...O=C	2	0.0227	0.0114	0.0027–0.0200
C-H...H-C	9	0.0444	0.0049	0.0028–0.0063
C-H... π	1	0.0112	0.0112	0.0112
C-H...H-N	1	0.0080	0.0080	0.0080
Intra-strand interaction of Sequence 1				
C-H...O=C	1	0.0119	0.0119	0.0119
C-H...N	3	0.0369	0.0123	0.0063–0.0161
C-H...H-C	2	0.0114	0.0057	0.0023–0.0091
C-H... π	1	0.0109	0.0109	0.0109
Intra-strand interaction of Sequence 2				
C-H...O=C	2	0.0159	0.0080	0.0033–0.0126
C-H...H-C	4	0.0198	0.0050	0.0033–0.0060
C-H... π	1	0.0118	0.0118	0.0118
C-H...C	1	0.0039	0.0039	0.0039

Table 2. Bond type, number of bond critical points (n(BCP)), and total electron density at the bond critical points ($\Sigma\rho(r_c)$) (a.u.) of amyloid parallel β -sheet and sequence 1 and sequence 2.

Tri-peptides **GPA** and **GPS** take typical PPII structures (Fig. 5a,b). In the case of **GPA** and **GPS**, no bond path was seen in water or chloroform. On the other hand, for the other β -strand-favoring sequences (**LVF**, **ITY**, **VVV**, **LVI**) (Fig. 5c–f), multiple bond paths between side chain and side chain, and between side chain and main chain are found (Fig. 5c–f). In the case of **LVF** (Fig. 5c), bond paths between side chain and main chain are detected in Leu, and the same kind of bond path is also seen in Leu dipeptide (Fig. 1). It is noteworthy that multiple bond paths including the H to H bond path, H to C bond path and C–H— π , between the benzene ring of side chain of Phe and the hydrogen or carbon atoms of the side chain of Leu were found, while the isolated Phe dipeptide did not have any bond bath (Fig. 3). Similarly, no through-space bond path is observed in the Tyr dipeptide (Supplementary Fig. 7), while in the case of **ITY** (Fig. 5d), there are multiple bond paths including the H to H bond path, H to C bond path, H to O bond path and C–H— π , between the phenol ring of side chain of Tyr (**Y**) and the hydrogen or carbon atoms of the side chain of Ile (**I**). In the case of **VVV** and **LVI** (Fig. 5e,f), we found bond paths between side chain and main chain, and also there are many H to H bond paths between the side chain and side chain. To conclude, in tripeptides, we can detect not only the side chain to main chain bond paths, which are similar to those in the dipeptide, but also multiple bond paths from side chain to side chain. This kind of side chain to side chain and side chain to main chain network of bond paths are characteristic to β -strand structures. We hypothesize that such network of bond paths may accompany with the β -strand structure in solution. Amino acids that do not show bond paths in dipeptides can form bond paths in tripeptides. Intriguingly, in tripeptides, the side chains of i and $i + 2$ amino acid residues tend to rotate in the same direction, i.e., N-facing (see Fig. 1). From the results of the present study of the dipeptides (e.g. Fig. 1), the N-facing of the side chain induced the formation of through-space bond path frequently in β -strand structures. Therefore, the N-facing of the side chains at the i and $i + 2$ amino acid residues in the β -strand structure of the tripeptide is consistent with the observation in the dipeptides.

Conformational preferences and bond paths of amyloid β -sheet. β -Sheet, consisting of β -strands, is one of the most common secondary structures in proteins (Fig. 6(a)). We clipped out a parallel β -sheet sequence containing the hydrophobic aromatic core of A β 40 (PDB 2LMP: K¹⁶-L¹⁷V¹⁸F¹⁹F²⁰-A²¹)⁴⁶ and applied QTAIM calculation directly without structure optimization. Figure 6(a) (left) shows the structure of the amyloid parallel β -sheet and Fig. 6(b) (right, molecular graph) delineates multiple inter-strand and intra-strand bond paths in the β -sheet. This complicated network contains side chain to side chain and side chain to main chain bond paths within one strand or within another strand, and also main chain to main chain bond paths between two strands. To analyze the amyloid parallel β -sheet structure further, **sequence 1** and **sequence 2** were calculated individually (Fig. 6c,d). The amino acid sequence of **sequence 1** and **sequence 2** are the same but the conformation is slightly different. The inter-strand and intra-strand interaction data are summarized in Table 2. **Sequence 1** is stabilized by one hydrogen-bonding bond path (C–H...O=C) and one hydrogen-benzene ring (C–H... π (C)) bond path²⁸, and there are several peripheral weak interactions such as three hydrogen(H)-nitrogen(N) bond paths and two hydrogen(H)-hydrogen(H) bond paths, among which the hydrogen-nitrogen bond path makes the greatest contribution based on number of bond paths and accumulation of average electron density at the bond critical point. **Sequence 2** is stabilized by two hydrogen-bonding bond paths (C=O...H-C) and one hydrogen-benzene ring (C–H... π (C)) bond path, also contains four hydrogen(H)-hydrogen(H) bond paths, and one hydrogen(H)-carbon(C) bond path. Among them, the hydrogen(H)-hydrogen(H) bond path is the most abundant, based on number of bond paths. When **sequence 1** and **sequence 2** were assembled, the intra-strand interactions were conserved. As regards these inter-strand interactions, the β -sheet is stabilized by six amide

Bond type	n(BCP)	$\Sigma\rho(r_c)$	Average $\rho(r_c)$	Range of $\rho(r_c)$
Inter-strand interaction of Amyloid anti-parallel β-sheet				
N-H...O=C	7	0.2097	0.0300	0.0066–0.0505
C-H...O=C	2	0.0210	0.0105	0.0072–0.0138
C=O...O=C	3	0.0550	0.0183	0.0129–0.0286
C-H...H-C	7	0.0467	0.0067	0.0046–0.0101
C=O...C	1	0.0062	0.0062	0.0062
C-H...H-N	1	0.0043	0.0043	0.0043
Intra-strand interaction of Sequence 3				
C-H...O=C	3	0.0352	0.0117	0.0088–0.0140
C-H...N	2	0.0442	0.0221	0.0170–0.0272
C-H...H-C	1	0.0089	0.0089	0.0089
C-H... π	1	0.0085	0.0085	0.0085
C=O...C	1	0.0120	0.0120	0.0120
Intra-strand interaction of Sequence 4				
C-H...O=C	4	0.0315	0.0079	0.0027–0.0149
C=O... π	2	0.0170	0.0085	0.0030–0.0140
C-H...N	3	0.0403	0.0134	0.0117–0.0161
C-H...H-N (sp ³)	1	0.0201	0.0201	0.0201

Table 3. Bond type, number of bond critical points (n(BCP)), and total electron density at the bond critical points ($\Sigma\rho(r_c)$) (a.u.) of *anti* parallel β -sheet and sequence 3 and sequence 4.

hydrogen-bonding bond paths, four hydrogen-bonding bond paths (C-H...O=C), and one hydrogen-benzene ring (C-H... π (C)) bond path, also coexisting through-space bond paths such as two oxygen(O)-oxygen(O) bond paths, and ten hydrogen-hydrogen bond paths (C-H...H-C and C-H...H-N). The six amide hydrogen-bonding bond paths may contribute most to stabilization of the β -sheet based on the number of bond paths and total electron density. However, other inter-strand bond-path networks, apart from hydrogen-bonding, are characteristic of β -strand/sheet structures, and are never examined in α -helical structure³⁶. The average and range of the electron density in each kind of atom pairs (Table 2) may indicate the magnitude of inter-strand interactions. In this context, the major contribution in forming the peptide assemblies comes from the conventional N-H—O=C hydrogen bonding, but the sum of various other weak interactions may be also contributed to or accompanied with overall β -strand/sheet structures.

Anti-parallel β -sheet was isolated from A β 40 (PDB 2LNQ_A:⁴⁷ K¹⁶-N²² (K¹⁶L¹⁷V¹⁸F¹⁹F²⁰A²¹N²²), 2LNQ_B:⁴⁷ K¹⁶-Q²² (K¹⁶L¹⁷V¹⁸F¹⁹F²⁰A²¹Q²²)) and QTAIM calculation was applied directly without structure optimization. Figure 6(e) shows the structure and bond paths (molecular graph). Various inter-strand and intra-strand bond paths were found. To analyze this complicated bond-path network, we separated the anti-parallel β -sheet, and calculated **Sequence 3** and **Sequence 4** individually (Fig. 6f,g). The inter-strand and intra-strand interaction data are collected in Supplementary Table 4. The results indicate that **Sequence 3** is stabilized by three hydrogen-bonding (C=O...H-C) bond paths and one C-H... π (C) bond path, accompanying topologically neighboring interactions such as two hydrogen(H)-nitrogen(N) bond paths, one hydrogen(H)-hydrogen(H) bond path, and one oxygen(O)-carbon(C) bond path, among which the two hydrogen-nitrogen bond paths are most characteristic in terms of the electron density. In **Sequence 4**, four hydrogen-bonding (C-H...O=C) bond paths and two C=O... π (C) bond paths stabilize the whole structure, accompanying topologically neighboring interactions such as three hydrogen(H)-nitrogen(N) bond paths, and one hydrogen(H)-hydrogen(H) bond path, among which the three hydrogen-nitrogen bond paths are most characteristic in terms of the electron density. As for inter-strand interactions, the anti-parallel β -sheet is stabilized by seven amide hydrogen-bonding bond paths, two hydrogen-bonding bond paths (C-H...O=C), together with topologically neighboring interactions of three oxygen-oxygen bond paths, eight hydrogen-hydrogen bond paths (C-H...H-C and C-H...H-N) and one oxygen(O)-carbon(C) bond path, while the intra-strand interactions in **Sequence 3** and **Sequence 4** are conserved. Based on bond-path strength (in terms of number and average electron density at the bond critical point), the amide hydrogen-bonding bond paths are the most important (Table 3). However, other topologically neighboring inter-strand bond-path networks, apart from amide hydrogen-bonding, are also characteristic of β -strand/sheet structures. Neither amyloid parallel β -sheet nor anti-parallel β -sheet showed π - π interaction although the hydrophobic aromatic core of A β 40 (17–20) (Leu¹⁷-Val¹⁸-Phe¹⁹-Phe²⁰) was proposed to be important in β -sheet formation^{10–15}. The QTAIM can detect through-space π - π interaction successfully in other systems⁵⁸. Therefore the present results are consistent with the previous reports^{16,17}, which excluded specific interactions involving π -electrons or aromatic character as forces that stabilize the whole fibril^{10–15}.

These results clearly suggest the importance of the weak bond-path network for the β -sheet structure. The van der Waals (vdW) interactions, one of the origins of hydrophobic interaction, have been regarded as dipole-induced non-directional intermolecular force, but the experiments in the crystals were inconsistent with this view. Instead, there are several reports to suggest that vdW interaction is directional, which can be described by bond path^{59,60}. This notion is consistent with our present results.

In conclusion, we carried out the QTAIM analysis to examine the postulate that hydrophobic interactions in peptide assembly can be represented in more visible manner in terms of through-space bond-paths, which are weak, but local and directional. We confirmed first that amino acids bearing a branched side chain, such as Val, Ile, Thr, and Cys, and aromatic amino acids, such as Tyr, Trp and Phe, have high propensities to form β -strand structure, whereas Ala, Gly and Pro have poor propensities^{19–21}, by means of the presence or absence of through-space bond path. Intriguingly, in tripeptides, the side chains of i and $i + 2$ amino acid residues tend to rotate in the same direction, i.e., N-facing (Fig. 1). From the results of the present study of the dipeptides, the N-facing of the side chain induced the formation of through-space bond path frequently in β -strand structures, but never induce in the PPII structures. Apart from amide hydrogen-bonding, other inter-strand bond-path networks arising from various types of weak bond paths (X–H—Y; X, Y = H, C, O, N, S), as well as *non-H–non-H* bond paths, are characteristic of β -strand/sheet structures. Similar bond paths from side chain to side chain and from side chain to main chain were found in a single β -strand and in di- and tripeptides. Therefore, weak interaction networks are based on a bottom-up approach from dipeptides, tripeptides to longer peptides. However, some of these through-space bond-path networks, particularly of aromatic amino acids such as Phe and Tyr were enhanced upon β -sheet formation: while there were no through-space bond paths in the Phe dipeptide and Tyr dipeptide. Multiple through-space bond paths were generated in a parallel β -sheet sequence containing the hydrophobic aromatic core of A β 40 (K¹⁶-L¹⁷V¹⁸F¹⁹F²⁰-A²¹) and anti-parallel β -sheet sequence of A β 40 (K¹⁶-N²² (K¹⁶L¹⁷V¹⁸F¹⁹F²⁰A²¹N²²) and K¹⁶-Q²² (K¹⁶L¹⁷V¹⁸F¹⁹F²⁰A²¹Q²²)). Intriguingly, neither amyloid parallel β -sheet nor anti-parallel β -sheet showed π - π interaction between aromatic amino acids in the sequences, which was previously claimed to be significant for amyloid β -sheet formation^{10–15}. Therefore the present bond-path analysis supported the previous experimental results^{16,17}, which excluded specific π - π interactions as forces that stabilize the whole A β fibril. Thus, the present bond-path analysis may be helpful in guiding *de novo* design of bioactive A β mimics and binding epitopes for protein aggregation and protein-protein interaction. We already demonstrated the validity of this kind of bond bath analysis in the experimental generation of β -strand stabilizer²³: the experimental observations were completely consistent with the through-space bond path analysis.

References

1. Marshall, K. E. *et al.* Hydrophobic, Aromatic, and Electrostatic Interactions Play a Central Role in Amyloid Fibril Formation and Stability. *Biochemistry*. **50**, 2061–2071 (2011).
2. Kreuzer, A. G. & Nowick, J. S. Elucidating the Structures of Amyloid Oligomers with Macrocyclic β -Hairpin Peptides: Insights into Alzheimer's Disease and Other Amyloid Diseases. *Acc. Chem. Res.* **51**, 706–718 (2018).
3. Ross, C. A. & Poirier, M. A. Protein aggregation and neurodegenerative disease. *Nat. Med.* **10**(suppl), S10–S17 (2004).
4. Selkoe, D. J. Alzheimer's Disease: Genes, Proteins, and Therapy. *Physiol. Rev.* **81**, 741–766 (2001).
5. Senguen, F. T. *et al.* Probing Aromatic, Hydrophobic, and Steric Effects on the Self-Assembly of an Amyloid- β Fragment Peptide. *Mol. Biosyst.* **7**, 486–496 (2011).
6. Bemporad, F. *et al.* Sequence and Structural Determinants of Amyloid Fibril Formation. *Acc. Chem. Res.* **39**, 620–627 (2006).
7. Hamley, I. W. Peptide Fibrillization. *Angew. Chem. Int. Ed.* **46**, 8128–8147 (2007).
8. Makin, O. S., Atkins, E., Sikorski, P., Johansson, J. & Serpell, L. C. Molecular basis for amyloid fibril formation and stability. *Proc. Natl. Acad. Sci.* **102**, 315–320 (2005).
9. Chiti, F., Stefani, M., Taddei, N., Ramponi, G. & Dobson, C. M. Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature*. **424**, 805–808 (2003).
10. Anand, U. & Mukherjee, M. Exploring the Self-Assembly of a Short Aromatic A β (16–24) Peptide. *Langmuir*. **29**, 2713–2721 (2013).
11. Gazit, E. A possible role for π -stacking in the self-assembly of amyloid fibrils. *FASEB J.* **16**, 77–83 (2002).
12. Jack, E., Newsome, M., Stockley, P. G., Radford, S. E. & Middleton, D. A. The Organization of Aromatic Side Groups in an Amyloid Fibril Probed by Solid-State 2H and 19F NMR Spectroscopy. *J. Am. Chem. Soc.* **128**, 8098–8099 (2006).
13. Wu, C., Lei, H. & Duan, Y. The Role of Phe in the Formation of Well-Ordered Oligomers of Amyloidogenic Hexapeptide (NFGAIL) Observed in Molecular Dynamics Simulations with Explicit Solvent. *Biophys. J.* **88**, 2897–2906 (2005).
14. Levy, M., Garmy, N., Gazit, E. & Fantini, J. The minimal amyloid-forming fragment of the islet amyloid polypeptide is a glycolipid-binding domain. *FEBS J.* **273**, 5724–5735 (2006).
15. Reches, M., Porat, Y. & Gazit, E. Amyloid Fibril Formation by Pentapeptide and Tetrapeptide Fragments of Human Calcitonin. *J. Biol. Chem.* **277**, 35475–35480 (2002).
16. Armstrong, A. H., Chen, J., McKoy, A. F. & Hecht, M. H. Mutations That Replace Aromatic Side Chains Promote Aggregation of the Alzheimers A β Peptide. *Biochemistry*. **50**, 4058–4067 (2011).
17. Bemporad, F., Taddei, N., Stefani, M. & Chiti, F. Assessing the role of aromatic residues in the amyloid aggregation of human muscle acylphosphatase. *Protein Sci.* **15**, 862–870 (2006).
18. Lopez del Amo, J. M. *et al.* Structural Properties of EGCG-Induced, Nontoxic Alzheimer's Disease A β Oligomers. *J. Mol. Biol.* **421**, 517–524 (2012).
19. Hill, T. A., Shepherd, N. E., Diness, F. & Fairlie, D. P. Constraining Cyclic Peptides to Mimic Protein Structure Motifs. *Angew. Chem. Int. Ed.* **53**, 13020–13041 (2014).
20. Merkel, J. S., Sturtevant, J. M. & Regan, L. Sidechain Interactions in Parallel β Sheets: The Energetics of Cross-Strand Pairings. *Structure*. **7**, 1333–1343 (1999).
21. Street, A. G. & Mayo, S. L. Intrinsic Beta-Sheet Propensities Result from van Der Waals Interactions between Side Chains and the Local Backbone. *Proc. Natl. Acad. Sci.* **96**, 9074–9076 (1999).
22. Grdadolnik, J., Mohacek-Grosev, V., Baldwin, R. L. & Avbelj, F. Populations of the Three Major Backbone Conformations in 19 Amino Acid Dipeptides. *Proc. Natl. Acad. Sci.* **108**, 1794–1798 (2011).
23. Zhai, L. *et al.* Application of C-Terminal 7-Azabicyclo[2.2.1]Heptane to Stabilize β -Strand-like Extended Conformation of a Neighboring α -Amino Acid. *J. Org. Chem.* **83**, 13063–13079 (2018).
24. Bader, R. F. W. *Atoms in Molecules. A Quantum Theory*; Oxford University Press: New York (1990).
25. Runtz, G. R., Bader, R. F. W. & Messer, R. R. Definition of bond paths and bond directions in terms of the molecular charge distribution. *Can. J. Chem.* **55**, 3040–3045 (1977).
26. Matta, C. F. & Boyd, R. J. *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*; Wiley-VCH: New York (2007).
27. Bader, R. F. W. Bond Paths Are Not Chemical Bonds. *J. Phys. Chem. A*. **113**, 10391–10396 (2009).
28. Bader, R. F. W. A Bond Path: A Universal Indicator of Bonded Interactions. *J. Phys. Chem. A*. **102**, 7314–7323 (1998).
29. Foroutan-Nejad, C., Shahbazian, S. & Marek, R. Toward a Consistent Interpretation of the QTAIM: Tortuous Link between Chemical Bonds, Interactions, and Bond/Line Paths. *Chem. Eur. J.* **20**, 10140–10152 (2014).

30. Keyvani, Z. A., Shahbazian, S. & Zahedi, M. To What Extent Are “Atoms in Molecules” Structures of Hydrocarbons Reproducible from the Promolecule Electron Densities? *Chem. Eur. J.* **22**, 5003–5009 (2016).
31. Shahbazian, S. Why Bond Critical Points Are Not “Bond” Critical Points. *Chem. Eur. J.* **24**, 5401–5405 (2018).
32. Lammertsma, K. & Ohwada, T. Three-Center, Two-Electron Systems. Origin of the Tilting of Their Substituents. *J. Am. Chem. Soc.* **118**, 7247–7254 (1996).
33. Popelier, P. L. A. & Bader, R. F. W. Effect of Twisting a Polypeptide on Its Geometry and Electron Distribution. *J. Phys. Chem.* **98**, 4473–4481 (1994).
34. Chang, C. & Bader, R. F. W. Theoretical Construction of a Polypeptide. *J. Phys. Chem.* **96**, 1654–1662 (1992).
35. Matta, C. F. & Bader, R. F. W. An Atoms-In-Molecules Study of the Genetically-Encoded Amino Acids: I. Effects of Conformation and of Tautomerization on Geometric, Atomic, and Bond Properties. *Proteins Struct. Funct. Genet.* **40**, 310–329 (2000).
36. Lapointe, S. M., Farrag, S., Bohórquez, H. J. & Boyd, R. J. QTAIM Study of an Alpha-Helix Hydrogen Bond Network. *J. Phys. Chem. B.* **113**, 10957–10964 (2009).
37. Schrödinger Release 2017-4: Maestro, Schrödinger, LLC, New York, NY (2017).
38. Frisch, M. J. *et al.* Gaussian 09, revision D.01, Gaussian, Inc., Wallingford, CT (2013).
39. Kang, Y. K., Byun, B. J. & Park, H. S. Conformational Preference and Cis-Trans Isomerization of 4-Methylproline Residues. *Biopolymers.* **95**, 51–61 (2011).
40. Marenich, A. V., Cramer, C. J. & Truhlar, D. G. Supporting Information (PART I) Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions Contents. *J. Phys. Chem. B* **113**, 6378 (2009).
41. Sawyer, L. *et al.* The Atomic Structure of Crystalline Porcine Pancreatic Elastase at 2.5 Å Resolution: Comparisons with the Structure of α -Chymotrypsin. *J. Mol. Biol.* **118**, 137–208 (1978).
42. Gulbis, J. M., Kelman, Z., Hurwitz, J., O'Donnell, M. & Kuriyan, J. Structure of the C-Terminal Region of P21(WAF1/CIP1) Complexed with Human PCNA. *Cell.* **87**, 297–306 (1996).
43. Yoder, M. D. & Jvrnakti, F. The Parallel β Helix and Other Coiled Folds. *FASEB J.* **9**, 335–342 (1995).
44. Matsuura, Y., Takano, T. & Dickerson, R. E. Structure of Cytochrome C551 from *Pseudomonas Aeruginosa* Refined at 1.6 Å Resolution and Comparison of the Two Redox Forms. *J. Mol. Biol.* **156**, 389–409 (1982).
45. Houde, D., Arndt, J., Domeier, W., Berkowitz, S. & Engen, J. R. Characterization of IgG1 Conformation and Conformational Dynamics by Hydrogen/Deuterium Exchange Mass Spectrometry. *Anal. Chem.* **81**, 2644–2651 (2009).
46. Paravastu, A. K., Leapman, R. D., Yau, W. & Tycko, R. Molecular structural basis for polymorphism in Alzheimer's β -amyloid fibrils. *Proc. Natl. Acad. Sci.* **105**, 18349–18354 (2008).
47. Qiang, W., Yau, W., Luo, Y., Mattson, M. P. & Tycko, R. Antiparallel β -sheet architecture in Iowa-mutant β -amyloid fibrils. *Proc. Natl. Acad. Sci.* **109**, 4443–4448 (2012).
48. Baerends, E. J. *et al.* ADF2018 and ADF2019 SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands (2019), <http://www.scm.com>.
49. Van Lenthe, E. & Baerends, E. J. Optimized Slater-Type Basis Sets for the Elements 1–118. *J. Comput. Chem.* **24**, 1142–1156 (2003).
50. Grimme, S. *et al.* When Do Interacting Atoms Form a Chemical Bond? Spectroscopic Measurements and Theoretical Analyses of Dideuteriophenanthrene. *Angew. Chem. Int. Ed.* **48**, 2592–2595 (2009).
51. Cerpa, E., Krapp, A., Vela, A. & Merino, G. The Implications of Symmetry of the External Potential on Bond Paths. *Chem. Eur. J.* **14**, 10232–10234 (2008).
52. Cerpa, E., Krapp, A., Flores-Moreno, R., Donald, K. & Merino, G. Influence of Endohedral Confinement on the Electronic Interaction between He Atoms: A He₂@C₂₀H₂₀ Case Study. *Chem. Eur. J.* **15**, 1985–1990 (2008).
53. Wick, C. R. & Clark, T. On Bond-Critical Points in QTAIM and Weak Interactions. *J. Mol. Model.* **24**, 1–9 (2018).
54. Clark, T. Polarization, Donor–Acceptor Interactions, and Covalent Contributions in Weak Interactions: A Clarification. *J. Mol. Model.* **23**, 297 (2017).
55. Monteiro, N. K. V. & Firme, C. L. Hydrogen-Hydrogen Bonds in Highly Branched Alkanes and in Alkane Complexes: A DFT, Ab Initio, QTAIM, and ELF Study. *J. Phys. Chem. A.* **118**, 1730–1740 (2014).
56. Matta, C. F., Hernández-Trujillo, J., Tang, T. H. & Bader, R. F. W. Hydrogen - Hydrogen Bonding: A Stabilizing Interaction in Molecules and Crystals. *Chem. Eur. J.* **9**, 1940–1951 (2003).
57. Vener, M. V., Manaev, A. V., Egorova, A. N. & Tsirelson, V. G. QTAIM Study of Strong H-Bonds with the O-H—A Fragment (A = O, N) in Three-Dimensional Periodical Crystals. *J. Phys. Chem. A.* **111**, 1155–1162 (2007).
58. Churchill, C. D. M., Rutledge, L. R. & Wetmore, S. D. Effects of the biological backbone on stacking interactions at DNA–protein interfaces: the interplay between the backbone $\dots\pi$ and $\pi\dots\pi$ components. *Phys. Chem. Chem. Phys.* **12**, 14515–14526 (2010).
59. Scherer, W. *et al.* Molecular Recognition in the Solid State: Topology of Experimental and Theoretical Charge Densities for Tetrasulfur Tetranitride. *Chem. Commun.* **7**, 635–636 (2000).
60. Gibbs, G. V. *et al.* Bond Paths and van Der Waals Interactions in Orpiment, As₂S₃. *J. Phys. Chem. A.* **114**, 6550–6557 (2010).

Acknowledgements

The computations were performed at the Research Center for Computational Science, Okazaki, Japan. This work was supported by JSPS KAKENHI Grant Numbers JP 26104508 and 16K08157 (Y.O.) and 26293002 and 18H02552 (T.O.).

Author Contributions

The whole project was conceived by Z.L., Y.O. and T.O. All calculations were carried by Z.L. and Y.O. gave technical support and conceptual advice. The data analysis was carried out by Z.L., Y.O. and T.O. The manuscript was written by Z.L. and T.O.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-47151-2>.

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) 2019