W O R L D W I D E

wwPDB

PROTEIN DATA BANK

wwPDB NMR Structure Validation Summary Report (i)

Jun 7, 2020 – 04:33 am BST

PDB ID	:	6OC9
Title	:	S8 phosphorylated beta amyloid 40 fibrils
Authors	:	Qiang, W.; Hu, Z.W.
Deposited on	:	2019-03-22

The following versions of software and data (see references (i)) were used in the production of this report:

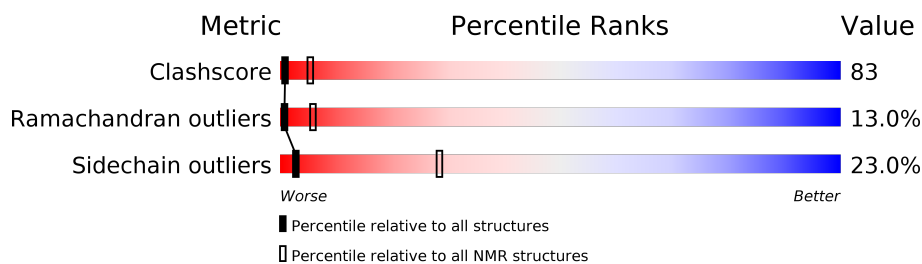|		
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
SOLID-STATE NMR

The overall completeness of chemical shifts assignment is 3%.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Percentile Ranks	Value
Clashscore		83
Ramachandran outliers		13.0%
Sidechain outliers		23.0%

Worse   *Better*

■ Percentile relative to all structures
▯ Percentile relative to all NMR structures

Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and a grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5%

Mol	Chain	Length	Quality of chain
1	A	40	20% / 50% / 13% / 18%
1	B	40	20% / 63% / 18%
1	C	40	18% / 40% / 10% / 33%
1	D	40	13% / 38% / 10% / 40%
1	E	40	20% / 28% / 10% / 43%
1	F	40	33% / 38% / 10% / 20%
1	G	40	25% / 58% / 18%
1	H	40	25% / 58% / 18%
1	I	40	18% / 70% / 13%

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	J	40	23% 63% 15%

2 Ensemble composition and analysis (i)

This entry contains 10 models. Model 10 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy.*

The following residues are included in the computation of the global validation metrics.

Well-defined (core) protein residues			
Well-defined core	Residue range (total)	Backbone RMSD (Å)	Medoid model
1	A:8-A:40, B:1-B:40, C:1-C:3, C:17-C:40, D:17-D:40, E:18-E:40, F:9-F:40, G:1-G:40, H:1-H:40, I:1-I:40, J:1-J:40 (339)	1.75	10

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 1 clusters and 2 single-model clusters were found.

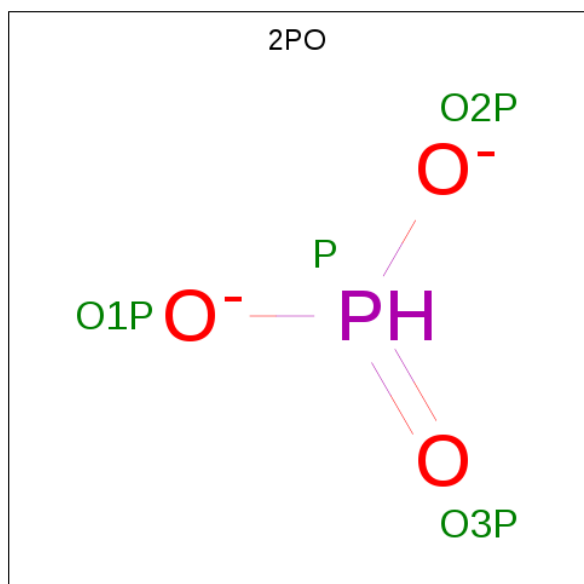| Cluster number	Models
1	1, 2, 4, 5, 6, 8, 9, 10
Single-model clusters	3; 7

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 6010 atoms, of which 2910 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Amyloid-beta precursor protein.

Mol	Chain	Residues	Atoms							Trace
1	A	40	Total	C	H	N	O	S		0
			597	194	291	53	58	1		
1	B	40	Total	C	H	N	O	S		0
			597	194	291	53	58	1		
1	C	40	Total	C	H	N	O	S		0
			597	194	291	53	58	1		
1	D	40	Total	C	H	N	O	S		0
			597	194	291	53	58	1		
1	E	40	Total	C	H	N	O	S		0
			597	194	291	53	58	1		
1	F	40	Total	C	H	N	O	S		0
			597	194	291	53	58	1		
1	G	40	Total	C	H	N	O	S		0
			597	194	291	53	58	1		
1	H	40	Total	C	H	N	O	S		0
			597	194	291	53	58	1		
1	I	40	Total	C	H	N	O	S		0
			597	194	291	53	58	1		
1	J	40	Total	C	H	N	O	S		0
			597	194	291	53	58	1		

- Molecule 2 is PHOSPHONATE (three-letter code: 2PO) (formula: HO_3P).

Mol	Chain	Residues	Atoms		
2	A	1	Total	O	P
			4	3	1
2	B	1	Total	O	P
			4	3	1
2	C	1	Total	O	P
			4	3	1
2	D	1	Total	O	P
			4	3	1
2	E	1	Total	O	P
			4	3	1
2	F	1	Total	O	P
			4	3	1
2	G	1	Total	O	P
			4	3	1
2	H	1	Total	O	P
			4	3	1
2	I	1	Total	O	P
			4	3	1
2	J	1	Total	O	P
			4	3	1

4 Residue-property plots ⓘ

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. Stretches of 2 or more consecutive residues without any outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which are present in the experimental sample, but not modelled in the final structure are shown in grey.
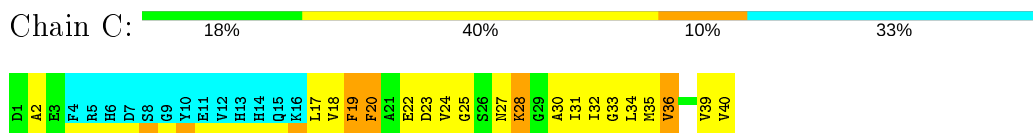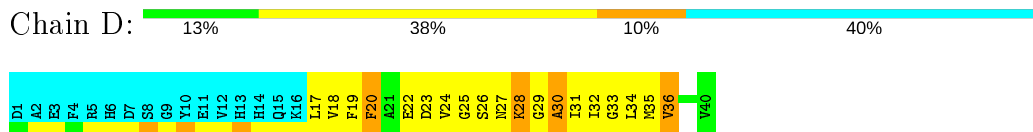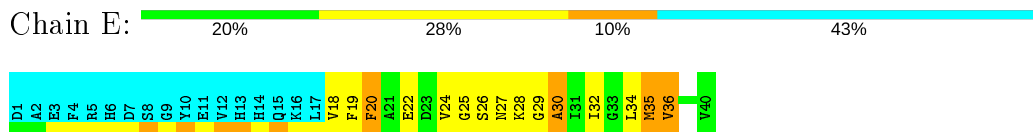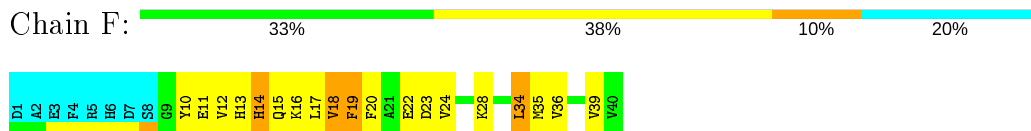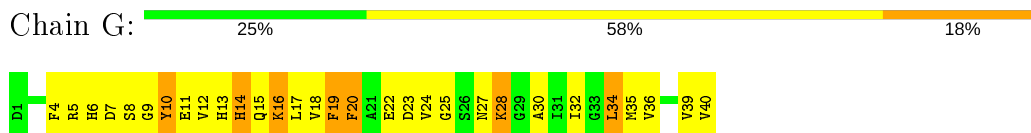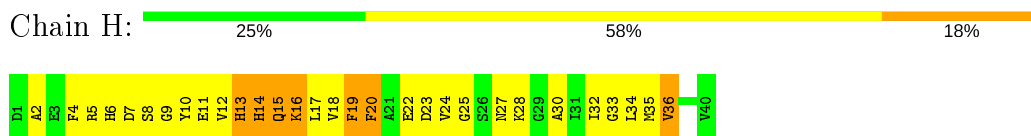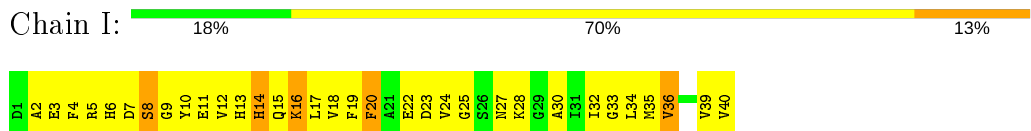
- Molecule 1: Amyloid-beta precursor protein
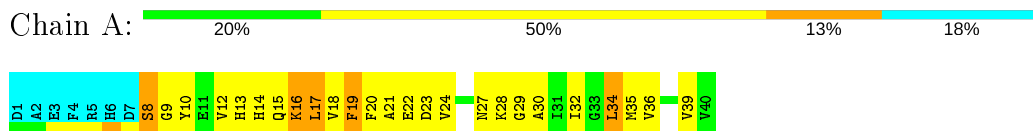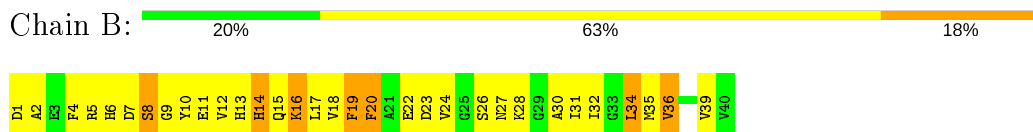
- Molecule 1: Amyloid-beta precursor protein

- Molecule 1: Amyloid-beta precursor protein
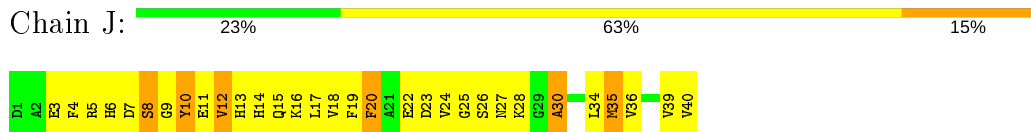

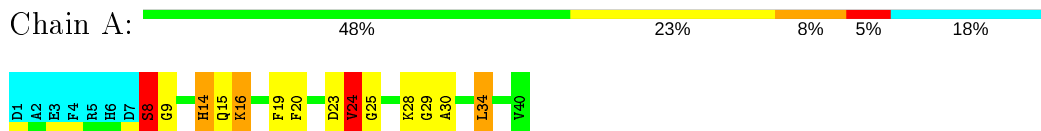- Molecule 1: Amyloid-beta precursor protein

- Molecule 1: Amyloid-beta precursor protein

- Molecule 1: Amyloid-beta precursor protein

Chain F: 33%   38%   10%   20%

D1 A2 E3 F4 R5 H6 D7 S8 G9 Y10 E11 V12 H13 H14 Q15 K16 L17 V18 F19 F20 A21 E22 D23 V24 K28 L34 M35 V36 V39 V40

• Molecule 1: Amyloid-beta precursor protein

Chain G: 25%   58%   18%

D1 F4 R5 H6 D7 S8 G9 Y10 E11 V12 H13 H14 Q15 K16 L17 V18 F19 F20 A21 E22 D23 V24 G25 S26 N27 K28 G29 A30 I31 I32 G33 L34 M35 V36 V39 V40

• Molecule 1: Amyloid-beta precursor protein

Chain H: 25%   58%   18%

D1 A2 E3 F4 R5 H6 D7 S8 G9 Y10 E11 V12 H13 H14 Q15 K16 L17 V18 F19 F20 A21 E22 D23 V24 G25 S26 N27 K28 G29 A30 I31 I32 G33 L34 M35 V36 V40

• Molecule 1: Amyloid-beta precursor protein

Chain I: 18%   70%   13%

D1 A2 E3 F4 R5 H6 D7 S8 G9 Y10 E11 V12 H13 H14 Q15 K16 L17 V18 F19 F20 A21 E22 D23 V24 G25 S26 N27 K28 G29 A30 I31 I32 G33 L34 M35 V36 V39 V40

• Molecule 1: Amyloid-beta precursor protein

Chain J: 23%   63%   15%

D1 A2 E3 F4 R5 H6 D7 S8 G9 Y10 E11 V12 H13 H14 Q15 K16 L17 V18 F19 F20 A21 E22 D23 V24 G25 S26 N27 K28 G29 A30 L34 M35 V36 V39 V40

4.2   Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 10. Colouring as in section 4.1 above.

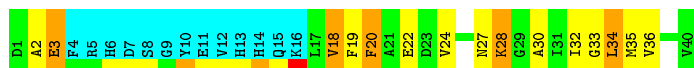• Molecule 1: Amyloid-beta precursor protein

Chain A: 48%   23%   8%   5%   18%

D1 A2 E3 F4 R5 H6 D7 S8 G9 H14 Q15 K16 F19 F20 D23 V24 G25 K28 G29 A30 L34 V40

• Molecule 1: Amyloid-beta precursor protein

Chain B: 50%   33%   18%

- Molecule 1: Amyloid-beta precursor protein

Chain C:

- Molecule 1: Amyloid-beta precursor protein

Chain D:

- Molecule 1: Amyloid-beta precursor protein

Chain E:

- Molecule 1: Amyloid-beta precursor protein
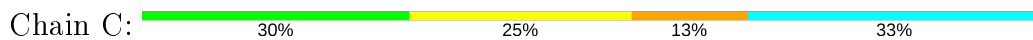
Chain F:

- Molecule 1: Amyloid-beta precursor protein
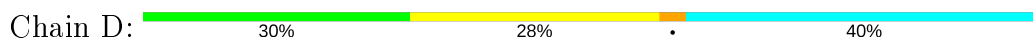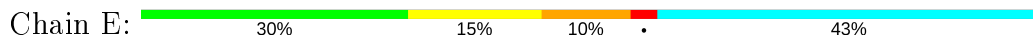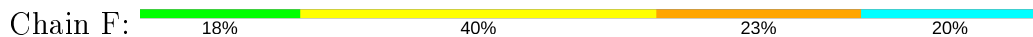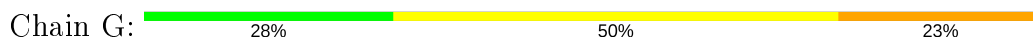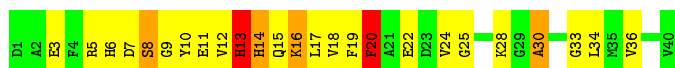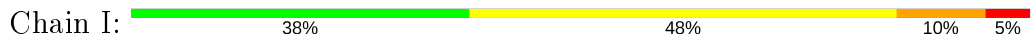
Chain G:

- Molecule 1: Amyloid-beta precursor protein

Chain H:

- Molecule 1: Amyloid-beta precursor protein

Chain I:

● Molecule 1: Amyloid-beta precursor protein

Chain J:

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *simulated annealing*.

Of the 40 calculated structures, 10 were deposited, based on the following criterion: *structures with the lowest energy*.

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
X-PLOR NIH	refinement	
X-PLOR NIH	structure calculation	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	123
Number of shifts mapped to atoms	123
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	3%

Note: This is a solid-state NMR structure, where hydrogen atoms are typically not assigned a chemical shift value, which may lead to lower completeness of assignment measure.

No validations of the models with respect to experimental NMR restraints is performed at this time.

6 Model quality (i)

6.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: 2PO

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

6.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in each chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	A	244	241	241	66±70
1	B	306	291	291	84±87
1	C	190	192	192	67±79
1	D	168	175	175	64±75
1	E	160	164	164	40±45
1	F	238	237	237	41±13
1	G	306	291	291	68±28
1	H	306	291	291	65±23
1	I	306	291	291	70±21
1	J	306	291	291	53±16
2	A	4	0	0	4±2
2	B	4	0	0	4±3
2	C	4	0	0	0±1
2	D	4	0	0	0±1
2	E	4	0	0	0±0
2	F	4	0	0	1±2
2	G	4	0	0	3±1
2	H	4	0	0	3±1
2	I	4	0	0	3±1
2	J	4	0	0	4±1
All	All	25700	24640	24633	4190

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 83.

5 of 3200 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:B:19:PHE:CD1	1:C:19:PHE:HB2	1.63	1.15	7	1
1:D:20:PHE:H	1:E:20:PHE:N	1.49	0.94	7	1
1:D:20:PHE:N	1:E:20:PHE:H	1.48	1.01	7	1
1:B:19:PHE:CG	1:C:19:PHE:HB2	1.43	1.48	7	1
1:A:17:LEU:O	1:B:17:LEU:CB	1.43	1.64	3	1

6.3 Torsion angles ⓘ

6.3.1 Protein backbone ⓘ

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the backbone conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	32/40 (80%)	22±3 (68±11%)	6±1 (20±4%)	4±3 (12±9%)	1	6
1	B	38/40 (95%)	24±4 (64±10%)	8±2 (22±6%)	5±4 (14±10%)	1	5
1	C	25/40 (62%)	19±3 (75±14%)	4±2 (16±7%)	2±2 (10±8%)	1	10
1	D	23/40 (58%)	17±4 (73±17%)	3±2 (14±10%)	3±3 (13±11%)	1	6
1	E	22/40 (55%)	17±2 (77±11%)	3±2 (13±8%)	2±2 (10±7%)	1	9
1	F	31/40 (78%)	23±1 (74±5%)	6±2 (18±6%)	3±1 (8±3%)	2	14
1	G	38/40 (95%)	26±3 (67±7%)	8±2 (20±4%)	5±2 (13±4%)	1	6
1	H	38/40 (95%)	25±2 (65±6%)	8±3 (22±7%)	5±1 (13±4%)	1	5
1	I	38/40 (95%)	24±2 (64±4%)	9±2 (23±5%)	5±2 (13±4%)	1	6
1	J	38/40 (95%)	23±4 (61±10%)	7±2 (19±6%)	8±3 (21±9%)	0	2
All	All	3230/4000 (81%)	2190 (68%)	619 (19%)	421 (13%)	1	6

5 of 168 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	H	16	LYS	9
1	J	30	ALA	9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	J	12	VAL	7
1	G	20	PHE	7
1	E	30	ALA	6

6.3.2   Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the sidechain conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	25/31 (81%)	20±2 (80±8%)	5±2 (20±8%)	3	33
1	B	31/31 (100%)	24±3 (78±9%)	7±3 (22±9%)	3	29
1	C	19/31 (61%)	13±1 (69±8%)	6±1 (31±8%)	1	15
1	D	17/31 (55%)	12±3 (73±19%)	5±3 (27±19%)	2	21
1	E	16/31 (52%)	12±3 (75±17%)	4±3 (25±17%)	2	25
1	F	24/31 (77%)	19±2 (80±7%)	5±2 (20±7%)	4	34
1	G	31/31 (100%)	24±2 (76±8%)	7±2 (24±8%)	2	27
1	H	31/31 (100%)	23±2 (74±6%)	8±2 (26±6%)	2	23
1	I	31/31 (100%)	25±1 (81±2%)	6±1 (19±2%)	4	36
1	J	31/31 (100%)	25±2 (79±6%)	7±2 (21±6%)	3	32
All	All	2560/3100 (83%)	1970 (77%)	590 (23%)	3	28

5 of 192 unique residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	G	28	LYS	10
1	J	35	MET	9
1	I	28	LYS	9
1	H	28	LYS	9
1	H	19	PHE	9

6.3.3   RNA (i)

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates (i)

There are no carbohydrates in this entry.

6.6 Ligand geometry (i)

10 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
2	2PO	F	101	-	0,3,3	0.00±0.00	-
2	2PO	H	101	-	0,3,3	0.00±0.00	-
2	2PO	C	101	-	0,3,3	0.00±0.00	-
2	2PO	E	101	-	0,3,3	0.00±0.00	-
2	2PO	B	101	-	0,3,3	0.00±0.00	-
2	2PO	J	101	-	0,3,3	0.00±0.00	-
2	2PO	G	101	-	0,3,3	0.00±0.00	-
2	2PO	A	101	-	0,3,3	0.00±0.00	-
2	2PO	D	101	-	0,3,3	0.00±0.00	-
2	2PO	I	101	-	0,3,3	0.00±0.00	-

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of angles that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
2	2PO	F	101	-	0,3,3	0.00±0.00	-
2	2PO	H	101	-	0,3,3	0.00±0.00	-

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
2	2PO	C	101	-	0,3,3	0.00±0.00	-
2	2PO	E	101	-	0,3,3	0.00±0.00	-
2	2PO	B	101	-	0,3,3	0.00±0.00	-
2	2PO	J	101	-	0,3,3	0.00±0.00	-
2	2PO	G	101	-	0,3,3	0.00±0.00	-
2	2PO	A	101	-	0,3,3	0.00±0.00	-
2	2PO	D	101	-	0,3,3	0.00±0.00	-
2	2PO	I	101	-	0,3,3	0.00±0.00	-

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

6.7 Other polymers (i)

There are no such molecules in this entry.

6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

7   Chemical shift validation (i)

The completeness of assignment taking into account all chemical shift lists is 3% for the well-defined parts and 3% for the entire structure.

7.1   Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *NMR-STAR_ 3.1_ for_ pABeta_ fibrils_ new1.txt*

7.1.1   Bookkeeping (i)

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	123
Number of shifts mapped to atoms	123
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

7.1.2   Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	32	1.12 ± 0.31	Should be applied
$^{13}C_\beta$	26	0.28 ± 0.44	None needed (< 0.5 ppm)
$^{13}C'$	32	1.72 ± 0.30	Should be applied
^{15}N	0	—	None (insufficient data)

7.1.3   Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 3%, i.e. 108 atoms were assigned a chemical shift out of a possible 3831. 2 out of 76 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	56/1695 (3%)	0/678 (0%)	56/678 (8%)	0/339 (0%)
Sidechain	52/1722 (3%)	0/983 (0%)	52/690 (8%)	0/49 (0%)

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	0/414 (0%)	0/229 (0%)	0/166 (0%)	0/19 (0%)
Overall	108/3831 (3%)	0/1890 (0%)	108/1534 (7%)	0/407 (0%)

Note: This is a solid-state NMR structure, where hydrogen atoms are typically not assigned a chemical shift value, which may lead to lower completeness of assignment measure.

7.1.4 Statistically unusual chemical shifts ⓘ

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots ⓘ

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.

Random coil index (RCI) for chain A: