



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 15, 2020 – 05:04 pm BST

PDB ID : 4HYG  
Title : Structure of a presenilin family intramembrane aspartate protease in C222 space group  
Authors : Li, X.; Dang, S.; Yan, C.; Wang, J.; Shi, Y.  
Deposited on : 2012-11-13  
Resolution : 3.32 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11



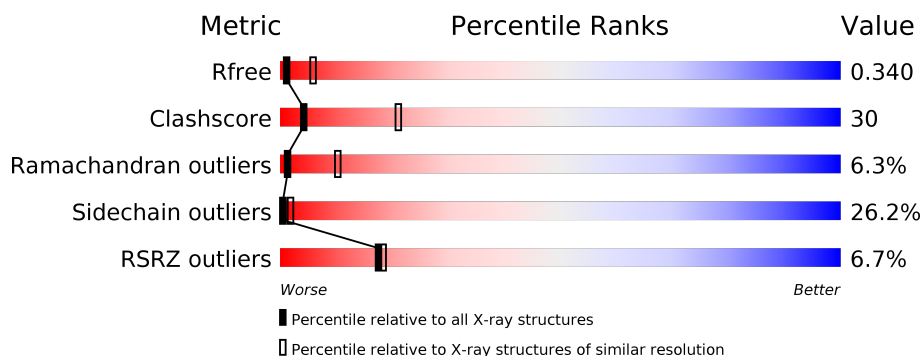
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.32 Å.

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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1089 (3.36-3.28)
Clashscore	141614	1137 (3.36-3.28)
Ramachandran outliers	138981	1115 (3.36-3.28)
Sidechain outliers	138945	1114 (3.36-3.28)
RSRZ outliers	127900	1059 (3.36-3.28)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria respectively. 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  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	301	<div> <div>3%</div> <div> <div>30%</div> <div>40%</div> <div>11%</div> <div>•</div> <div>18%</div> </div> </div>
1	B	301	<div> <div>5%</div> <div> <div>31%</div> <div>37%</div> <div>12%</div> <div>•</div> <div>18%</div> </div> </div>
1	C	301	<div> <div>6%</div> <div> <div>34%</div> <div>34%</div> <div>9%</div> <div>•</div> <div>21%</div> </div> </div>
1	D	301	<div> <div>7%</div> <div> <div>33%</div> <div>34%</div> <div>11%</div> <div>•</div> <div>21%</div> </div> </div>



## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 7115 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Putative uncharacterized protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	246	Total	C	N	O	S	0	0	0
			1802	1219	276	296	11			
1	B	247	Total	C	N	O	S	0	0	0
			1810	1226	277	296	11			
1	C	238	Total	C	N	O	S	0	0	0
			1763	1197	267	288	11			
1	D	237	Total	C	N	O	S	0	0	0
			1740	1180	264	285	11			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	40	ASN	ASP	ENGINEERED MUTATION	UNP A3CWV0
A	42	SER	GLU	ENGINEERED MUTATION	UNP A3CWV0
A	147	GLU	ALA	ENGINEERED MUTATION	UNP A3CWV0
A	148	PRO	VAL	ENGINEERED MUTATION	UNP A3CWV0
A	229	VAL	ALA	ENGINEERED MUTATION	UNP A3CWV0
B	40	ASN	ASP	ENGINEERED MUTATION	UNP A3CWV0
B	42	SER	GLU	ENGINEERED MUTATION	UNP A3CWV0
B	147	GLU	ALA	ENGINEERED MUTATION	UNP A3CWV0
B	148	PRO	VAL	ENGINEERED MUTATION	UNP A3CWV0
B	229	VAL	ALA	ENGINEERED MUTATION	UNP A3CWV0
C	40	ASN	ASP	ENGINEERED MUTATION	UNP A3CWV0
C	42	SER	GLU	ENGINEERED MUTATION	UNP A3CWV0
C	147	GLU	ALA	ENGINEERED MUTATION	UNP A3CWV0
C	148	PRO	VAL	ENGINEERED MUTATION	UNP A3CWV0
C	229	VAL	ALA	ENGINEERED MUTATION	UNP A3CWV0
D	40	ASN	ASP	ENGINEERED MUTATION	UNP A3CWV0
D	42	SER	GLU	ENGINEERED MUTATION	UNP A3CWV0
D	147	GLU	ALA	ENGINEERED MUTATION	UNP A3CWV0
D	148	PRO	VAL	ENGINEERED MUTATION	UNP A3CWV0
D	229	VAL	ALA	ENGINEERED MUTATION	UNP A3CWV0











## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	168.30Å 201.89Å 117.53Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.11 – 3.32 49.11 – 3.32	Depositor EDS
% Data completeness (in resolution range)	82.0 (49.11-3.32) 82.4 (49.11-3.32)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.10 (at 3.33Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, $R_{free}$	0.298 , 0.344 0.301 , 0.340	Depositor DCC
$R_{free}$ test set	1260 reflections (5.11%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	111.9	Xtriage
Anisotropy	1.048	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 145.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	7115	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	165.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.73% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.



## 5 Model quality

### 5.1 Standard geometry

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z  >5	RMSZ	# Z  >5
1	A	0.56	0/1839	0.84	1/2511 (0.0%)
1	B	0.52	0/1847	0.87	3/2523 (0.1%)
1	C	0.51	0/1799	0.82	0/2457
1	D	0.53	0/1777	0.86	3/2427 (0.1%)
All	All	0.53	0/7262	0.85	7/9918 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	5
1	B	0	3
1	C	0	2
1	D	0	1
All	All	0	11

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	120	LEU	CA-CB-CG	8.27	134.33	115.30
1	B	35	LEU	CA-CB-CG	8.20	134.15	115.30
1	D	146	LEU	CB-CG-CD1	5.87	120.98	111.00
1	B	59	LEU	CA-CB-CG	5.57	128.10	115.30
1	D	249	LEU	CA-CB-CG	5.40	127.72	115.30

There are no chirality outliers.

5 of 11 planarity outliers are listed below:



Mol	Chain	Res	Type	Group
1	A	124	TRP	Peptide
1	A	145	SER	Peptide
1	A	146	LEU	Peptide
1	A	246	ALA	Peptide
1	A	272	GLN	Peptide

## 5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the 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 within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1802	0	1949	123	0
1	B	1810	0	1965	109	0
1	C	1763	0	1903	95	0
1	D	1740	0	1881	119	0
All	All	7115	0	7698	437	0

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 30.

The worst 5 of 437 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:246:ALA:HA	1:B:249:LEU:HD12	1.43	1.00
1:D:139:ALA:HA	1:D:226:ILE:HG12	1.53	0.89
1:A:36:VAL:HG12	1:A:37:ALA:H	1.40	0.87
1:D:265:PHE:HB3	1:D:271:PRO:HG2	1.56	0.87
1:D:228:VAL:HG22	1:D:251:ALA:HB1	1.58	0.86

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 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 X-ray entries followed by that with respect to entries of similar resolution.



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	238/301 (79%)	196 (82%)	30 (13%)	12 (5%)	2	14
1	B	239/301 (79%)	198 (83%)	24 (10%)	17 (7%)	1	8
1	C	228/301 (76%)	188 (82%)	25 (11%)	15 (7%)	1	9
1	D	229/301 (76%)	191 (83%)	23 (10%)	15 (7%)	1	9
All	All	934/1204 (78%)	773 (83%)	102 (11%)	59 (6%)	1	10

5 of 59 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	36	VAL
1	A	99	PRO
1	A	122	PRO
1	A	246	ALA
1	B	42	SER

### 5.3.2 Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

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	188/236 (80%)	140 (74%)	48 (26%)	0	2
1	B	189/236 (80%)	137 (72%)	52 (28%)	0	1
1	C	185/236 (78%)	132 (71%)	53 (29%)	0	1
1	D	183/236 (78%)	141 (77%)	42 (23%)	1	3
All	All	745/944 (79%)	550 (74%)	195 (26%)	0	2

5 of 195 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	248	THR
1	C	65	LEU
1	D	158	LEU

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
1	B	257	VAL
1	B	288	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	267	ASN
1	B	279	ASN
1	D	20	GLN
1	D	232	HIS
1	D	270	ASN

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

### 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

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

### 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

### 5.6 Ligand geometry [i](#)

There are no ligands in this entry.

### 5.7 Other polymers [i](#)

There are no such residues in this entry.

### 5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.



## 6 Fit of model and data [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	A	246/301 (81%)	0.06	10 (4%) 37 36	105, 136, 202, 270	0
1	B	247/301 (82%)	0.05	16 (6%) 18 19	91, 145, 212, 269	0
1	C	238/301 (79%)	0.20	17 (7%) 16 17	108, 165, 239, 404	0
1	D	237/301 (78%)	0.36	22 (9%) 8 9	140, 183, 267, 341	0
All	All	968/1204 (80%)	0.17	65 (6%) 17 18	91, 158, 233, 404	0

The worst 5 of 65 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	6	TRP	15.3
1	C	6	TRP	8.8
1	C	224	PRO	6.5
1	A	5	ASP	5.8
1	B	272	GLN	5.8

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

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

### 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

### 6.4 Ligands [i](#)

There are no ligands in this entry.



## 6.5 Other polymers ⓘ

There are no such residues in this entry.