



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 07:34 pm GMT

PDB ID : 4HYD  
Title : Structure of a presenilin family intramembrane aspartate protease in C2221 space group  
Authors : Li, X.; Dang, S.; Yan, C.; Wang, J.; Shi, Y.  
Deposited on : 2012-11-13  
Resolution : 3.80 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

<http://wwpdb.org/validation/2016/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)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

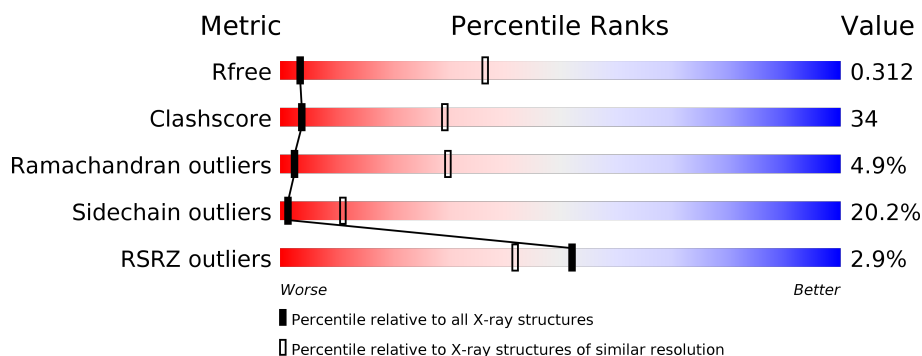
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.80 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))
$R_{free}$	100719	1019 (4.08-3.52)
Clashscore	112137	1030 (4.04-3.56)
Ramachandran outliers	110173	1011 (4.06-3.54)
Sidechain outliers	110143	1005 (4.06-3.54)
RSRZ outliers	101464	1032 (4.08-3.52)

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. 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>4%</div> <div> <div>34%</div> <div>37%</div> <div>10%</div> <div>18%</div> </div> </div>
1	B	301	<div> <div>%</div> <div> <div>36%</div> <div>37%</div> <div>8%</div> <div>18%</div> </div> </div>
1	C	301	<div> <div>4%</div> <div> <div>34%</div> <div>37%</div> <div>11%</div> <div>18%</div> </div> </div>
1	D	301	<div> <div>%</div> <div> <div>34%</div> <div>38%</div> <div>9%</div> <div>18%</div> </div> </div>

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 7199 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.

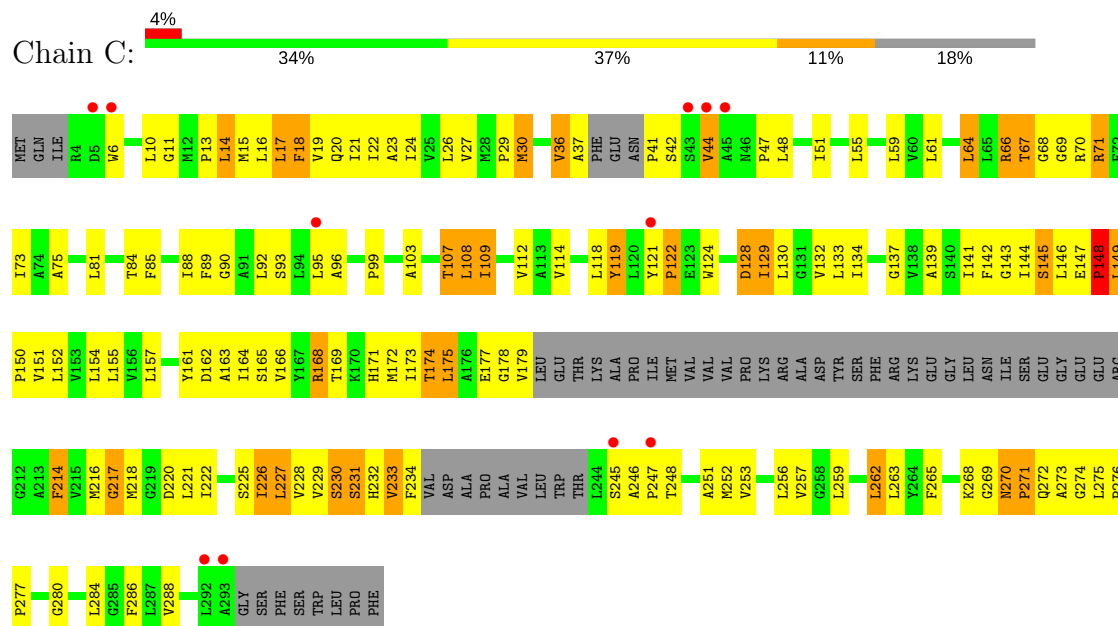
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	246	Total	C	N	O	S	0	0	0
			1802	1219	276	296	11			
1	B	246	Total	C	N	O	S	0	0	0
			1799	1216	276	296	11			
1	C	246	Total	C	N	O	S	0	0	0
			1799	1216	276	296	11			
1	D	246	Total	C	N	O	S	0	0	0
			1799	1216	276	296	11			

There are 20 discrepancies between the modelled and reference sequences:

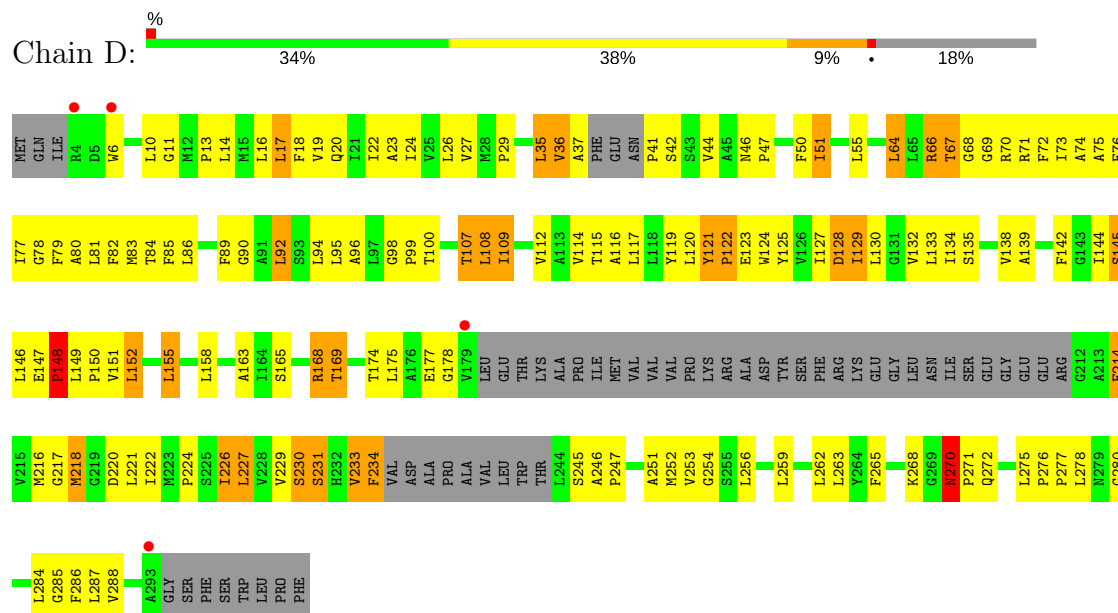
Chain	Residue	Modelled	Actual	Comment	Reference
A	40	ASN	ASP	ENGINEERED MUTATION	UNP A3CWV0
A	42	SER	GLU	ENGINEERED MUTATION	UNP A3CWV0
A	147	GLU	ALA	ENGINEERED MUTATION	UNP A3CWV0
A	148	PRO	VAL	ENGINEERED MUTATION	UNP A3CWV0
A	229	VAL	ALA	ENGINEERED MUTATION	UNP A3CWV0
B	40	ASN	ASP	ENGINEERED MUTATION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



• Molecule 1: Putative uncharacterized protein



• Molecule 1: Putative uncharacterized protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	114.08Å 170.03Å 181.40Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	31.58 – 3.80 31.58 – 3.79	Depositor EDS
% Data completeness (in resolution range)	84.7 (31.58-3.80) 83.9 (31.58-3.79)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.18 (at 3.75Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, $R_{free}$	0.307 , 0.308 0.302 , 0.312	Depositor DCC
$R_{free}$ test set	648 reflections (4.38%)	DCC
Wilson B-factor (Å <sup>2</sup> )	122.8	Xtriage
Anisotropy	0.572	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.19 , 102.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.42$ , $\langle L^2 \rangle = 0.25$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	7199	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	184.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 14.35% 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 [i](#)

### 5.1 Standard geometry [i](#)

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	0.49	0/1839	0.77	0/2511
1	B	0.47	0/1836	0.76	0/2507
1	C	0.44	0/1836	0.74	0/2507
1	D	0.51	0/1836	0.79	0/2507
All	All	0.48	0/7347	0.77	0/10032

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	1
1	B	0	1
1	C	0	1
1	D	0	1
All	All	0	4

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	145	SER	Peptide
1	B	145	SER	Peptide
1	C	145	SER	Peptide
1	D	145	SER	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	128	3
1	B	1799	0	1940	136	0
1	C	1799	0	1940	140	0
1	D	1799	0	1940	119	0
All	All	7199	0	7769	503	3

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 34.

The worst 5 of 503 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:36:VAL:CG2	1:B:144:ILE:HG22	1.65	1.25
1:C:36:VAL:CG2	1:C:144:ILE:HG22	1.68	1.21
1:C:36:VAL:HG12	1:C:37:ALA:H	1.01	1.17
1:B:36:VAL:HG22	1:B:144:ILE:HG22	1.18	1.15
1:C:36:VAL:HG13	1:C:145:SER:HA	1.15	1.13

All (3) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:292:LEU:CD2	1:A:292:LEU:CD2[3_654]	1.64	0.56
1:A:292:LEU:CD1	1:A:292:LEU:CD2[3_654]	1.90	0.30
1:A:292:LEU:CG	1:A:292:LEU:CD2[3_654]	2.00	0.20

## 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	30
1	B	238/301 (79%)	198 (83%)	29 (12%)	11 (5%)	3	31
1	C	238/301 (79%)	196 (82%)	31 (13%)	11 (5%)	3	31
1	D	238/301 (79%)	198 (83%)	27 (11%)	13 (6%)	2	28
All	All	952/1204 (79%)	788 (83%)	117 (12%)	47 (5%)	2	30

5 of 47 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	36	VAL
1	A	96	ALA
1	B	36	VAL
1	C	36	VAL
1	D	36	VAL

### 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%)	145 (77%)	43 (23%)	1	7
1	B	187/236 (79%)	152 (81%)	35 (19%)	2	13
1	C	187/236 (79%)	148 (79%)	39 (21%)	1	10
1	D	187/236 (79%)	153 (82%)	34 (18%)	2	14
All	All	749/944 (79%)	598 (80%)	151 (20%)	1	11

5 of 151 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	214	PHE
1	C	55	LEU
1	D	177	GLU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	218	MET
1	B	263	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	20	GLN

### 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%)	-0.42	12 (4%) 30 24	102, 158, 255, 434	0
1	B	246/301 (81%)	-0.41	2 (0%) 86 79	144, 184, 265, 350	0
1	C	246/301 (81%)	-0.37	11 (4%) 34 27	147, 196, 279, 349	0
1	D	246/301 (81%)	-0.52	4 (1%) 72 63	97, 154, 257, 361	0
All	All	984/1204 (81%)	-0.43	29 (2%) 52 42	97, 176, 268, 434	0

The worst 5 of 29 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	6	TRP	5.9
1	C	293	ALA	5.8
1	A	123	GLU	5.7
1	D	293	ALA	4.1
1	B	234	PHE	4.1

### 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 [i](#)

There are no such residues in this entry.