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

wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 05:21 PM GMT

PDB ID : 4HYC

Title : Structure of a presenilin family intramembrane aspartate protease in P2 space group

Authors : Li, X.; Dang, S.; Yan, C.; Wang, J.; Shi, Y.

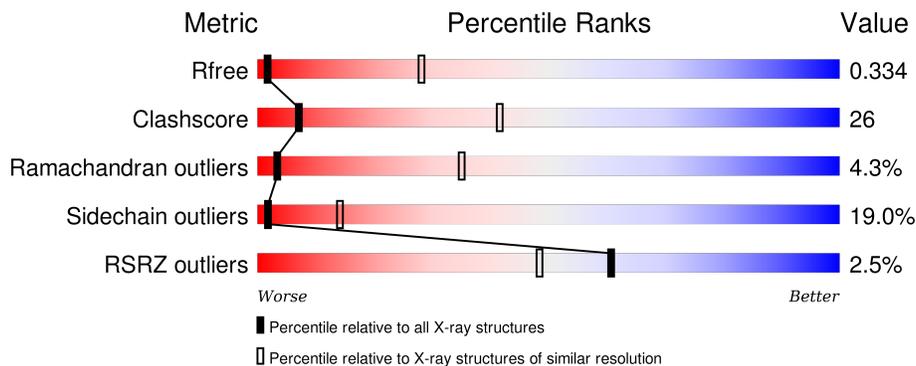Deposited on : 2012-11-13

Resolution : 3.95 Å(reported)

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.95 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	Percentile Ranks	Value
Rfree		0.334
Clashscore		26
Ramachandran outliers		4.3%
Sidechain outliers		19.0%
RSRZ outliers		2.5%

Worse   *Better*

■ Percentile relative to all X-ray structures
▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	1011 (4.38-3.54)
Clashscore	102246	1025 (4.32-3.60)
Ramachandran outliers	100387	1008 (4.34-3.58)
Sidechain outliers	100360	1027 (4.36-3.56)
RSRZ outliers	91569	1015 (4.38-3.54)

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 >=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 <=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	% — 38% / 36% / 7% / · / 18%
1	B	301	2% — 40% / 32% / 9% / 18%
1	C	301	2% — 40% / 34% / 7% / 18%
1	D	301	% — 36% / 37% / 9% / 18%
1	E	301	2% — 42% / 31% / 9% / 18%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	301	3% 37% 35% 9% 18%
1	G	301	% 37% 36% 9% 18%
1	H	301	5% 40% 34% 8% 18%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 14416 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 1802	C 1219	N 276	O 296	S 11	0	0	0
1	B	246	Total 1802	C 1219	N 276	O 296	S 11	0	0	0
1	C	246	Total 1802	C 1219	N 276	O 296	S 11	0	0	0
1	D	246	Total 1802	C 1219	N 276	O 296	S 11	0	0	0
1	E	246	Total 1802	C 1219	N 276	O 296	S 11	0	0	0
1	F	246	Total 1802	C 1219	N 276	O 296	S 11	0	0	0
1	G	246	Total 1802	C 1219	N 276	O 296	S 11	0	0	0
1	H	246	Total 1802	C 1219	N 276	O 296	S 11	0	0	0

There are 40 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

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
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
E	40	ASN	ASP	ENGINEERED MUTATION	UNP A3CWV0
E	42	SER	GLU	ENGINEERED MUTATION	UNP A3CWV0
E	147	GLU	ALA	ENGINEERED MUTATION	UNP A3CWV0
E	148	PRO	VAL	ENGINEERED MUTATION	UNP A3CWV0
E	229	VAL	ALA	ENGINEERED MUTATION	UNP A3CWV0
F	40	ASN	ASP	ENGINEERED MUTATION	UNP A3CWV0
F	42	SER	GLU	ENGINEERED MUTATION	UNP A3CWV0
F	147	GLU	ALA	ENGINEERED MUTATION	UNP A3CWV0
F	148	PRO	VAL	ENGINEERED MUTATION	UNP A3CWV0
F	229	VAL	ALA	ENGINEERED MUTATION	UNP A3CWV0
G	40	ASN	ASP	ENGINEERED MUTATION	UNP A3CWV0
G	42	SER	GLU	ENGINEERED MUTATION	UNP A3CWV0
G	147	GLU	ALA	ENGINEERED MUTATION	UNP A3CWV0
G	148	PRO	VAL	ENGINEERED MUTATION	UNP A3CWV0
G	229	VAL	ALA	ENGINEERED MUTATION	UNP A3CWV0
H	40	ASN	ASP	ENGINEERED MUTATION	UNP A3CWV0
H	42	SER	GLU	ENGINEERED MUTATION	UNP A3CWV0
H	147	GLU	ALA	ENGINEERED MUTATION	UNP A3CWV0
H	148	PRO	VAL	ENGINEERED MUTATION	UNP A3CWV0
H	229	VAL	ALA	ENGINEERED MUTATION	UNP A3CWV0

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-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. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

• Molecule 1: Putative uncharacterized protein

• Molecule 1: Putative uncharacterized protein

• Molecule 1: Putative uncharacterized protein

Chain C:

• Molecule 1: Putative uncharacterized protein

Chain D:

• Molecule 1: Putative uncharacterized protein

Chain E:

• Molecule 1: Putative uncharacterized protein

Chain F:

• Molecule 1: Putative uncharacterized protein

Chain G:

• Molecule 1: Putative uncharacterized protein

Chain H:

L158 A163 I164 S165 R168 T169 K170 H171 M172 I173 T174 L175 A176 E177 G178 V179 LEU GLU THR LYS ALA PRO ILE MET VAL VAL VAL PRO LYS ARG ALA ASP TYR SER PHE ARG LYS GLU GLY LEU ASN ILE SER GLU GLY GLU GLU ARG G212 A213 F214 V215 M216 G217 M218 G219 D220 L221

P224 S225 I226 L227 V228 V229 S230 S231 H232 V233 F234 VAL ASP ALA PRO ALA VAL LEU TRP THR L244 S245 A246 P247 A251 M252 S255 L256 V257 G258 L259 L262 L263 Y264 F265 V266 N267 K268 G269 N270 P271 L275 P276 P277 L278 L284 G285 F286 L287 A293 GLY SER PHE

SER TRP LEU PRO PHE

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	127.22Å 115.83Å 137.17Å 90.00° 100.59° 90.00°	Depositor
Resolution (Å)	40.16 – 3.95	Depositor
	40.16 – 3.95	EDS
% Data completeness	78.3 (40.16-3.95)	Depositor
(in resolution range)	78.5 (40.16-3.95)	EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.81 (at 4.00Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, R_{free}	0.308 , 0.341	Depositor
	0.306 , 0.334	DCC
R_{free} test set	1339 reflections (4.93%)	DCC
Wilson B-factor (Å^2)	123.3	Xtriage
Anisotropy	0.236	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.24 , 77.6	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< |L| > = 0.46, < L^2 > = 0.29$	Xtriage
Outliers	1 of 27265 reflections (0.004%)	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	14416	wwPDB-VP
Average B, all atoms (Å^2)	141.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.14% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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	0.42	0/1839	0.68	0/2511
1	B	0.44	0/1839	0.68	0/2511
1	C	0.43	0/1839	0.69	0/2511
1	D	0.43	0/1839	0.68	0/2511
1	E	0.44	0/1839	0.71	0/2511
1	F	0.42	0/1839	0.68	1/2511 (0.0%)
1	G	0.44	0/1839	0.71	0/2511
1	H	0.40	0/1839	0.65	1/2511 (0.0%)
All	All	0.43	0/14712	0.69	2/20088 (0.0%)

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	2
1	B	0	1
1	C	0	2
1	D	0	1
1	E	0	1
1	F	0	2
1	G	0	1
1	H	0	1
All	All	0	11

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	F	59	LEU	CA-CB-CG	5.51	127.97	115.30
1	H	59	LEU	CA-CB-CG	5.07	126.95	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	B	145	SER	Peptide
1	C	124	TRP	Peptide
1	C	145	SER	Peptide

5.2 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 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	109	0
1	B	1802	0	1949	98	0
1	C	1802	0	1949	101	0
1	D	1802	0	1949	109	0
1	E	1802	0	1949	100	0
1	F	1802	0	1949	109	0
1	G	1802	0	1949	101	0
1	H	1802	0	1949	103	0
All	All	14416	0	15592	789	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 26.

The worst 5 of 789 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:A:36:VAL:CG2	1:A:144:ILE:HG22	1.28	1.64
1:C:36:VAL:CG2	1:C:144:ILE:HG22	1.30	1.57
1:F:36:VAL:CG2	1:F:144:ILE:HG22	1.51	1.40
1:A:36:VAL:HG13	1:A:145:SER:CA	1.59	1.32
1:H:36:VAL:CG2	1:H:144:ILE:HG22	1.63	1.29

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

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%)	202 (85%)	26 (11%)	10 (4%)	3	35
1	B	238/301 (79%)	196 (82%)	31 (13%)	11 (5%)	3	33
1	C	238/301 (79%)	201 (84%)	27 (11%)	10 (4%)	3	35
1	D	238/301 (79%)	198 (83%)	29 (12%)	11 (5%)	3	33
1	E	238/301 (79%)	198 (83%)	30 (13%)	10 (4%)	3	35
1	F	238/301 (79%)	199 (84%)	29 (12%)	10 (4%)	3	35
1	G	238/301 (79%)	198 (83%)	30 (13%)	10 (4%)	3	35
1	H	238/301 (79%)	198 (83%)	30 (13%)	10 (4%)	3	35
All	All	1904/2408 (79%)	1590 (84%)	232 (12%)	82 (4%)	3	35

5 of 82 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	36	VAL
1	A	96	ALA
1	A	230	SER
1	B	36	VAL
1	B	96	ALA

5.3.2 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 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%)	152 (81%)	36 (19%)	2	14

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles		
1	B	188/236 (80%)	150 (80%)	38 (20%)	1	12	
1	C	188/236 (80%)	152 (81%)	36 (19%)	2	14	
1	D	188/236 (80%)	152 (81%)	36 (19%)	2	14	
1	E	188/236 (80%)	152 (81%)	36 (19%)	2	14	
1	F	188/236 (80%)	153 (81%)	35 (19%)	2	15	
1	G	188/236 (80%)	152 (81%)	36 (19%)	2	14	
1	H	188/236 (80%)	155 (82%)	33 (18%)	2	18	
All	All	1504/1888 (80%)	1218 (81%)	286 (19%)	2	14	

5 of 286 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	169	THR
1	E	129	ILE
1	H	128	ASP
1	D	214	PHE
1	D	287	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	G	171	HIS

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, 95th 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(Å²)	Q<0.9
1	A	246/301 (81%)	-0.38	4 (1%)	74	63	81, 117, 188, 281	0
1	B	246/301 (81%)	-0.27	5 (2%)	68	57	92, 120, 194, 266	0
1	C	246/301 (81%)	-0.33	5 (2%)	68	57	81, 117, 194, 279	0
1	D	246/301 (81%)	-0.27	4 (1%)	74	63	99, 124, 189, 250	0
1	E	246/301 (81%)	-0.23	5 (2%)	68	57	86, 137, 203, 289	0
1	F	246/301 (81%)	-0.24	9 (3%)	45	34	108, 150, 221, 269	0
1	G	246/301 (81%)	-0.31	4 (1%)	74	63	83, 126, 197, 287	0
1	H	246/301 (81%)	-0.12	14 (5%)	27	19	142, 173, 247, 299	0
All	All	1968/2408 (81%)	-0.27	50 (2%)	61	48	81, 133, 210, 299	0

The worst 5 of 50 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	H	35	LEU	14.0
1	H	34	GLY	7.4
1	F	35	LEU	6.8
1	E	179	VAL	5.5
1	E	68	GLY	4.8

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 (i)

There are no such residues in this entry.