Full wwPDB X-ray Structure Validation Report ⓘ

May 21, 2020 – 08:24 pm BST

PDB ID	:	4NOM
Title	:	Crystal structure of asparaginyl endopeptidase (AEP)/Legumain activated at pH 4.5
Authors	:	Zhao, L.; Hua, T.; Ru, H.; Ni, X.; Shaw, N.; Jiao, L.; Ding, W.; Qu, L.; Ouyang, S.; Liu, Z.J.
Deposited on	:	2013-11-19
Resolution	:	2.01 Å(reported)

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

We welcome your comments at *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 ①) were used in the production of this report:

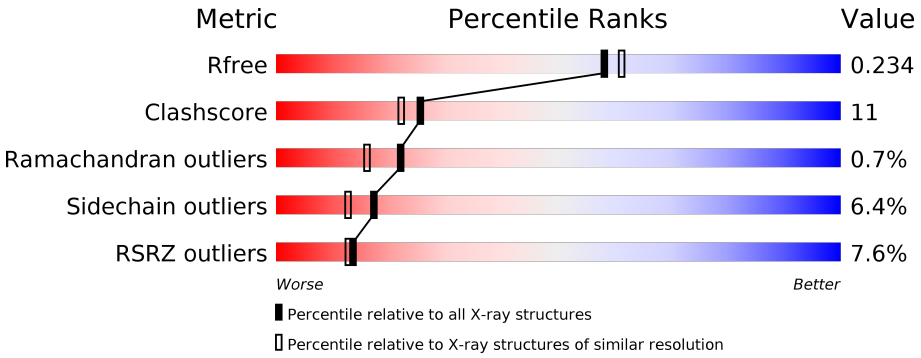
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.01 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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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 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 <=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	441	7% 74% 15% • 8%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3424 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 Legumain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	407	Total	C	N	O	S	0	0	0
			3266	2063	563	618	22			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	436	HIS	-	EXPRESSION TAG	UNP O89017
A	437	HIS	-	EXPRESSION TAG	UNP O89017
A	438	HIS	-	EXPRESSION TAG	UNP O89017
A	439	HIS	-	EXPRESSION TAG	UNP O89017
A	440	HIS	-	EXPRESSION TAG	UNP O89017
A	441	HIS	-	EXPRESSION TAG	UNP O89017

- Molecule 2 is water.

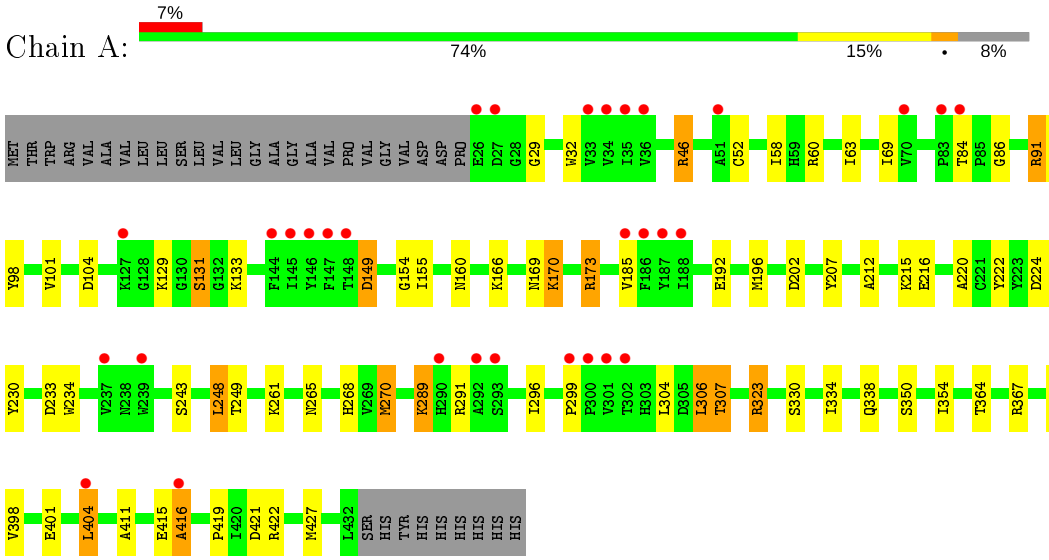
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	158	Total	O	0	0
			158	158		

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes 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: Legumain

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	57.95Å 167.72Å 51.93Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	32.62 – 2.01 32.62 – 2.01	Depositor EDS
% Data completeness (in resolution range)	92.3 (32.62-2.01) 85.1 (32.62-2.01)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.50 (at 2.01Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.3_928)	Depositor
R, R_{free}	0.204 , 0.243 0.199 , 0.234	Depositor DCC
R_{free} test set	1617 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	28.2	Xtriage
Anisotropy	0.725	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 44.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	3424	wwPDB-VP
Average B, all atoms (Å^2)	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.74% 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, 0.333 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	0.38	0/3349	0.54	0/4545

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

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	416	ALA	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	3266	0	3148	68	0
2	A	158	0	0	12	0
All	All	3424	0	3148	68	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 11.

All (68) 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:225:GLU:HB3	1:A:323:ARG:NH2	1.00	1.32
1:A:225:GLU:CB	1:A:323:ARG:NH2	1.94	1.31
1:A:155:ILE:N	1:A:196:MET:HE1	1.47	1.30
1:A:155:ILE:N	1:A:196:MET:CE	1.97	1.27
1:A:196:MET:SD	2:A:623:HOH:O	1.96	1.23
1:A:155:ILE:CA	1:A:196:MET:HE1	1.84	1.07
1:A:422:ARG:NH2	2:A:637:HOH:O	1.88	1.06
1:A:225:GLU:CB	1:A:323:ARG:HH21	1.58	1.03
1:A:154:GLY:C	1:A:196:MET:CE	2.37	0.93
1:A:154:GLY:C	1:A:196:MET:HE1	1.89	0.92
1:A:225:GLU:CD	1:A:323:ARG:HH21	1.74	0.90
1:A:225:GLU:CG	1:A:323:ARG:HH21	1.90	0.84
1:A:86:GLY:O	2:A:644:HOH:O	1.96	0.84
1:A:224:ASP:OD2	2:A:651:HOH:O	1.95	0.83
1:A:225:GLU:HB3	1:A:323:ARG:HH22	0.97	0.78
1:A:225:GLU:OE1	1:A:323:ARG:NH2	2.18	0.76
1:A:155:ILE:N	1:A:196:MET:HE2	2.00	0.76
1:A:154:GLY:C	1:A:196:MET:HE2	2.07	0.74
1:A:91:ARG:NH2	1:A:299:PRO:O	2.20	0.74
1:A:160:ASN:OD1	2:A:616:HOH:O	2.04	0.73
1:A:166:LYS:HE2	1:A:170:LYS:HE3	1.70	0.72
1:A:202:ASP:OD1	2:A:638:HOH:O	2.09	0.70
1:A:265:ASN:ND2	2:A:615:HOH:O	2.22	0.66
1:A:154:GLY:CA	1:A:196:MET:HE2	2.26	0.66
1:A:225:GLU:HB3	1:A:323:ARG:HH21	0.98	0.63
1:A:155:ILE:CA	1:A:196:MET:CE	2.63	0.61
1:A:155:ILE:H	1:A:196:MET:CE	2.04	0.61
1:A:169:ASN:OD1	1:A:173:ARG:NH2	2.30	0.60
1:A:169:ASN:ND2	2:A:646:HOH:O	2.35	0.60
1:A:350:SER:O	1:A:354:ILE:HG12	2.02	0.59
1:A:84:THR:HG21	1:A:227:ARG:HH21	1.67	0.59
1:A:98:TYR:O	1:A:101:VAL:HG22	2.03	0.59
1:A:212:ALA:HB1	1:A:216:GLU:HG3	1.85	0.58
1:A:225:GLU:CB	1:A:323:ARG:HH22	1.86	0.58
1:A:155:ILE:C	1:A:196:MET:CE	2.74	0.56
1:A:52:CYS:HB3	1:A:97:VAL:HG13	1.89	0.54
1:A:304:LEU:HB3	1:A:306:LEU:HD13	1.89	0.54
1:A:185:VAL:HA	1:A:207:TYR:O	2.10	0.52

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:52:CYS:CB	1:A:97:VAL:HG13	2.40	0.52
1:A:84:THR:HG21	1:A:227:ARG:NH2	2.25	0.51
1:A:398:VAL:HG13	2:A:649:HOH:O	2.12	0.50
1:A:69:ILE:HG13	1:A:133:LYS:HD2	1.93	0.49
1:A:58:ILE:HG23	1:A:63:ILE:HD12	1.94	0.49
1:A:155:ILE:HA	1:A:196:MET:HE1	1.85	0.49
1:A:155:ILE:C	1:A:196:MET:HE1	2.30	0.49
1:A:376:THR:HB	1:A:415:GLU:OE1	2.13	0.48
1:A:243:SER:HA	1:A:248:LEU:HD11	1.95	0.47
1:A:401:GLU:O	1:A:404:LEU:HB2	2.14	0.47
1:A:60:ARG:O	1:A:291:ARG:NH2	2.47	0.47
1:A:155:ILE:C	1:A:196:MET:HE3	2.35	0.47
1:A:29:GLY:O	1:A:289:LYS:HD3	2.15	0.46
1:A:46:ARG:HD3	1:A:220:ALA:CB	2.46	0.45
1:A:230:TYR:OH	1:A:307:THR:HG21	2.16	0.45
1:A:192:GLU:OE1	2:A:588:HOH:O	2.20	0.44
1:A:330:SER:O	1:A:334:ILE:HG12	2.17	0.44
1:A:104:ASP:HA	1:A:131:SER:HB2	1.99	0.44
1:A:419:PRO:HD2	1:A:422:ARG:HD3	2.01	0.43
1:A:354:ILE:CD1	1:A:427:MET:HA	2.48	0.43
1:A:131:SER:HB3	1:A:133:LYS:H	1.83	0.42
1:A:32:TRP:O	2:A:511:HOH:O	2.21	0.42
1:A:215:LYS:HE2	1:A:215:LYS:HB3	1.79	0.42
1:A:92:PRO:HD3	1:A:222:TYR:CG	2.54	0.42
1:A:46:ARG:HD2	1:A:233:ASP:OD1	2.20	0.42
1:A:289:LYS:HB2	1:A:289:LYS:HE2	1.91	0.41
1:A:398:VAL:CG1	2:A:649:HOH:O	2.67	0.41
1:A:411:ALA:O	1:A:415:GLU:HG2	2.21	0.41
1:A:268:HIS:O	1:A:270:MET:HG2	2.20	0.40
1:A:364:THR:HG23	1:A:367:ARG:HH21	1.87	0.40

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	405/441 (92%)	392 (97%)	10 (2%)	3 (1%)	22 16

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	149	ASP
1	A	46	ARG
1	A	416	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	359/387 (93%)	336 (94%)	23 (6%)	17 13

All (23) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	91	ARG
1	A	93	ASN
1	A	97	VAL
1	A	129	LYS
1	A	131	SER
1	A	149	ASP
1	A	170	LYS
1	A	173	ARG
1	A	226	GLU
1	A	234	TRP
1	A	248	LEU
1	A	249	THR
1	A	261	LYS
1	A	270	MET
1	A	289	LYS
1	A	296	ILE

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	306	LEU
1	A	307	THR
1	A	323	ARG
1	A	338	GLN
1	A	370	SER
1	A	404	LEU
1	A	421	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

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

5.5 Carbohydrates

There are no carbohydrates in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

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^{th} 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^2)	Q<0.9
1	A	407/441 (92%)	0.32	31 (7%) 13 13	25, 42, 71, 97	0

All (31) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	302	THR	4.1
1	A	26	GLU	3.9
1	A	301	VAL	3.7
1	A	290	HIS	3.7
1	A	416	ALA	3.6
1	A	300	PRO	3.5
1	A	36	VAL	3.5
1	A	145	ILE	3.4
1	A	187	TYR	3.4
1	A	147	PHE	3.4
1	A	148	THR	3.1
1	A	51	ALA	3.1
1	A	299	PRO	2.9
1	A	127	LYS	2.9
1	A	185	VAL	2.8
1	A	146	TYR	2.8
1	A	35	ILE	2.7
1	A	27	ASP	2.7
1	A	144	PHE	2.7
1	A	34	VAL	2.6
1	A	404	LEU	2.5
1	A	186	PHE	2.5
1	A	188	ILE	2.5
1	A	293	SER	2.3
1	A	84	THR	2.3
1	A	239	TRP	2.3
1	A	83	PRO	2.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	33	VAL	2.2
1	A	237	VAL	2.1
1	A	70	VAL	2.1
1	A	292	ALA	2.1

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.