WORLDWIDE PDB PROTEIN DATA BANK

Full wwPDB X-ray Structure Validation Report (i)

May 16, 2020 – 06:51 pm BST

PDB ID	:	3IUJ
Title	:	apPEP_WT2 opened state
Authors	:	Chiu, T.K.
Deposited on	:	2009-08-31
Resolution	:	1.80 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 (i) symbol.

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

|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
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 1.80 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.
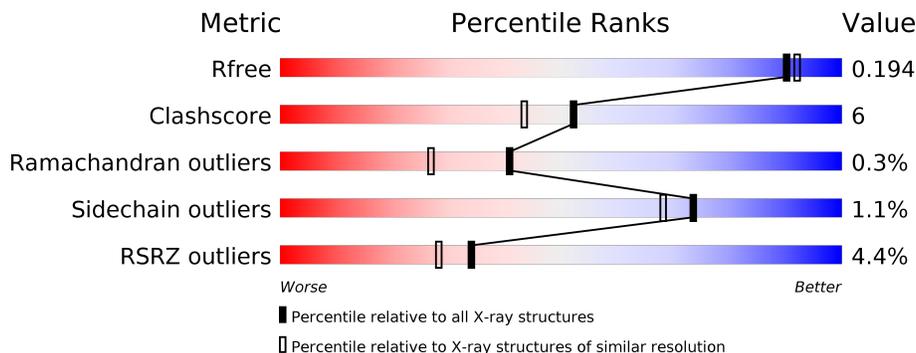
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)
RSRZ outliers	127900	5850 (1.80-1.80)

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	693	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 5948 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 Prolyl endopeptidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	670	Total	C	N	O	Se	0	0	0
			5296	3361	915	1012	8			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	EXPRESSION TAG	UNP Q9X6R4
A	-1	SER	-	EXPRESSION TAG	UNP Q9X6R4
A	0	HIS	-	EXPRESSION TAG	UNP Q9X6R4
A	106	GLN	LYS	SEE REMARK 999	UNP Q9X6R4
A	325	GLN	HIS	SEE REMARK 999	UNP Q9X6R4
A	326	GLN	ARG	SEE REMARK 999	UNP Q9X6R4
A	334	SER	THR	SEE REMARK 999	UNP Q9X6R4
A	335	GLY	ALA	SEE REMARK 999	UNP Q9X6R4
A	348	ARG	PRO	SEE REMARK 999	UNP Q9X6R4
A	577	THR	ALA	SEE REMARK 999	UNP Q9X6R4

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	652	Total	O	0	0
			652	652		

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 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: Prolyl endopeptidase
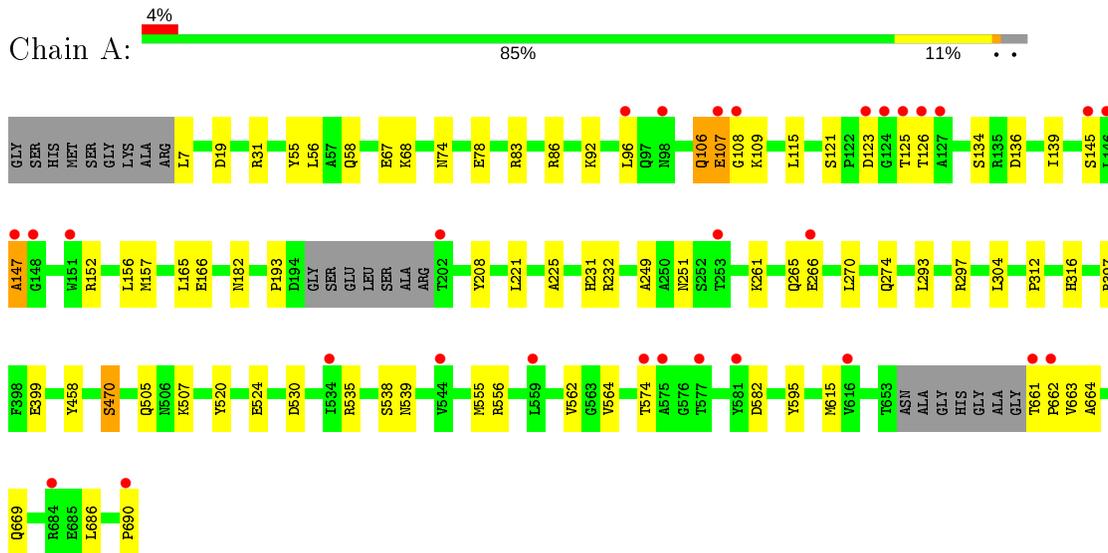
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants	62.64Å 85.13Å 148.17Å	Depositor
a, b, c, α, β, γ	90.00° 90.00° 90.00°	
Resolution (Å)	33.97 – 1.80	Depositor
	33.97 – 1.80	EDS
% Data completeness	93.9 (33.97-1.80)	Depositor
(in resolution range)	96.9 (33.97-1.80)	EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.07	Depositor
$<I/\sigma(I)>$ [1]	2.66 (at 1.81Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.181 , 0.218	Depositor
	0.192 , 0.194	DCC
R_{free} test set	6780 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å^2)	27.9	Xtriage
Anisotropy	0.398	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 42.1	EDS
L-test for twinning [2]	$<|L|> = 0.48$, $<L^2> = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	5948	wwPDB-VP
Average B, all atoms (Å^2)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.92% 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 (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	0.44	0/5429	0.69	1/7371 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	564	VAL	N-CA-C	-5.48	96.21	111.00

There are no chirality outliers.

There are no planarity outliers.

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	5296	0	5068	57	0
2	A	652	0	0	7	0
All	All	5948	0	5068	57	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 6.

All (57) 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:136:ASP:H	1:A:182:ASN:HD21	1.17	0.90

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:686:LEU:HD12	1:A:690:PRO:HG3	1.55	0.88
1:A:661:THR:HG23	1:A:662:PRO:CD	2.20	0.72
1:A:86:ARG:NH1	1:A:106:GLN:HE22	1.92	0.67
1:A:125:THR:HB	1:A:147:ALA:HB2	1.75	0.67
1:A:136:ASP:N	1:A:182:ASN:HD21	1.93	0.66
1:A:145:SER:HB3	1:A:152:ARG:CZ	2.25	0.66
1:A:67:GLU:HG3	2:A:1599:HOH:O	1.97	0.65
1:A:56:LEU:HD21	1:A:615:MSE:HE1	1.81	0.61
1:A:221:LEU:HD21	1:A:225:ALA:HB3	1.84	0.60
1:A:136:ASP:H	1:A:182:ASN:ND2	1.96	0.60
1:A:661:THR:HG23	1:A:662:PRO:HD3	1.82	0.59
1:A:86:ARG:NH1	1:A:106:GLN:NE2	2.51	0.59
1:A:106:GLN:OE1	1:A:109:LYS:HD2	2.06	0.56
1:A:535:ARG:HG3	2:A:1259:HOH:O	2.06	0.55
1:A:106:GLN:HG3	1:A:107:GLU:HG2	1.88	0.55
1:A:507:LYS:NZ	1:A:539:ASN:HD21	2.04	0.55
1:A:221:LEU:HD21	1:A:225:ALA:CB	2.37	0.54
1:A:139:ILE:HD12	1:A:165:LEU:HD22	1.88	0.54
1:A:156:LEU:HD12	1:A:166:GLU:HG3	1.90	0.53
1:A:193:PRO:HD3	1:A:208:TYR:CZ	2.43	0.53
1:A:265:GLN:HE21	1:A:265:GLN:HA	1.74	0.53
1:A:107:GLU:HG3	1:A:108:GLY:H	1.74	0.52
1:A:316:HIS:HD2	2:A:1552:HOH:O	1.93	0.51
1:A:669:GLN:NE2	2:A:1377:HOH:O	2.41	0.51
1:A:96:LEU:HD22	1:A:96:LEU:N	2.27	0.50
1:A:19:ASP:OD2	1:A:31:ARG:NH2	2.45	0.49
1:A:274:GLN:NE2	1:A:297:ARG:HH11	2.12	0.48
1:A:686:LEU:CD1	1:A:690:PRO:HG3	2.37	0.48
1:A:123:ASP:OD1	1:A:125:THR:HG23	2.14	0.47
1:A:397:ARG:HD3	1:A:399:GLU:OE2	2.14	0.47
1:A:261:LYS:HB2	1:A:270:LEU:HD23	1.96	0.47
1:A:55:TYR:O	1:A:58:GLN:HG2	2.15	0.47
1:A:121:SER:HB2	1:A:126:THR:OG1	2.14	0.47
1:A:530:ASP:HA	1:A:556:ARG:HB2	1.97	0.47
1:A:538:SER:HA	1:A:562:VAL:O	2.15	0.46
1:A:115:LEU:HD13	1:A:157:MSE:HE1	1.96	0.46
1:A:458:TYR:CZ	1:A:539:ASN:HB3	2.52	0.45
1:A:78:GLU:HG3	1:A:92:LYS:HG2	1.99	0.44
1:A:555:MSE:H	1:A:555:MSE:SE	2.51	0.44
1:A:661:THR:O	1:A:664:ALA:N	2.51	0.43
1:A:125:THR:HB	1:A:147:ALA:CB	2.44	0.43

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:83:ARG:HG3	1:A:83:ARG:HH11	1.83	0.43
1:A:134:SER:HB2	1:A:182:ASN:HD22	1.82	0.43
1:A:265:GLN:HA	1:A:265:GLN:NE2	2.34	0.42
1:A:505:GLN:HA	1:A:595:TYR:CE2	2.55	0.42
1:A:68:LYS:HE3	2:A:1525:HOH:O	2.20	0.42
1:A:507:LYS:HZ1	1:A:539:ASN:HD21	1.65	0.42
1:A:74:ASN:HA	2:A:1160:HOH:O	2.19	0.41
1:A:562:VAL:O	1:A:562:VAL:HG12	2.20	0.41
1:A:232:ARG:H	1:A:251:ASN:HD22	1.67	0.41
1:A:470:SER:HB3	2:A:1607:HOH:O	2.20	0.41
1:A:520:TYR:CZ	1:A:524:GLU:HG3	2.55	0.41
1:A:266:GLU:H	1:A:266:GLU:HG2	1.68	0.41
1:A:231:HIS:CG	1:A:249:ALA:HB1	2.56	0.40
1:A:293:LEU:O	1:A:304:LEU:HD12	2.20	0.40
1:A:661:THR:C	1:A:663:VAL:N	2.74	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	664/693 (96%)	648 (98%)	14 (2%)	2 (0%)	41	27

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	107	GLU
1	A	147	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	551/556 (99%)	545 (99%)	6 (1%)	73 68

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	7	LEU
1	A	106	GLN
1	A	312	PRO
1	A	470	SER
1	A	574	THR
1	A	582	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (10) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	98	ASN
1	A	105	GLN
1	A	182	ASN
1	A	251	ASN
1	A	265	GLN
1	A	274	GLN
1	A	487	ASN
1	A	504	GLN
1	A	506	ASN
1	A	539	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, 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(Å^2)	Q<0.9
1	A	662/693 (95%)	-0.07	29 (4%)	34	28	20, 29, 51, 72	0

All (29) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	661	THR	7.1
1	A	124	GLY	6.2
1	A	147	ALA	6.2
1	A	125	THR	6.1
1	A	107	GLU	5.5
1	A	574	THR	4.3
1	A	146	LEU	4.3
1	A	148	GLY	3.9
1	A	123	ASP	3.8
1	A	96	LEU	3.5
1	A	108	GLY	3.1
1	A	98	ASN	3.1
1	A	534	ILE	3.0
1	A	202	THR	3.0
1	A	559	LEU	2.8
1	A	616	VAL	2.8
1	A	127	ALA	2.6
1	A	266	GLU	2.6
1	A	145	SER	2.5
1	A	690	PRO	2.4
1	A	151	TRP	2.4
1	A	544	VAL	2.4
1	A	253	THR	2.3
1	A	581	TYR	2.2
1	A	575	ALA	2.2
1	A	126	THR	2.1
1	A	662	PRO	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	577	THR	2.0
1	A	684	ARG	2.0

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.