Full wwPDB X-ray Structure Validation Report ⓘ

May 16, 2020 – 12:53 am BST

PDB ID	:	4EP8
Title	:	Initial Urease Structure for Radiation Damage Experiment at 100 K
Authors	:	Warkentin, M.; Badeau, R.; Hopkins, J.B.; Thorne, R.E.
Deposited on	:	2012-04-17
Resolution	:	1.55 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 ⓘ) were used in the production of this report:

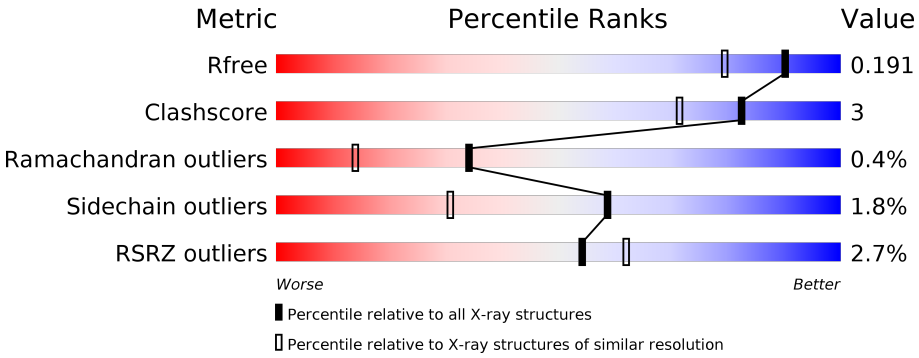
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 1.55 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	1483 (1.56-1.56)
Clashscore	141614	1529 (1.56-1.56)
Ramachandran outliers	138981	1498 (1.56-1.56)
Sidechain outliers	138945	1495 (1.56-1.56)
RSRZ outliers	127900	1465 (1.56-1.56)

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	C	566	3% 93% 6%
2	B	101	3% 94% 6%
3	A	100	% 93% 7%

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 6688 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 Urease subunit alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	566	Total	C	N	O	S	0	0	0
			4229	2652	741	813	23			

- Molecule 2 is a protein called Urease subunit beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	101	Total	C	N	O	S	0	0	0
			785	496	150	136	3			

- Molecule 3 is a protein called Urease subunit gamma.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	100	Total	C	N	O	S	0	0	0
			776	491	134	146	5			

- Molecule 4 is NICKEL (II) ION (three-letter code: NI) (formula: Ni).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	C	2	Total	Ni	0	0
			2	2		

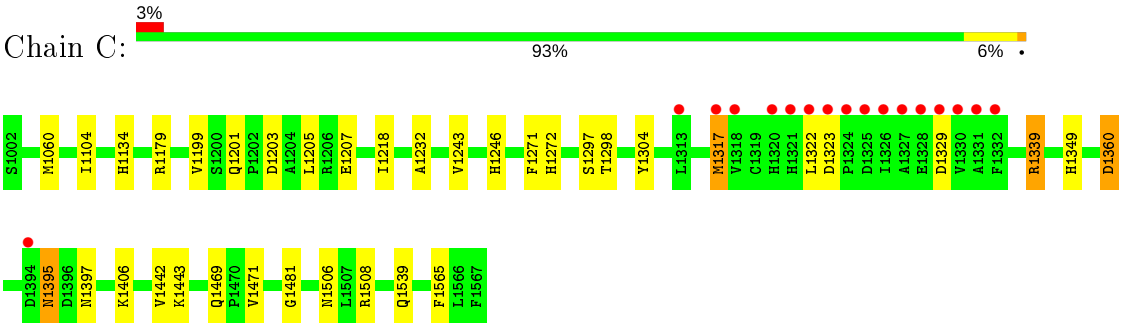
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	C	586	Total	O	0	0
			586	586		
5	B	160	Total	O	0	0
			160	160		
5	A	150	Total	O	0	0
			150	150		

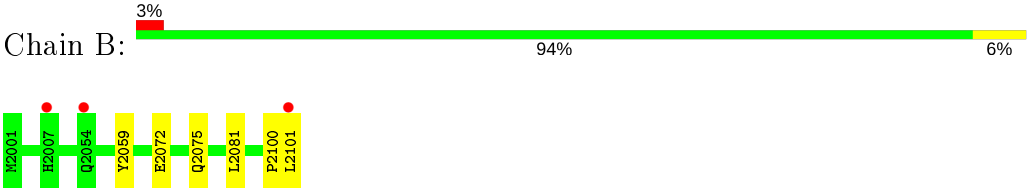
3 Residue-property plots (i)

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.

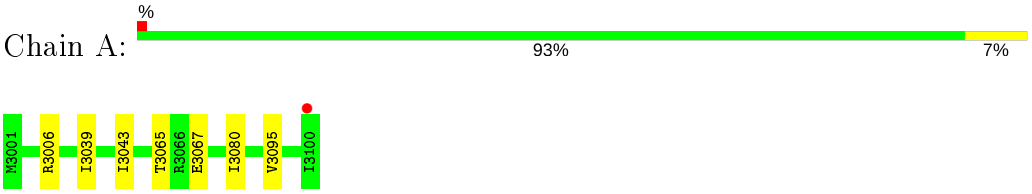
- Molecule 1: Urease subunit alpha



- Molecule 2: Urease subunit beta



- Molecule 3: Urease subunit gamma



4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 21 3	Depositor
Cell constants a, b, c, α, β, γ	168.89Å 168.89Å 168.89Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	15.62 – 1.55 15.55 – 1.55	Depositor EDS
% Data completeness (in resolution range)	(Not available) (15.62-1.55) 98.8 (15.55-1.55)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.73 (at 1.55Å)	Xtriage
Refinement program	BUSTER-TNT, BUSTER 2.8.0	Depositor
R, R_{free}	0.178 , 0.196 0.170 , 0.191	Depositor DCC
R_{free} test set	5704 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	19.1	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 53.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.52$, $< L^2 > = 0.36$	Xtriage
Estimated twinning fraction	0.006 for -l,-k,-h	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	6688	wwPDB-VP
Average B, all atoms (Å^2)	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.26% 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

Bond lengths and bond angles in the following residue types are not validated in this section: NI, KCX

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	C	0.56	0/4301	0.68	0/5860
2	B	0.51	0/805	0.67	0/1087
3	A	0.56	0/787	0.61	0/1061
All	All	0.55	0/5893	0.67	0/8008

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	C	4229	0	4175	22	0
2	B	785	0	772	4	0
3	A	776	0	804	3	0
4	C	2	0	0	0	0
5	A	150	0	0	0	0
5	B	160	0	0	0	0
5	C	586	0	0	0	0
All	All	6688	0	5751	29	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 3.

All (29) 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)
2:B:2072:GLU:H	2:B:2075:GLN:HE21	1.31	0.78
1:C:1201:GLN:HE21	1:C:1203:ASP:H	1.36	0.72
1:C:1395:ASN:HD22	1:C:1397:ASN:H	1.45	0.64
1:C:1442:VAL:HG13	1:C:1565:PHE:CZ	2.33	0.63
1:C:1218:ILE:HD13	1:C:1243:VAL:HG13	1.80	0.62
1:C:1297:SER:OG	1:C:1349:HIS:HE1	1.85	0.59
1:C:1218:ILE:HD11	1:C:1232:ALA:CB	2.33	0.58
1:C:1322:LEU:HB3	1:C:1329:ASP:HB3	1.90	0.53
1:C:1442:VAL:HG12	1:C:1443:LYS:HG3	1.92	0.52
3:A:3043:ILE:HD11	3:A:3095:VAL:HG21	1.92	0.51
1:C:1203:ASP:O	1:C:1207:GLU:HG3	2.11	0.51
1:C:1349:HIS:HD2	1:C:1406:LYS:HZ2	1.60	0.49
1:C:1349:HIS:HD2	1:C:1406:LYS:NZ	2.11	0.48
2:B:2072:GLU:H	2:B:2075:GLN:NE2	2.06	0.48
3:A:3065:THR:OG1	3:A:3067:GLU:HG2	2.14	0.47
2:B:2100:PRO:O	2:B:2101:LEU:HB2	2.14	0.46
2:B:2059:TYR:HB3	2:B:2081:LEU:HB3	1.97	0.46
1:C:1395:ASN:ND2	1:C:1397:ASN:H	2.13	0.44
3:A:3039:ILE:HB	3:A:3080:ILE:HD13	1.99	0.44
1:C:1317:MET:HG3	1:C:1322:LEU:HB2	2.00	0.43
1:C:1134:HIS:CD2	1:C:1360:ASP:HA	2.53	0.43
1:C:1134:HIS:CE1	1:C:1271:PHE:CD2	3.07	0.43
1:C:1304:TYR:OH	1:C:1339:ARG:HG3	2.19	0.42
1:C:1298:THR:CG2	1:C:1360:ASP:HB2	2.50	0.42
1:C:1469:GLN:NE2	1:C:1471:VAL:H	2.17	0.42
1:C:1271:PHE:O	1:C:1272:HIS:C	2.59	0.41
1:C:1442:VAL:HG13	1:C:1565:PHE:HZ	1.84	0.41
1:C:1199:VAL:HG12	1:C:1205:LEU:HD21	2.03	0.41
1:C:1317:MET:HA	1:C:1322:LEU:HD12	2.03	0.41

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	C	563/566 (100%)	535 (95%)	25 (4%)	3 (0%)	29 9
2	B	99/101 (98%)	94 (95%)	5 (5%)	0	100 100
3	A	98/100 (98%)	97 (99%)	1 (1%)	0	100 100
All	All	760/767 (99%)	726 (96%)	31 (4%)	3 (0%)	34 14

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	1360	ASP
1	C	1323	ASP
1	C	1481	GLY

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	C	442/442 (100%)	432 (98%)	10 (2%)	50 21
2	B	78/78 (100%)	78 (100%)	0	100 100
3	A	85/85 (100%)	84 (99%)	1 (1%)	71 49
All	All	605/605 (100%)	594 (98%)	11 (2%)	59 31

All (11) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	1060	MET
1	C	1104	ILE
1	C	1179	ARG
1	C	1246	HIS
1	C	1317	MET
1	C	1339	ARG
1	C	1395	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	1506	ASN
1	C	1508	ARG
1	C	1539	GLN
3	A	3006	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (12) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	1142	GLN
1	C	1166	HIS
1	C	1201	GLN
1	C	1349	HIS
1	C	1362	GLN
1	C	1395	ASN
1	C	1469	GLN
1	C	1499	ASN
2	B	2007	HIS
2	B	2016	ASN
2	B	2075	GLN
3	A	3081	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)

1 non-standard protein/DNA/RNA residue is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
1	KCX	C	1217	1,4	7,11,12	1.58	2 (28%)	4,12,14	1.25	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
1	KCX	C	1217	1,4	-	0/7/10/12	-

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	1217	KCX	CB-CA	2.80	1.57	1.53
1	C	1217	KCX	CE-NZ	2.05	1.50	1.45

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers (i)

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	C	565/566 (99%)	-0.37	17 (3%) 50 58	14, 19, 34, 84	0
2	B	101/101 (100%)	-0.10	3 (2%) 50 58	17, 23, 38, 50	0
3	A	100/100 (100%)	-0.46	1 (1%) 82 86	14, 19, 30, 38	0
All	All	766/767 (99%)	-0.34	21 (2%) 54 62	14, 19, 34, 84	0

All (21) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	1326	ILE	8.7
1	C	1322	LEU	8.2
1	C	1323	ASP	7.2
1	C	1327	ALA	6.0
1	C	1324	PRO	5.7
1	C	1318	VAL	5.1
1	C	1325	ASP	5.1
1	C	1330	VAL	4.4
2	B	2101	LEU	4.3
2	B	2007	HIS	3.6
1	C	1320	HIS	3.6
1	C	1331	ALA	3.5
1	C	1321	HIS	3.5
1	C	1317	MET	3.0
2	B	2054	GLN	3.0
1	C	1329	ASP	2.6
1	C	1313	LEU	2.3
1	C	1332	PHE	2.3
3	A	3100	ILE	2.2
1	C	1394	ASP	2.2
1	C	1328	GLU	2.0

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

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
1	KCX	C	1217	12/13	0.92	0.08	15,18,19,20	0

6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	NI	C	1601	1/1	0.99	0.15	27,27,27,27	0
4	NI	C	1602	1/1	0.99	0.17	24,24,24,24	0

6.5 Other polymers (i)

There are no such residues in this entry.