Preliminary Full wwPDB X-ray Structure Validation Report (i)

May 31, 2018 – 10:09 AM BST

Deposition ID : D_1200010294
PDB ID : *(not yet assigned)*

This is a Preliminary Full wwPDB X-ray Structure Validation Report.

This report is produced by the wwPDB Deposition System during initial deposition but before annotation of the structure.

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 (1)) were used in the production of this report:

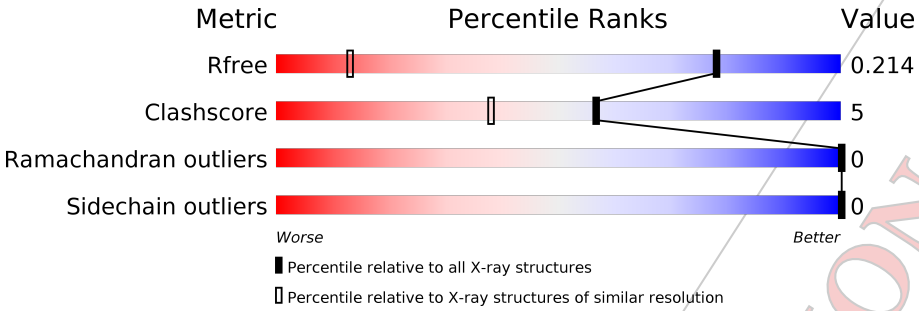
MolProbity	:	4.02b-467
Mogul	:	1.7.3 (157068), CSD as539be (2018)
Xtriage (Phenix)	:	1.13
EDS	:	rb-20031172
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac	:	5.8.0158
CCP4	:	7.0 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20031172

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 1.23 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}	111664	1699 (1.28-1.20)
Clashscore	122126	1783 (1.28-1.20)
Ramachandran outliers	120053	1724 (1.28-1.20)
Sidechain outliers	120020	1722 (1.28-1.20)

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%

Mol	Chain	Length	Quality of chain
1	A	129	92% 8%

2 Entry composition

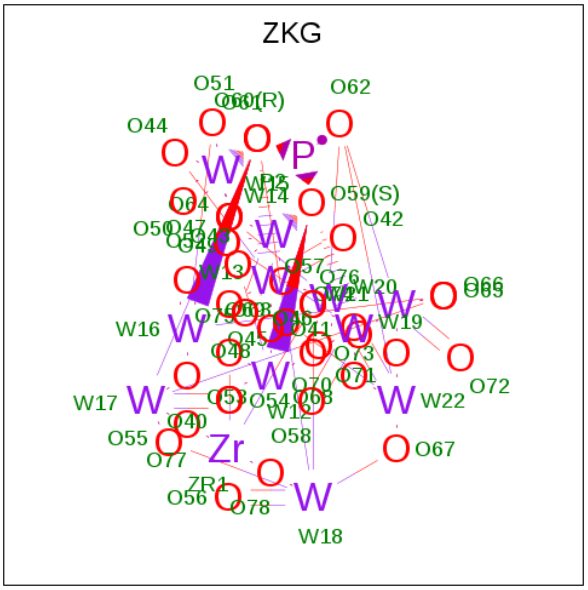
There are 4 unique types of molecules in this entry. The entry contains 2363 atoms, of which 988 are hydrogens and 0 are deuteriums.

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 Lysozyme C.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	129	2036	640	988	203	195	10	0	12	0

- Molecule 2 is ZIRCONIUM(IV) PHOSPHOTUNGSTATE KEGGIN (three-letter code: ZKG) (formula: $O_{39}PW_{11}Zr$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	O	P	W	Zr		
2	B	1	52	39	1	11	1	0	0
			Total	O	P	W	Zr		
2	B	1	104	78	2	22	2	0	1

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Cl		
3	C	1	1	1	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	C	1	Total Cl 1 1	0	0
3	C	1	Total Cl 1 1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	S	158	Total O 168 168	0	8

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. 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. 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: Lysozyme C

Chain A:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	77.70Å 77.70Å 37.87Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	38.85 – 1.23 38.85 – 1.23	Depositor EDS
% Data completeness (in resolution range)	99.8 (38.85-1.23) 95.6 (38.85-1.23)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.97 (at 1.23Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.147 , 0.171 0.192 , 0.214	Depositor DCC
R_{free} test set	1742 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å^2)	12.7	Xtriage
Anisotropy	0.170	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 54.1	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	0.94	EDS
Total number of atoms	2363	wwPDB-VP
Average B, all atoms (Å^2)	19.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.48% 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: ZKG, CL

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	0.46	0/1116	0.70	1/1508 (0.1%)

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	18	ASP	CB-CG-OD1	5.99	123.69	118.30

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	A	1048	988	948	8	0
2	B	156	0	0	4	0
3	C	3	0	0	1	0
4	S	168	0	0	7	0
All	All	1375	988	948	12	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 5.

All (12) 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:1:ZKG:O46	4:S:179:HOH:O	1.86	0.92
2:B:2[A]:ZKG:O77	4:S:43[A]:HOH:O	1.97	0.81
1:A:129:LEU:O	4:S:58[A]:HOH:O	1.99	0.80
2:B:1:ZKG:O75	4:S:179:HOH:O	2.10	0.70
3:C:2:CL:CL	4:S:196:HOH:O	2.50	0.67
1:A:65:ASN:OD1	2:B:2[B]:ZKG:ZR1	1.56	0.59
1:A:122:ALA:HA	1:A:125:ARG:HG3	1.86	0.58
1:A:14:ARG:NH2	4:S:102:HOH:O	2.38	0.50
1:A:33:LYS:HG2	1:A:123:TRP:CH2	2.53	0.43
1:A:122:ALA:HA	1:A:125:ARG:CG	2.49	0.42
1:A:62:TRP:HD1	4:S:193:HOH:O	2.04	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	A	139/129 (108%)	139 (100%)	0	0	100	100

There are no Ramachandran outliers to report.

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	117/105 (111%)	117 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

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

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)

Of 6 ligands modelled in this entry, 3 are monoatomic - leaving 3 for Mogul analysis.

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	$\#\|Z\| > 2$
2	ZKG	B	1	4	65,74,74	1.07	5 (7%)	6,222,222	2.04	2 (33%)
2	ZKG	B	2[A]	4	65,74,74	1.06	5 (7%)	6,222,222	1.51	1 (16%)
2	ZKG	B	2[B]	-	65,74,74	1.04	5 (7%)	6,222,222	1.82	2 (33%)

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
2	ZKG	B	1	4	-	0/0/420/420	0/0/23/23
2	ZKG	B	2[A]	4	-	0/0/420/420	0/0/23/23
2	ZKG	B	2[B]	-	-	0/0/420/420	0/0/23/23

All (15) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1	ZKG	W14-O43	-2.66	1.80	1.93
2	B	2[B]	ZKG	W15-O44	-2.55	1.81	1.93
2	B	2[A]	ZKG	W18-O61	-2.46	2.28	2.36
2	B	2[B]	ZKG	W14-O43	-2.43	1.81	1.93
2	B	2[A]	ZKG	W14-O43	-2.37	1.82	1.93
2	B	2[A]	ZKG	W15-O44	-2.29	1.82	1.93
2	B	1	ZKG	W15-O44	-2.19	1.83	1.93
2	B	2[B]	ZKG	W13-O44	2.06	2.03	1.93
2	B	2[B]	ZKG	P2-O59	2.09	1.60	1.54
2	B	1	ZKG	W12-O43	2.12	2.03	1.93
2	B	1	ZKG	P2-O59	2.17	1.60	1.54
2	B	2[B]	ZKG	W12-O43	2.18	2.04	1.93
2	B	2[A]	ZKG	W13-O44	2.32	2.04	1.93
2	B	2[A]	ZKG	W12-O43	2.38	2.04	1.93
2	B	1	ZKG	W13-O44	2.41	2.05	1.93

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	1	ZKG	O62-P2-O61	-3.86	104.28	110.57
2	B	2[B]	ZKG	O62-P2-O61	-3.62	104.67	110.57
2	B	2[A]	ZKG	O62-P2-O61	-3.03	105.64	110.57
2	B	1	ZKG	O60-P2-O61	2.34	114.38	110.57
2	B	2[B]	ZKG	O60-P2-O61	2.46	114.59	110.57

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	1	ZKG	2	0
2	B	2[A]	ZKG	1	0
2	B	2[B]	ZKG	1	0

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

6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.