



wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 22, 2023 – 12:13 am GMT

PDB ID : 7Z8A
Title : Sam68
Authors : Nadal, M.; Fuentes-Prior, P.
Deposited on : 2022-03-16
Resolution : 2.06 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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

MolProbity : 4.02b-467
Mogul : 1.8.4, CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.31.3
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) : 2.31.3

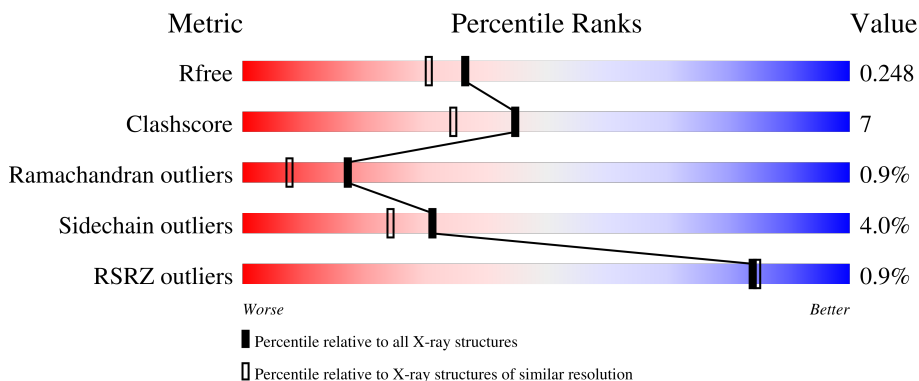
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.06 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	2684 (2.08-2.04)
Clashscore	141614	2801 (2.08-2.04)
Ramachandran outliers	138981	2768 (2.08-2.04)
Sidechain outliers	138945	2768 (2.08-2.04)
RSRZ outliers	127900	2646 (2.08-2.04)

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 of 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 $\leq 5\%$. 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	AAA	115	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	NO3	AAA	1004	-	-	-	X

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	NO3	AAA	1007	-	-	-	X

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 1034 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.

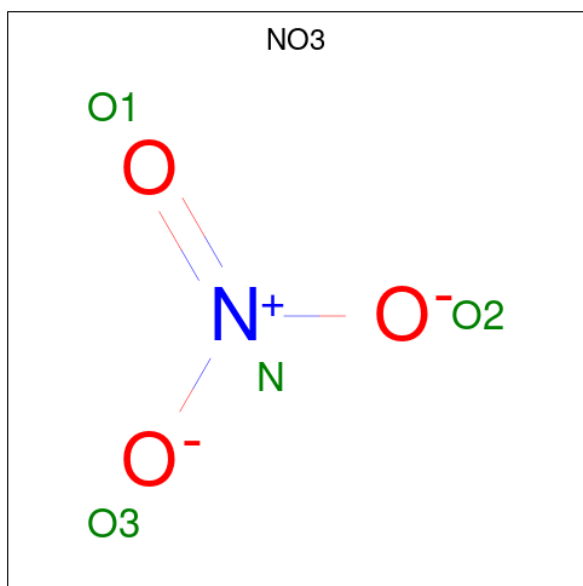
- Molecule 1 is a protein called KHDR1 protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	AAA	115	918	593	157	160	8	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	249	CYS	GLU	conflict	UNP A0A7K8A7M1

- Molecule 2 is NITRATE ION (three-letter code: NO3) (formula: NO₃).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	N	O		
2	AAA	1	4	1	3	0	0
2	AAA	1	4	1	3	0	0
2	AAA	1	4	1	3	0	0

Continued on next page...

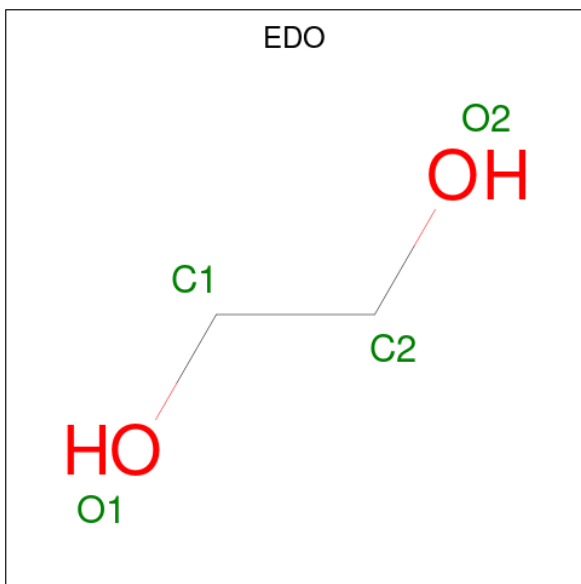
Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	AAA	1	Total	N	O	0	0
			4	1	3		
2	AAA	1	Total	N	O	0	0
			4	1	3		
2	AAA	1	Total	N	O	0	0
			4	1	3		
2	AAA	1	Total	N	O	0	0
			4	1	3		

- Molecule 3 is LITHIUM ION (three-letter code: LI) (formula: Li).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	AAA	1	Total	Li	0	0
			1	1		

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C₂H₆O₂).



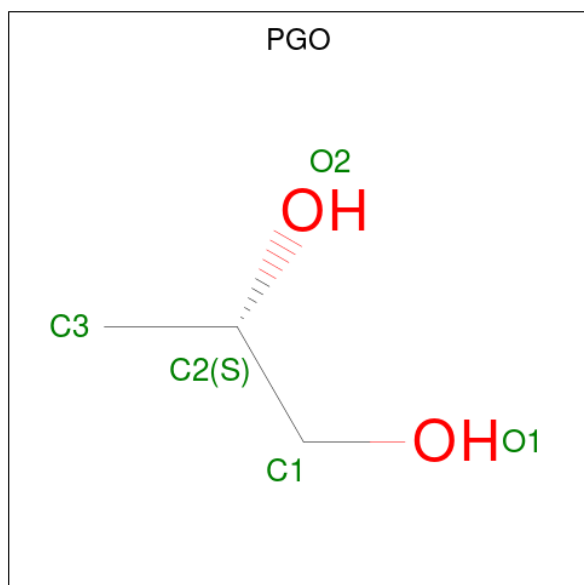
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	AAA	1	Total	C	O	0	0
			4	2	2		
4	AAA	1	Total	C	O	0	0
			4	2	2		
4	AAA	1	Total	C	O	0	0
			4	2	2		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	AAA	1	Total C O 4 2 2	0	0
4	AAA	1	Total C O 4 2 2	0	0

- Molecule 5 is S-1,2-PROPANEDIOL (three-letter code: PGO) (formula: C₃H₈O₂).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	AAA	1	Total C O 5 3 2	0	0
5	AAA	1	Total C O 5 3 2	0	0
5	AAA	1	Total C O 5 3 2	0	0


- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	AAA	48	Total O 48 48	0	0

3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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: KHDR1 protein

Chain AAA:  %



4 Data and refinement statistics

Property	Value	Source
Space group	P 41 2 2	Depositor
Cell constants a, b, c, α , β , γ	68.56Å 68.56Å 83.19Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.53 – 2.06 48.48 – 2.06	Depositor EDS
% Data completeness (in resolution range)	99.5 (48.53-2.06) 99.5 (48.48-2.06)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.72 (at 2.07Å)	Xtrriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.204 , 0.246 0.205 , 0.248	Depositor DCC
R_{free} test set	625 reflections (4.89%)	wwPDB-VP
Wilson B-factor (Å ²)	40.0	Xtrriage
Anisotropy	0.103	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 47.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	1034	wwPDB-VP
Average B, all atoms (Å ²)	45.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.22% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: EDO, NO3, PGO, LI

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	AAA	0.48	0/938	0.92	2/1254 (0.2%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	AAA	213	ARG	NE-CZ-NH1	-7.33	116.64	120.30
1	AAA	213	ARG	NE-CZ-NH2	6.80	123.70	120.30

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	AAA	918	0	949	12	1
2	AAA	32	0	0	2	0
3	AAA	1	0	0	0	0
4	AAA	20	0	30	4	1
5	AAA	15	0	24	0	0
6	AAA	48	0	0	1	0
All	All	1034	0	1003	13	1

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 7.

The worst 5 of 13 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)
4:AAA:1012:EDO:H11	6:AAA:1120:HOH:O	1.45	1.15
1:AAA:234:PHE:HD1	4:AAA:1011:EDO:H21	1.29	0.96
1:AAA:234:PHE:CD1	4:AAA:1011:EDO:H21	2.15	0.80
1:AAA:162:ILE:HD12	1:AAA:227:LEU:HG	1.77	0.67
1:AAA:154:MET:O	1:AAA:234:PHE:HA	2.07	0.55

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:AAA:224:ASN:ND2	4:AAA:1013:EDO:O1[7_554]	1.89	0.31

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	AAA	113/115 (98%)	110 (97%)	2 (2%)	1 (1%)	17 8

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	AAA	205	ASP

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	AAA	99/99 (100%)	95 (96%)	4 (4%)	31 24

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	AAA	150	SER
1	AAA	196	SER
1	AAA	208	LYS
1	AAA	259	MET

Sometimes 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 monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 17 ligands modelled in this entry, 1 is monoatomic - leaving 16 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	NO3	AAA	1008	-	1,3,3	0.10	0	0,3,3	-	-
2	NO3	AAA	1005	-	1,3,3	0.05	0	0,3,3	-	-
2	NO3	AAA	1007	-	1,3,3	0.03	0	0,3,3	-	-
2	NO3	AAA	1002	-	1,3,3	0.06	0	0,3,3	-	-
2	NO3	AAA	1004	-	1,3,3	0.11	0	0,3,3	-	-
5	PGO	AAA	1017	-	3,4,4	0.17	0	1,4,4	0.94	0
2	NO3	AAA	1003	-	1,3,3	0.07	0	0,3,3	-	-
4	EDO	AAA	1011	-	3,3,3	0.09	0	2,2,2	0.72	0
5	PGO	AAA	1015	-	3,4,4	0.15	0	1,4,4	0.80	0
4	EDO	AAA	1010	-	3,3,3	0.04	0	2,2,2	0.03	0
4	EDO	AAA	1012	-	3,3,3	0.37	0	2,2,2	0.90	0
4	EDO	AAA	1013	-	3,3,3	0.48	0	2,2,2	0.92	0
4	EDO	AAA	1014	-	3,3,3	0.19	0	2,2,2	0.42	0
2	NO3	AAA	1006	-	1,3,3	0.10	0	0,3,3	-	-
2	NO3	AAA	1001	-	1,3,3	0.31	0	0,3,3	-	-
5	PGO	AAA	1016	-	3,4,4	0.30	0	1,4,4	1.32	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
5	PGO	AAA	1017	-	-	1/2/2/2	-
4	EDO	AAA	1011	-	-	1/1/1/1	-
5	PGO	AAA	1015	-	-	1/2/2/2	-
4	EDO	AAA	1010	-	-	0/1/1/1	-
4	EDO	AAA	1013	-	-	1/1/1/1	-
4	EDO	AAA	1012	-	-	1/1/1/1	-
4	EDO	AAA	1014	-	-	0/1/1/1	-
5	PGO	AAA	1016	-	-	2/2/2/2	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	AAA	1016	PGO	O1-C1-C2-C3
5	AAA	1016	PGO	O1-C1-C2-O2
4	AAA	1013	EDO	O1-C1-C2-O2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
5	AAA	1017	PGO	O1-C1-C2-C3
4	AAA	1012	EDO	O1-C1-C2-O2

There are no ring outliers.

6 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	AAA	1008	NO3	1	0
2	AAA	1002	NO3	1	0
4	AAA	1011	EDO	2	0
4	AAA	1012	EDO	1	0
4	AAA	1013	EDO	0	1
4	AAA	1014	EDO	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 [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	AAA	115/115 (100%)	-0.11	1 (0%) 84 85	25, 38, 69, 80	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AAA	259	MET	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 monosaccharides 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, 95th 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
2	NO3	AAA	1002	4/4	0.75	0.33	76,78,81,88	0
2	NO3	AAA	1007	4/4	0.75	0.55	87,93,99,113	0
2	NO3	AAA	1004	4/4	0.79	0.44	102,107,109,110	0
2	NO3	AAA	1005	4/4	0.80	0.32	85,89,98,113	0
5	PGO	AAA	1017	5/5	0.80	0.19	61,70,84,90	0
4	EDO	AAA	1014	4/4	0.83	0.56	68,70,72,92	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
4	EDO	AAA	1011	4/4	0.83	0.23	58,59,61,68	0
5	PGO	AAA	1015	5/5	0.85	0.17	58,71,77,80	0
2	NO3	AAA	1008	4/4	0.89	0.28	73,78,82,103	0
5	PGO	AAA	1016	5/5	0.90	0.13	49,56,60,64	0
2	NO3	AAA	1006	4/4	0.90	0.32	78,82,87,93	0
2	NO3	AAA	1003	4/4	0.91	0.28	71,72,74,86	0
3	LI	AAA	1009	1/1	0.92	0.34	16,16,16,16	0
4	EDO	AAA	1010	4/4	0.93	0.15	76,78,79,80	0
4	EDO	AAA	1013	4/4	0.93	0.45	56,57,63,65	0
4	EDO	AAA	1012	4/4	0.98	0.29	42,45,47,55	0
2	NO3	AAA	1001	4/4	0.99	0.19	46,49,49,51	4

6.5 Other polymers [\(i\)](#)

There are no such residues in this entry.