



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 22, 2023 – 03:07 am GMT

PDB ID : 7ZAB  
Title : Sam68  
Authors : Nadal, M.; Puestes-Prior, P.  
Deposited on : 2022-03-22  
Resolution : 2.46 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](mailto: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

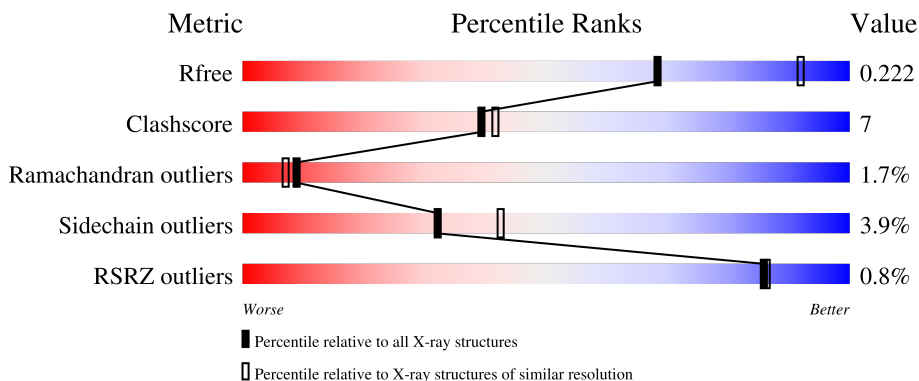
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*


The reported resolution of this entry is 2.46 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	1544 (2.48-2.44)
Clashscore	141614	1613 (2.48-2.44)
Ramachandran outliers	138981	1598 (2.48-2.44)
Sidechain outliers	138945	1598 (2.48-2.44)
RSRZ outliers	127900	1523 (2.48-2.44)

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  $\geq 3$ , 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	118	 87% 10% ..

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
4	LI	AAA	1005	-	-	-	X

## 2 Entry composition [i](#)

There are 9 unique types of molecules in this entry. The entry contains 1008 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 Isoform 2 of KH domain-containing, RNA-binding, signal transduction-associated protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	AAA	118	948	611	161	166	10	0	1	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	249	CYS	GLU	engineered mutation	UNP Q07666

- Molecule 2 is IODIDE ION (three-letter code: IOD) (formula: I).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	AAA	1	Total	I	0	0
			1	1		

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	AAA	3	Total	Cl	0	0
			3	3		

- Molecule 4 is LITHIUM ION (three-letter code: LI) (formula: Li).

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

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	AAA	1	Total	C	O	0	0
			6	3	3		

- Molecule 6 is (4S)-2-METHYL-2,4-PENTANEDIOL (three-letter code: MPD) (formula:  $C_6H_{14}O_2$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	AAA	1	Total	C	O	0	0
			8	6	2		

- Molecule 7 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula:  $C_2H_6O_2$ ).



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

- Molecule 8 is DIMETHYL SULFOXIDE (three-letter code: DMS) (formula: C<sub>2</sub>H<sub>6</sub>OS).



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


- Molecule 9 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
9	AAA	13	13	13	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: Isoform 2 of KH domain-containing, RNA-binding, signal transduction-associated protein 1

Chain AAA:  %



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	97.84Å 97.84Å 82.51Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	43.80 – 2.46 43.76 – 2.46	Depositor EDS
% Data completeness (in resolution range)	99.7 (43.80-2.46) 99.8 (43.76-2.46)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.28 (at 2.45Å)	Xtrriage
Refinement program	REFMAC 5.8.0267	Depositor
R, $R_{free}$	0.206 , 0.223 0.206 , 0.222	Depositor DCC
$R_{free}$ test set	776 reflections (5.16%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	85.8	Xtrriage
Anisotropy	0.002	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 70.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	1008	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	84.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.75% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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: DMS, CL, EDO, LI, MPD, IOD, GOL

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	0.53	0/969	0.96	0/1294

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 [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	948	0	977	15	0
2	AAA	1	0	0	0	0
3	AAA	3	0	0	0	0
4	AAA	1	0	0	0	0
5	AAA	6	0	8	2	0
6	AAA	8	0	14	5	0
7	AAA	24	0	36	0	0
8	AAA	4	0	6	0	0
9	AAA	13	0	0	0	0
All	All	1008	0	1041	15	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 7.

The worst 5 of 15 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:AAA:148:LEU:HD22	6:AAA:1007:MPD:H12	1.27	1.12
1:AAA:238:CYS:HB3	6:AAA:1007:MPD:H32	1.26	1.06
1:AAA:148:LEU:CD2	6:AAA:1007:MPD:H12	2.08	0.83
1:AAA:155:LYS:HD3	5:AAA:1006:GOL:H32	1.59	0.82
1:AAA:162:ILE:HD12	1:AAA:227:LEU:HG	1.85	0.59

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	AAA	116/118 (98%)	110 (95%)	4 (3%)	2 (2%)	<b>9</b> <b>7</b>

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	AAA	205	ASP
1	AAA	204	ARG

### 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.

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
-----	-------	----------	-----------	----------	-------------

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	AAA	103/102 (101%)	99 (96%)	4 (4%)	32 42

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

Mol	Chain	Res	Type
1	AAA	146	LEU
1	AAA	196	SER
1	AAA	204	ARG
1	AAA	206	LYS

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 14 ligands modelled in this entry, 5 are monoatomic - leaving 9 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
7	EDO	AAA	1012	-	3,3,3	0.62	0	2,2,2	0.80	0
7	EDO	AAA	1009	-	3,3,3	0.05	0	2,2,2	0.40	0
7	EDO	AAA	1011	-	3,3,3	0.22	0	2,2,2	1.17	0
8	DMS	AAA	1014	-	3,3,3	0.79	0	3,3,3	0.62	0
7	EDO	AAA	1008	-	3,3,3	0.16	0	2,2,2	0.40	0
5	GOL	AAA	1006	-	5,5,5	0.10	0	5,5,5	0.62	0
6	MPD	AAA	1007	-	7,7,7	1.37	2 (28%)	9,10,10	2.05	2 (22%)
7	EDO	AAA	1013	-	3,3,3	0.40	0	2,2,2	0.37	0
7	EDO	AAA	1010	-	3,3,3	0.14	0	2,2,2	0.62	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
7	EDO	AAA	1012	-	-	1/1/1/1	-
7	EDO	AAA	1009	-	-	1/1/1/1	-
7	EDO	AAA	1011	-	-	1/1/1/1	-
7	EDO	AAA	1008	-	-	1/1/1/1	-
5	GOL	AAA	1006	-	-	2/4/4/4	-
6	MPD	AAA	1007	-	-	0/5/5/5	-
7	EDO	AAA	1013	-	-	1/1/1/1	-
7	EDO	AAA	1010	-	-	1/1/1/1	-

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	AAA	1007	MPD	O2-C2	-2.77	1.37	1.44
6	AAA	1007	MPD	C1-C2	-2.07	1.46	1.52

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	AAA	1007	MPD	O2-C2-C1	-5.16	91.53	108.08
6	AAA	1007	MPD	O2-C2-CM	2.39	115.75	108.08

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	AAA	1006	GOL	O1-C1-C2-C3
5	AAA	1006	GOL	O1-C1-C2-O2
7	AAA	1008	EDO	O1-C1-C2-O2
7	AAA	1009	EDO	O1-C1-C2-O2
7	AAA	1010	EDO	O1-C1-C2-O2

There are no ring outliers.

2 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	AAA	1006	GOL	2	0
6	AAA	1007	MPD	5	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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	AAA	118/118 (100%)	0.17	1 (0%) 86   86	58, 75, 121, 152	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AAA	260[A]	MET	2.7

### 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
4	LI	AAA	1005	1/1	0.58	0.47	17,17,17,17	0
7	EDO	AAA	1010	4/4	0.68	0.25	117,121,132,135	0
8	DMS	AAA	1014	4/4	0.71	0.38	105,120,162,191	0
7	EDO	AAA	1012	4/4	0.77	0.26	97,99,112,127	0
3	CL	AAA	1003	1/1	0.79	0.19	114,114,114,114	0
7	EDO	AAA	1013	4/4	0.83	0.38	95,111,121,155	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
3	CL	AAA	1004	1/1	0.85	0.28	108,108,108,108	0
5	GOL	AAA	1006	6/6	0.86	0.26	99,105,135,158	0
7	EDO	AAA	1008	4/4	0.87	0.18	90,100,111,128	0
6	MPD	AAA	1007	8/8	0.89	0.34	43,76,91,119	0
3	CL	AAA	1002	1/1	0.90	0.06	111,111,111,111	0
7	EDO	AAA	1009	4/4	0.91	0.40	101,107,114,120	0
7	EDO	AAA	1011	4/4	0.93	0.23	77,88,98,109	0
2	IOD	AAA	1001	1/1	0.99	0.26	75,75,75,75	1

## 6.5 Other polymers [i](#)

There are no such residues in this entry.