wwPDB X-ray Structure Validation Summary Report (i)

Jan 22, 2023 – 12:14 am GMT

PDB ID	:	7Z9B
Title	:	Sam68
Authors	:	Nadal, M.; Fuentes-Prior, P.
Deposited on	:	2022-03-20
Resolution	:	3.33 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 (i) 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 (i)) were used in the production of this report:

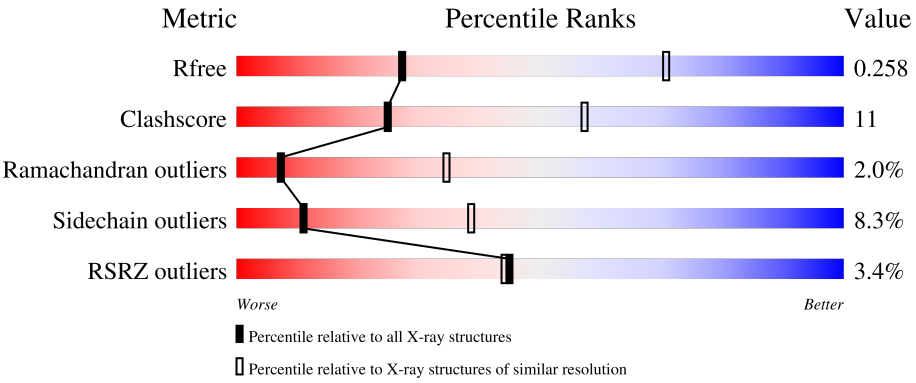
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.31.3
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.3

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.33 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	1060 (3.38-3.30)
Clashscore	141614	1111 (3.38-3.30)
Ramachandran outliers	138981	1090 (3.38-3.30)
Sidechain outliers	138945	1089 (3.38-3.30)
RSRZ outliers	127900	1028 (3.38-3.30)

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 <=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	116	3% 72% 22%
1	BBB	116	3% 72% 21% 6%
1	CCC	116	4% 72% 24%

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	EDO	CCC	1102	-	-	X	-

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 2832 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.

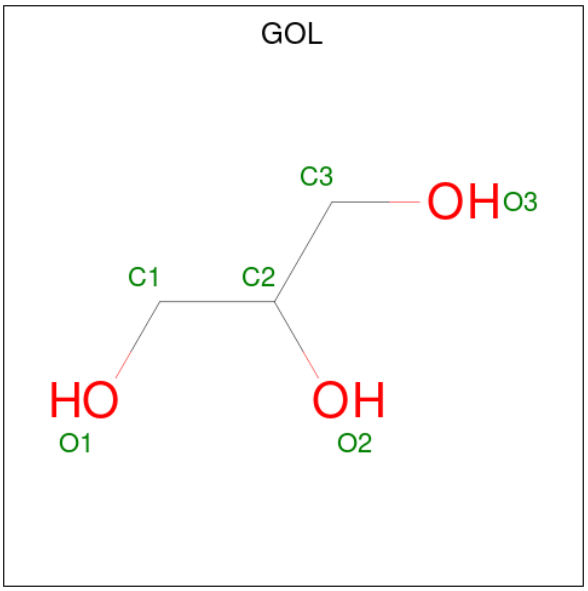
- Molecule 1 is a protein called KH domain-containing, RNA-binding, signal transduction-associated protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAA	116	Total	C	N	O	S	0	0	0
			928	599	159	163	7			
1	BBB	116	Total	C	N	O	S	0	0	0
			929	599	159	164	7			
1	CCC	116	Total	C	N	O	S	0	0	0
			929	599	159	164	7			

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

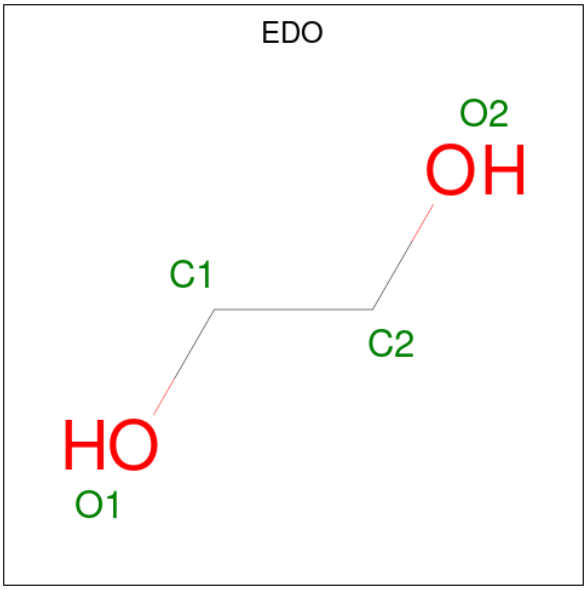
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	AAA	1	Total	I	0	0
			1	1		
2	BBB	1	Total	I	0	0
			1	1		

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



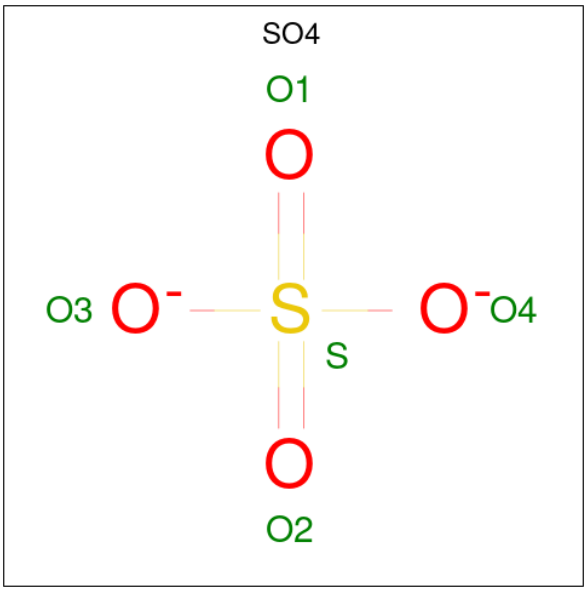
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	AAA	1	Total C O 6 3 3	0	0
3	AAA	1	Total C O 6 3 3	0	0
3	BBB	1	Total C O 6 3 3	0	0

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



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

- Molecule 5 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	CCC	1	Total O S 5 4 1	0	0

- Molecule 6 is water.

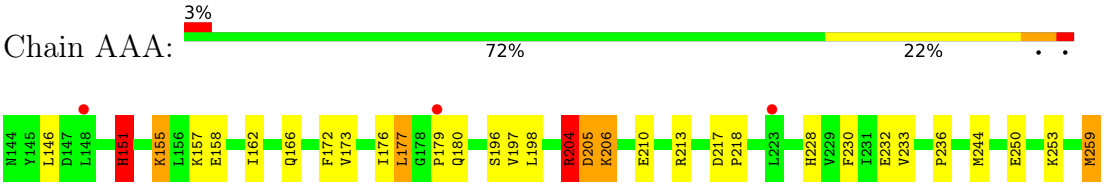
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	AAA	4	Total O 4 4	0	0
6	BBB	3	Total O 3 3	0	0
6	CCC	2	Total O 2 2	0	0

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.

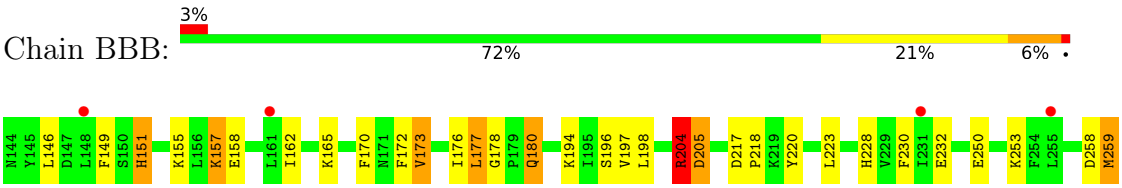
• Molecule 1: KH domain-containing, RNA-binding, signal transduction-associated protein 1

Chain AAA:



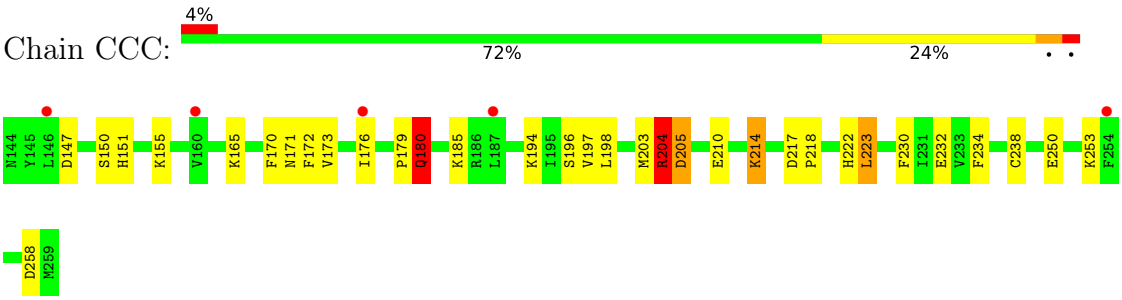
• Molecule 1: KH domain-containing, RNA-binding, signal transduction-associated protein 1

Chain BBB:



• Molecule 1: KH domain-containing, RNA-binding, signal transduction-associated protein 1

Chain CCC:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	136.73Å 237.59Å 82.32Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	118.80 – 3.33 118.80 – 3.33	Depositor EDS
% Data completeness (in resolution range)	99.4 (118.80-3.33) 98.6 (118.80-3.33)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.38 (at 3.33Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.217 , 0.250 0.229 , 0.258	Depositor DCC
R_{free} test set	1054 reflections (5.29%)	wwPDB-VP
Wilson B-factor (Å^2)	102.2	Xtriage
Anisotropy	0.236	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 92.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.28$	Xtriage
Estimated twinning fraction	0.206 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.260 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	2832	wwPDB-VP
Average B, all atoms (Å^2)	118.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.12% 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: EDO, GOL, SO4, IOD

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	0.43	0/948	1.02	3/1269 (0.2%)
1	BBB	0.45	0/949	1.05	3/1269 (0.2%)
1	CCC	0.41	0/949	1.00	2/1269 (0.2%)
All	All	0.43	0/2846	1.02	8/3807 (0.2%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	CCC	0	1

There are no bond length outliers.

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	AAA	151	HIS	CB-CA-C	9.06	128.51	110.40
1	BBB	151	HIS	CB-CA-C	7.87	126.14	110.40
1	CCC	151	HIS	CB-CA-C	7.63	125.67	110.40
1	BBB	204	ARG	CG-CD-NE	-6.28	98.61	111.80
1	CCC	179	PRO	C-N-CA	6.22	137.25	121.70

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	CCC	203	MET	Peptide

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	928	0	956	23	0
1	BBB	929	0	956	21	0
1	CCC	929	0	956	22	0
2	AAA	1	0	0	0	0
2	BBB	1	0	0	0	0
3	AAA	12	0	16	0	0
3	BBB	6	0	8	0	0
4	AAA	4	0	6	0	0
4	CCC	8	0	12	5	0
5	CCC	5	0	0	0	0
6	AAA	4	0	0	2	0
6	BBB	3	0	0	1	0
6	CCC	2	0	0	1	0
All	All	2832	0	2910	62	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 11.

The worst 5 of 62 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:BBB:259:MET:HG2	1:BBB:259:MET:O	1.44	1.09
1:AAA:210:GLU:OE2	1:AAA:213:ARG:NH2	1.84	1.09
1:BBB:173:VAL:O	1:BBB:177:LEU:HD22	1.68	0.92
1:AAA:173:VAL:O	1:AAA:177:LEU:HD22	1.70	0.92
1:BBB:259:MET:O	1:BBB:259:MET:CG	2.25	0.84

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	114/116 (98%)	105 (92%)	7 (6%)	2 (2%)	8 37
1	BBB	114/116 (98%)	105 (92%)	7 (6%)	2 (2%)	8 37
1	CCC	114/116 (98%)	104 (91%)	7 (6%)	3 (3%)	5 29
All	All	342/348 (98%)	314 (92%)	21 (6%)	7 (2%)	7 34

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	AAA	205	ASP
1	BBB	204	ARG
1	BBB	205	ASP
1	CCC	204	ARG
1	CCC	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	100/100 (100%)	92 (92%)	8 (8%)	12 40
1	BBB	100/100 (100%)	91 (91%)	9 (9%)	9 34
1	CCC	100/100 (100%)	92 (92%)	8 (8%)	12 40
All	All	300/300 (100%)	275 (92%)	25 (8%)	11 38

5 of 25 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	BBB	194	LYS
1	CCC	165	LYS
1	CCC	238	CYS
1	BBB	259	MET
1	CCC	180	GLN

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

5.3.3 RNA

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

There are no monosaccharides in this entry.

5.6 Ligand geometry

Of 9 ligands modelled in this entry, 2 are monoatomic - leaving 7 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	$\#\|Z\| > 2$
4	EDO	AAA	1004	-	3,3,3	0.40	0	2,2,2	0.48	0
3	GOL	AAA	1002	-	5,5,5	0.12	0	5,5,5	0.41	0
3	GOL	BBB	302	-	5,5,5	0.13	0	5,5,5	0.42	0
5	SO4	CCC	1101	-	4,4,4	0.37	0	6,6,6	0.10	0
3	GOL	AAA	1003	-	5,5,5	0.16	0	5,5,5	0.56	0
4	EDO	CCC	1102	-	3,3,3	0.16	0	2,2,2	0.22	0
4	EDO	CCC	1103	-	3,3,3	0.09	0	2,2,2	0.38	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
4	EDO	AAA	1004	-	-	0/1/1/1	-
3	GOL	AAA	1002	-	-	0/4/4/4	-
3	GOL	BBB	302	-	-	2/4/4/4	-
3	GOL	AAA	1003	-	-	2/4/4/4	-
4	EDO	CCC	1102	-	-	0/1/1/1	-
4	EDO	CCC	1103	-	-	0/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	BBB	302	GOL	O1-C1-C2-C3
3	BBB	302	GOL	O1-C1-C2-O2
3	AAA	1003	GOL	O2-C2-C3-O3
3	AAA	1003	GOL	O1-C1-C2-C3

There are no ring outliers.

2 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	CCC	1102	EDO	4	0
4	CCC	1103	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

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	AAA	116/116 (100%)	0.65	3 (2%)	56	54	82, 111, 147, 173	0
1	BBB	116/116 (100%)	0.60	4 (3%)	45	44	81, 114, 153, 173	0
1	CCC	116/116 (100%)	0.67	5 (4%)	35	36	83, 115, 149, 161	0
All	All	348/348 (100%)	0.64	12 (3%)	45	44	81, 114, 150, 173	0

The worst 5 of 12 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AAA	148	LEU	3.5
1	CCC	160	VAL	3.2
1	BBB	148	LEU	2.9
1	AAA	223	LEU	2.8
1	CCC	146	LEU	2.6

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

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
5	SO4	CCC	1101	5/5	0.56	0.14	220,232,244,252	0
4	EDO	CCC	1103	4/4	0.67	0.25	100,101,114,116	0
4	EDO	AAA	1004	4/4	0.72	0.40	104,112,119,122	0
4	EDO	CCC	1102	4/4	0.73	0.20	106,112,112,116	0
3	GOL	AAA	1002	6/6	0.84	0.15	119,120,128,138	0
3	GOL	AAA	1003	6/6	0.87	0.18	118,129,136,143	0
3	GOL	BBB	302	6/6	0.90	0.11	115,131,137,146	0
2	IOD	AAA	1001	1/1	0.98	0.27	110,110,110,110	1
2	IOD	BBB	301	1/1	0.99	0.31	106,106,106,106	0

6.5 Other polymers (i)

There are no such residues in this entry.