



wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 17, 2023 – 01:31 PM EST

PDB ID : 4WAN
Title : Crystal structure of Msl5 protein in complex with RNA at 1.8 Å
Authors : Jacewicz, A.; Smith, P.; Chico, L.; Schwer, B.; Shuman, S.
Deposited on : 2014-08-29
Resolution : 1.80 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 : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.36
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) : 2.36

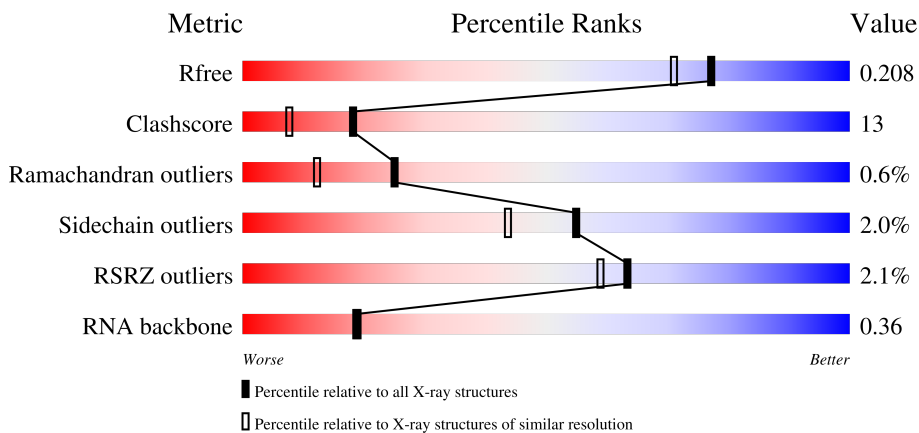
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.80 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.




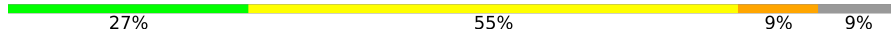
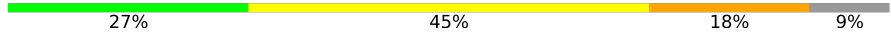

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)
RSRZ outliers	127900	5850 (1.80-1.80)
RNA backbone	3102	1060 (2.40-1.20)

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 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	A	129	 3% 76% 19% • •
1	C	129	 66% 30% • •
1	E	129	 5% 68% 26% • 5%
1	G	129	 3% 69% 19% • 10%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	B	11	
2	D	11	
2	F	11	
2	H	11	

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 4934 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 Branchpoint-bridging protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	124	Total 957	C 602	N 173	O 178	S 4	0	2	1
1	C	125	Total 962	C 602	N 172	O 184	S 4	0	0	0
1	E	122	Total 932	C 581	N 171	O 176	S 4	0	1	0
1	G	116	Total 906	C 572	N 164	O 167	S 3	0	2	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	143	SER	-	expression tag	UNP Q12186
C	143	SER	-	expression tag	UNP Q12186
E	143	SER	-	expression tag	UNP Q12186
G	143	SER	-	expression tag	UNP Q12186

- Molecule 2 is a RNA chain called RNA (5'-R(P*UP*AP*CP*UP*AP*AP*CP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	B	9	Total 151	C 66	N 25	O 52	P 8	0	0	1
2	D	10	Total 163	C 71	N 25	O 58	P 9	0	0	1
2	F	10	Total 162	C 71	N 25	O 57	P 9	0	0	1
2	H	9	Total 151	C 66	N 25	O 52	P 8	0	0	1

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



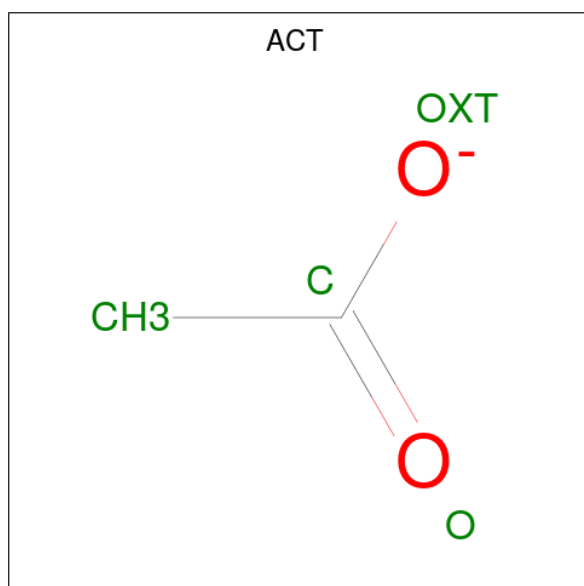
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	O	S	0	0
			5	4	1		
3	B	1	Total	O	S	0	0
			5	4	1		
3	G	1	Total	O	S	0	0
			5	4	1		
3	G	1	Total	O	S	0	0
			5	4	1		

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			6	3	3		
4	C	1	Total	C	O	0	0
			6	3	3		
4	C	1	Total	C	O	0	0
			6	3	3		
4	E	1	Total	C	O	0	0
			6	3	3		
4	E	1	Total	C	O	0	0
			6	3	3		

- Molecule 5 is ACETATE ION (three-letter code: ACT) (formula: C₂H₃O₂).



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

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	108	Total	O	0	3
			111	111		
6	B	30	Total	O	0	0
			30	30		
6	C	108	Total	O	0	3
			111	111		

Continued on next page...

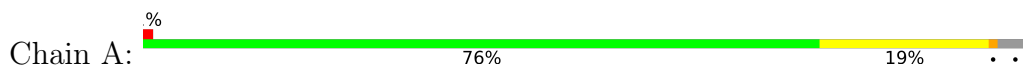
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	D	34	Total 36	O 36	0	2
6	E	92	Total 95	O 95	0	3
6	F	17	Total 17	O 17	0	0
6	G	74	Total 78	O 78	0	4
6	H	14	Total 14	O 14	0	0

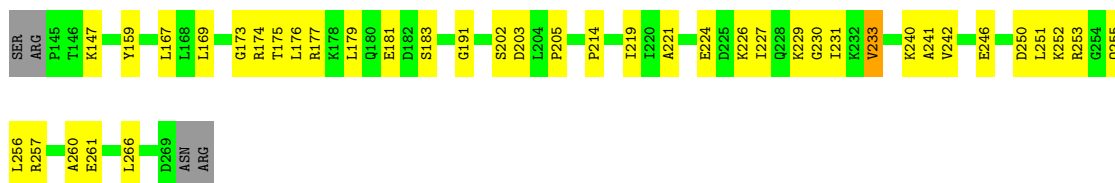
3 Residue-property plots [i](#)

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.

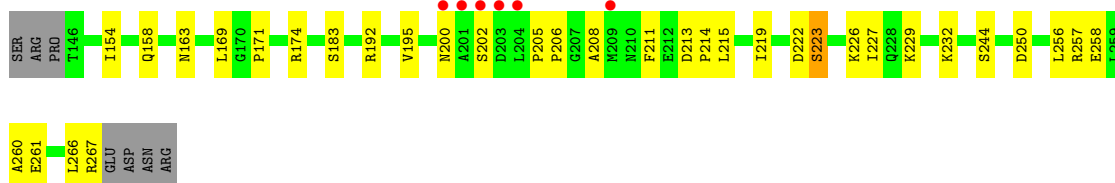
- Molecule 1: Branchpoint-bridging protein



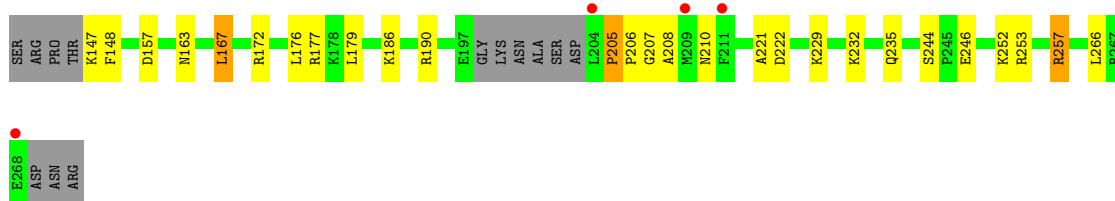
- Molecule 1: Branchpoint-bridging protein



- Molecule 1: Branchpoint-bridging protein

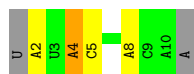


- Molecule 1: Branchpoint-bridging protein



- Molecule 2: RNA (5'-R(P*UP*AP*CP*UP*AP*AP*CP*A)-3')

Chain B: 



- Molecule 2: RNA (5'-R(P*UP*AP*CP*UP*AP*AP*CP*A)-3')

Chain D: 



- Molecule 2: RNA (5'-R(P*UP*AP*CP*UP*AP*AP*CP*A)-3')

Chain F: 



- Molecule 2: RNA (5'-R(P*UP*AP*CP*UP*AP*AP*CP*A)-3')

Chain H: 



4 Data and refinement statistics i

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	53.38Å 130.97Å 53.37Å 90.00° 117.43° 90.00°	Depositor
Resolution (Å)	45.61 – 1.80 45.61 – 1.80	Depositor EDS
% Data completeness (in resolution range)	99.2 (45.61-1.80) 99.2 (45.61-1.80)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.64 (at 1.81Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE: 1.8.4_1496)	Depositor
R, R_{free}	0.175 , 0.206 0.177 , 0.208	Depositor DCC
R_{free} test set	2988 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å ²)	26.1	Xtriage
Anisotropy	0.035	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 58.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.45$, $\langle L^2 \rangle = 0.27$	Xtriage
Estimated twinning fraction	0.036 for l,k,-h-l 0.036 for -h-l,k,h 0.043 for -h-l,-k,l 0.039 for h,-k,-h-l 0.457 for l,-k,h	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	4934	wwPDB-VP
Average B, all atoms (Å ²)	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.49% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: SO4, GOL, ACT

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	0.52	0/977	0.55	0/1316
1	C	0.52	0/976	0.52	0/1314
1	E	0.48	0/945	0.54	0/1273
1	G	0.47	0/924	0.54	0/1243
2	B	1.09	1/167 (0.6%)	1.01	0/259
2	D	1.03	1/180 (0.6%)	1.01	0/279
2	F	0.97	1/179 (0.6%)	0.93	0/277
2	H	1.01	1/167 (0.6%)	0.89	0/259
All	All	0.61	4/4515 (0.1%)	0.63	0/6220

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	H	2	A	O3'-P	-10.70	1.48	1.61
2	B	2	A	O3'-P	-10.67	1.48	1.61
2	D	1	U	O3'-P	-10.59	1.48	1.61
2	F	1	U	O3'-P	-10.45	1.48	1.61

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	A	957	0	981	20	1
1	C	962	0	975	28	1
1	E	932	0	938	32	0
1	G	906	0	934	27	0
2	B	151	0	75	4	0
2	D	163	0	81	5	0
2	F	162	0	81	9	0
2	H	151	0	75	7	0
3	A	5	0	0	0	0
3	B	5	0	0	1	0
3	G	10	0	0	1	0
4	A	6	0	8	0	0
4	C	12	0	16	2	0
4	E	12	0	16	0	0
5	A	4	0	3	0	0
5	C	4	0	3	1	0
6	A	111	0	0	7	0
6	B	30	0	0	1	0
6	C	111	0	0	9	1
6	D	36	0	0	0	0
6	E	95	0	0	6	1
6	F	17	0	0	1	0
6	G	78	0	0	10	0
6	H	14	0	0	1	0
All	All	4934	0	4186	116	2

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 13.

The worst 5 of 116 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:A:151:LYS:NZ	1:A:213:ASP:OD2	2.00	0.94
1:G:172:ARG:NH1	6:G:433:HOH:O	2.14	0.79
1:A:186:LYS:NZ	6:A:452:HOH:O	2.17	0.77
1:E:267:ARG:NH1	2:F:3:U:OP2	2.18	0.77
1:C:246:GLU:O	6:C:504:HOH:O	2.06	0.73

All (2) 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)
6:C:408:HOH:O	6:E:412:HOH:O[1_655]	2.15	0.05
1:A:150:ASP:OD1	1:C:202:SER:OG[2_646]	2.18	0.02

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	124/129 (96%)	120 (97%)	2 (2%)	2 (2%)	9	2
1	C	123/129 (95%)	119 (97%)	4 (3%)	0	100	100
1	E	120/129 (93%)	116 (97%)	4 (3%)	0	100	100
1	G	114/129 (88%)	107 (94%)	6 (5%)	1 (1%)	17	6
All	All	481/516 (93%)	462 (96%)	16 (3%)	3 (1%)	25	12

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	246	GLU
1	G	205	PRO
1	A	247	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	A	104/112 (93%)	102 (98%)	2 (2%)	57	46
1	C	105/112 (94%)	103 (98%)	2 (2%)	57	46

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	E	100/112 (89%)	98 (98%)	2 (2%)	55	44
1	G	99/112 (88%)	97 (98%)	2 (2%)	55	44
All	All	408/448 (91%)	400 (98%)	8 (2%)	55	44

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	257	ARG
1	G	167	LEU
1	E	223	SER
1	C	250	ASP
1	E	250	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	E	200	ASN

5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	B	6/11 (54%)	1 (16%)	0
2	D	8/11 (72%)	2 (25%)	0
2	F	7/11 (63%)	2 (28%)	0
2	H	7/11 (63%)	2 (28%)	0
All	All	28/44 (63%)	7 (25%)	0

5 of 7 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	B	4	A
2	D	2	A
2	D	4	A
2	F	3	U
2	F	4	A

There are no RNA pucker outliers to report.

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](#)

11 ligands are 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
5	ACT	C	303	-	3,3,3	0.76	0	3,3,3	1.48	0
4	GOL	C	301	-	5,5,5	0.44	0	5,5,5	0.44	0
3	SO4	B	101	-	4,4,4	0.17	0	6,6,6	0.09	0
3	SO4	A	301	-	4,4,4	0.17	0	6,6,6	0.13	0
4	GOL	C	302	-	5,5,5	0.46	0	5,5,5	0.75	0
3	SO4	G	301	-	4,4,4	0.16	0	6,6,6	0.25	0
3	SO4	G	302	-	4,4,4	0.19	0	6,6,6	0.30	0
5	ACT	A	303	-	3,3,3	0.76	0	3,3,3	1.38	0
4	GOL	E	302	-	5,5,5	0.48	0	5,5,5	0.10	0
4	GOL	E	301	-	5,5,5	0.38	0	5,5,5	0.31	0
4	GOL	A	302	-	5,5,5	0.40	0	5,5,5	0.44	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	GOL	C	301	-	-	2/4/4/4	-
4	GOL	C	302	-	-	2/4/4/4	-
4	GOL	E	302	-	-	2/4/4/4	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	GOL	E	301	-	-	1/4/4/4	-
4	GOL	A	302	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	302	GOL	C1-C2-C3-O3
4	C	302	GOL	C1-C2-C3-O3
4	C	302	GOL	O2-C2-C3-O3
4	C	301	GOL	O1-C1-C2-C3
4	E	302	GOL	C1-C2-C3-O3

There are no ring outliers.

4 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	C	303	ACT	1	0
3	B	101	SO4	1	0
4	C	302	GOL	2	0
3	G	302	SO4	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	A	124/129 (96%)	-0.58	1 (0%) 86 84	13, 24, 51, 75	0
1	C	125/129 (96%)	-0.59	0 100 100	14, 23, 62, 89	0
1	E	122/129 (94%)	-0.28	6 (4%) 29 24	18, 31, 84, 118	0
1	G	116/129 (89%)	-0.33	4 (3%) 45 39	19, 33, 88, 108	0
2	B	9/11 (81%)	-0.85	0 100 100	19, 24, 64, 75	0
2	D	10/11 (90%)	-0.51	0 100 100	21, 39, 104, 134	0
2	F	10/11 (90%)	-0.61	0 100 100	34, 50, 105, 106	0
2	H	9/11 (81%)	-0.81	0 100 100	29, 43, 74, 87	0
All	All	525/560 (93%)	-0.46	11 (2%) 63 59	13, 28, 83, 134	0

The worst 5 of 11 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	203	ASP	5.0
1	A	268	GLU	4.9
1	E	201	ALA	4.1
1	G	204	LEU	3.5
1	G	268	GLU	3.0

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
5	ACT	C	303	4/4	0.89	0.10	48,49,50,53	0
4	GOL	A	302	6/6	0.94	0.07	27,28,30,32	0
4	GOL	C	302	6/6	0.94	0.09	28,29,33,36	0
4	GOL	E	302	6/6	0.94	0.08	35,37,38,40	0
3	SO4	B	101	5/5	0.94	0.10	73,73,73,74	0
4	GOL	E	301	6/6	0.95	0.07	42,45,46,47	0
5	ACT	A	303	4/4	0.96	0.07	28,29,29,30	0
3	SO4	A	301	5/5	0.96	0.10	55,57,59,60	0
4	GOL	C	301	6/6	0.97	0.07	26,28,34,38	0
3	SO4	G	302	5/5	0.97	0.08	38,41,44,49	0
3	SO4	G	301	5/5	0.97	0.06	32,40,41,41	0

6.5 Other polymers [i](#)

There are no such residues in this entry.