Full wwPDB X-ray Structure Validation Report ⓘ

Aug 30, 2020 – 11:34 AM BST

PDB ID	:	3KMH
Title	:	Crystal Structure of a Novel Sugar Isomerase from E. coli O157:H7
Authors	:	van Staalduinen, L.M.; Jia, Z.; Montreal-Kingston Bacterial Structural Genomics Initiative (BSGI)
Deposited on	:	2009-11-10
Resolution	:	1.58 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 following versions of software and data (see references ⓘ) were used in the production of this report:

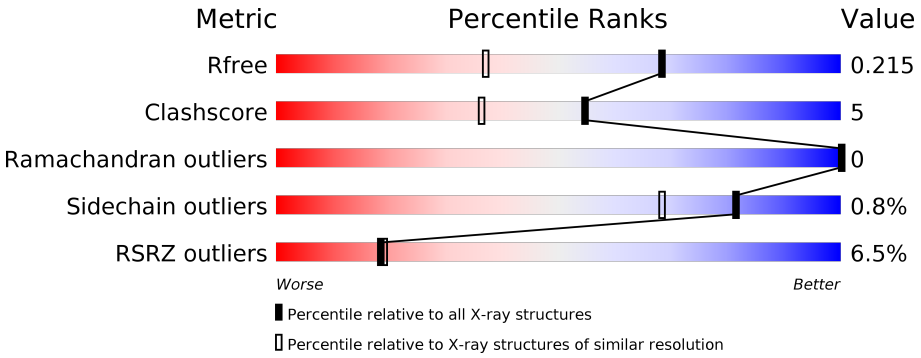
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.13
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.13

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 1.58 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	5534 (1.60-1.56)
Clashscore	141614	5861 (1.60-1.56)
Ramachandran outliers	138981	5708 (1.60-1.56)
Sidechain outliers	138945	5703 (1.60-1.56)
RSRZ outliers	127900	5431 (1.60-1.56)

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 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	A	246	6% 83% 7% 11%
1	B	246	6% 82% 8% 9%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 4026 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 D-lyxose isomerase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	219	Total	C	N	O	S	Se	0	0	0
			1740	1099	306	327	4	4			
1	B	223	Total	C	N	O	S	Se	0	0	0
			1771	1119	311	333	4	4			

There are 38 discrepancies between the modelled and reference sequences:

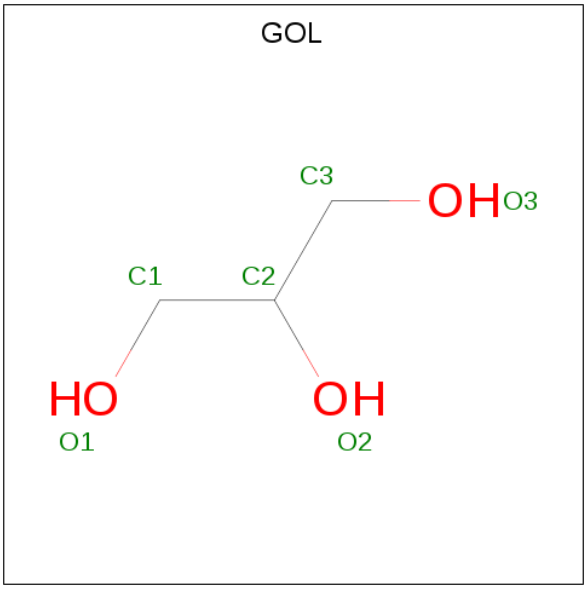
Chain	Residue	Modelled	Actual	Comment	Reference
A	-18	MSE	-	expression tag	UNP Q8X5Q7
A	-17	GLY	-	expression tag	UNP Q8X5Q7
A	-16	SER	-	expression tag	UNP Q8X5Q7
A	-15	SER	-	expression tag	UNP Q8X5Q7
A	-14	HIS	-	expression tag	UNP Q8X5Q7
A	-13	HIS	-	expression tag	UNP Q8X5Q7
A	-12	HIS	-	expression tag	UNP Q8X5Q7
A	-11	HIS	-	expression tag	UNP Q8X5Q7
A	-10	HIS	-	expression tag	UNP Q8X5Q7
A	-9	HIS	-	expression tag	UNP Q8X5Q7
A	-8	SER	-	expression tag	UNP Q8X5Q7
A	-7	SER	-	expression tag	UNP Q8X5Q7
A	-6	GLY	-	expression tag	UNP Q8X5Q7
A	-5	LEU	-	expression tag	UNP Q8X5Q7
A	-4	VAL	-	expression tag	UNP Q8X5Q7
A	-3	PRO	-	expression tag	UNP Q8X5Q7
A	-2	ARG	-	expression tag	UNP Q8X5Q7
A	-1	GLY	-	expression tag	UNP Q8X5Q7
A	0	SER	-	expression tag	UNP Q8X5Q7
B	-18	MSE	-	expression tag	UNP Q8X5Q7
B	-17	GLY	-	expression tag	UNP Q8X5Q7
B	-16	SER	-	expression tag	UNP Q8X5Q7
B	-15	SER	-	expression tag	UNP Q8X5Q7
B	-14	HIS	-	expression tag	UNP Q8X5Q7
B	-13	HIS	-	expression tag	UNP Q8X5Q7

Continued on next page...

Continued from previous page...

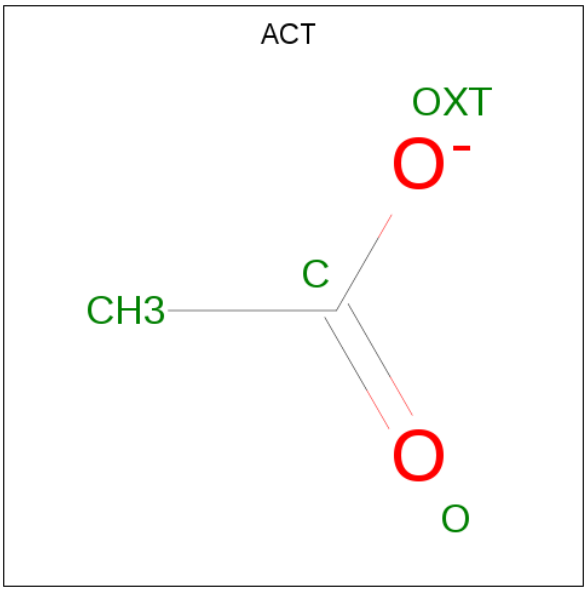
Chain	Residue	Modelled	Actual	Comment	Reference
B	-12	HIS	-	expression tag	UNP Q8X5Q7
B	-11	HIS	-	expression tag	UNP Q8X5Q7
B	-10	HIS	-	expression tag	UNP Q8X5Q7
B	-9	HIS	-	expression tag	UNP Q8X5Q7
B	-8	SER	-	expression tag	UNP Q8X5Q7
B	-7	SER	-	expression tag	UNP Q8X5Q7
B	-6	GLY	-	expression tag	UNP Q8X5Q7
B	-5	LEU	-	expression tag	UNP Q8X5Q7
B	-4	VAL	-	expression tag	UNP Q8X5Q7
B	-3	PRO	-	expression tag	UNP Q8X5Q7
B	-2	ARG	-	expression tag	UNP Q8X5Q7
B	-1	GLY	-	expression tag	UNP Q8X5Q7
B	0	SER	-	expression tag	UNP Q8X5Q7

- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C O 6 3 3	0	0
2	B	1	Total C O 6 3 3	0	0

- Molecule 3 is ACETATE ION (three-letter code: ACT) (formula: $C_2H_3O_2$).

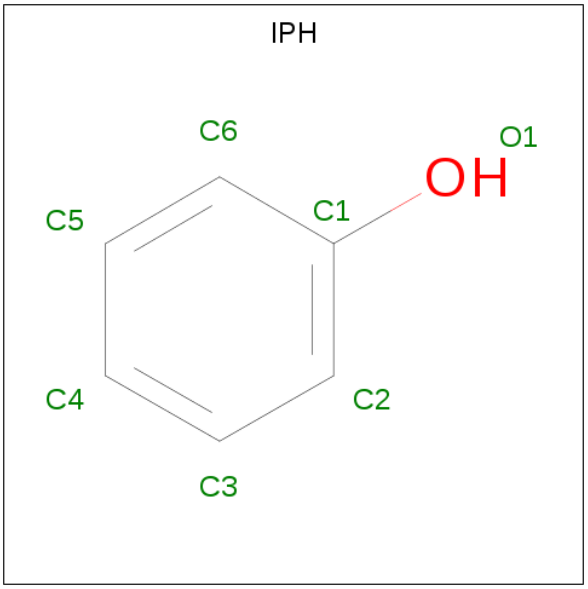


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 4 2 2	0	0

- Molecule 4 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	1	Total Mn 1 1	0	0
4	A	1	Total Mn 1 1	0	0

- Molecule 5 is PHENOL (three-letter code: IPH) (formula: C_6H_6O).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	B	1	Total C O 7 6 1	0	0

- Molecule 6 is water.

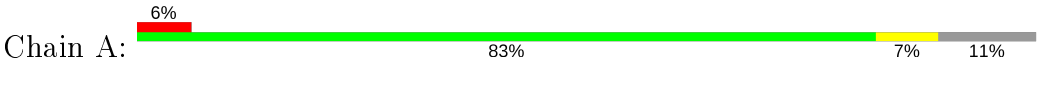
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	245	Total O 245 245	0	0
6	B	245	Total O 245 245	0	0

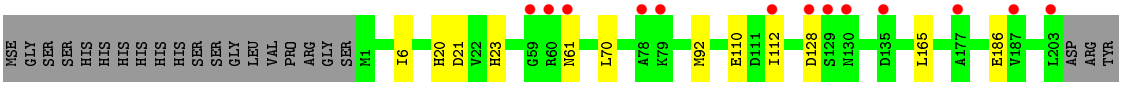
3 Residue-property plots (i)

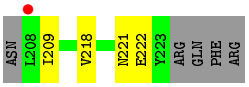
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: D-lyxose isomerase

Chain A:

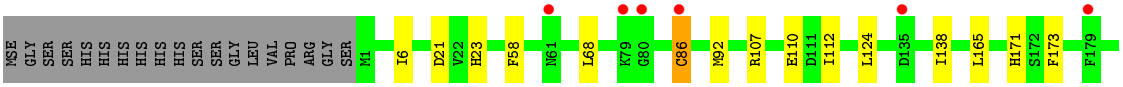


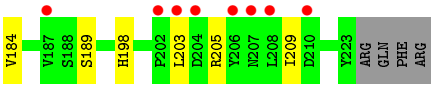




- Molecule 1: D-lyxose isomerase

Chain B:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	53.35Å 75.31Å 61.78Å 90.00° 106.29° 90.00°	Depositor
Resolution (Å)	29.50 – 1.58 29.65 – 1.58	Depositor EDS
% Data completeness (in resolution range)	96.9 (29.50-1.58) 96.7 (29.65-1.58)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.06	Depositor
$< I/\sigma(I) >$ [1]	6.19 (at 1.57Å)	Xtriage
Refinement program	REFMAC 5.5.0072	Depositor
R, R_{free}	0.191 , 0.213 0.190 , 0.215	Depositor DCC
R_{free} test set	3175 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å^2)	18.1	Xtriage
Anisotropy	0.189	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.41 , 54.0	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	0.96	EDS
Total number of atoms	4026	wwPDB-VP
Average B, all atoms (Å^2)	22.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.96% 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: GOL, IPH, MN, 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	0.35	0/1784	0.52	0/2419
1	B	0.40	1/1817 (0.1%)	0.55	0/2466
All	All	0.37	1/3601 (0.0%)	0.54	0/4885

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	86	CYS	CB-SG	6.97	1.94	1.82

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	1740	0	1643	15	0
1	B	1771	0	1665	21	0
2	A	6	0	8	0	0
2	B	6	0	8	0	0
3	A	4	0	3	1	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
5	B	7	0	6	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	A	245	0	0	2	0
6	B	245	0	0	4	0
All	All	4026	0	3333	36	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 (36) 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:B:68:LEU:HD11	1:B:92:MSE:HE3	1.34	1.05
1:B:6:ILE:HD11	1:B:209:ILE:HD12	1.74	0.67
1:B:107:ARG:NH1	6:B:248:HOH:O	2.29	0.65
1:A:21:ASP:OD1	1:A:23:HIS:HE1	1.82	0.61
1:B:112:ILE:HD13	1:B:165:LEU:CD1	2.34	0.58
1:B:110:GLU:HG3	1:B:112:ILE:HD11	1.87	0.57
1:A:221:ASN:HB2	6:A:273:HOH:O	2.07	0.55
1:B:107:ARG:HH12	1:B:189:SER:HB2	1.70	0.55
1:B:112:ILE:HD13	1:B:165:LEU:HD12	1.90	0.54
1:A:112:ILE:HD13	1:A:165:LEU:CD1	2.37	0.54
1:A:20:HIS:HD2	6:B:359:HOH:O	1.90	0.53
1:B:92:MSE:HE2	1:B:184:VAL:HB	1.93	0.51
1:A:70:LEU:HD23	1:A:92:MSE:HE2	1.92	0.50
1:B:6:ILE:CD1	1:B:209:ILE:HD12	2.42	0.49
1:B:198:HIS:HD2	6:B:335:HOH:O	1.95	0.49
1:B:110:GLU:HG3	1:B:112:ILE:CD1	2.42	0.49
1:B:112:ILE:CD1	1:B:165:LEU:HD12	2.41	0.49
1:B:58:PHE:HB3	1:B:203:LEU:HD13	1.95	0.48
1:A:112:ILE:N	1:A:112:ILE:HD12	2.28	0.48
1:A:92:MSE:CE	1:A:186:GLU:OE1	2.62	0.47
1:A:23:HIS:HD2	3:A:229:ACT:OXT	1.96	0.47
1:A:112:ILE:CD1	1:A:165:LEU:HD12	2.45	0.46
1:A:6:ILE:HD11	1:A:209:ILE:HD12	1.96	0.46
1:A:92:MSE:HE1	1:A:186:GLU:OE1	2.16	0.46
1:A:112:ILE:HD13	1:A:165:LEU:HD12	1.97	0.46
1:B:112:ILE:HD12	1:B:112:ILE:N	2.31	0.45
1:A:20:HIS:HE1	6:A:381:HOH:O	1.99	0.45
1:A:110:GLU:HG3	1:A:112:ILE:HD11	1.99	0.45
1:B:205:ARG:HE	5:B:228:IPH:C1	2.30	0.45
1:B:92:MSE:HE1	1:B:173:PHE:CE1	2.53	0.44
1:B:21:ASP:OD1	1:B:23:HIS:HE1	2.01	0.43

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:124:LEU:HB2	1:B:138:ILE:HD13	2.02	0.42
1:B:110:GLU:CG	1:B:112:ILE:HD11	2.50	0.41
1:A:218:VAL:HG13	1:A:222:GLU:HG3	2.02	0.41
1:B:198:HIS:HE1	6:B:422:HOH:O	2.03	0.41
1:B:165:LEU:HD13	1:B:171:HIS:CG	2.56	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	215/246 (87%)	214 (100%)	1 (0%)	0	100	100
1	B	221/246 (90%)	218 (99%)	3 (1%)	0	100	100
All	All	436/492 (89%)	432 (99%)	4 (1%)	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	186/205 (91%)	184 (99%)	2 (1%)	73	55
1	B	189/205 (92%)	188 (100%)	1 (0%)	88	80
All	All	375/410 (92%)	372 (99%)	3 (1%)	81	68

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	61	ASN
1	A	128	ASP
1	B	86	CYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (11) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	20	HIS
1	A	23	HIS
1	A	61	ASN
1	A	147	GLN
1	A	191	ASN
1	A	221	ASN
1	B	12	HIS
1	B	20	HIS
1	B	23	HIS
1	B	191	ASN
1	B	198	HIS

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 6 ligands modelled in this entry, 2 are monoatomic - leaving 4 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	$\#\|Z\| > 2$
5	IPH	B	228	-	7,7,7	0.52	0	8,8,8	0.28	0
2	GOL	B	229	-	5,5,5	0.37	0	5,5,5	0.24	0
3	ACT	A	229	-	1,3,3	1.32	0	0,3,3	0.00	-
2	GOL	A	228	-	5,5,5	0.37	0	5,5,5	0.16	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	IPH	B	228	-	-	-	0/1/1/1
2	GOL	B	229	-	-	3/4/4/4	-
2	GOL	A	228	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	229	GOL	C1-C2-C3-O3
2	A	228	GOL	C1-C2-C3-O3
2	B	229	GOL	O2-C2-C3-O3
2	A	228	GOL	O2-C2-C3-O3
2	B	229	GOL	O1-C1-C2-C3

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	B	228	IPH	1	0
3	A	229	ACT	1	0

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

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	A	215/246 (87%)	0.37	14 (6%)	18 19	12, 20, 32, 37	0
1	B	219/246 (89%)	0.34	14 (6%)	19 19	12, 19, 30, 37	0
All	All	434/492 (88%)	0.35	28 (6%)	18 19	12, 19, 31, 37	0

All (28) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	208	LEU	7.1
1	B	79	LYS	5.6
1	A	203	LEU	4.1
1	B	206	TYR	3.9
1	B	208	LEU	3.7
1	A	61	ASN	3.6
1	B	203	LEU	3.4
1	A	79	LYS	3.4
1	B	202	PRO	3.4
1	A	129	SER	3.0
1	B	187	VAL	2.9
1	B	135	ASP	2.9
1	A	59	GLY	2.8
1	B	207	ASN	2.8
1	B	179	PHE	2.8
1	A	112	ILE	2.7
1	B	80	GLY	2.7
1	B	61	ASN	2.7
1	A	130	ASN	2.6
1	A	128	ASP	2.5
1	B	204	ASP	2.5
1	A	60	ARG	2.5
1	B	86	CYS	2.5
1	A	177	ALA	2.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	187	VAL	2.2
1	B	210	ASP	2.1
1	A	135	ASP	2.0
1	A	78	ALA	2.0

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, 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
2	GOL	A	228	6/6	0.61	0.24	47,47,47,48	0
2	GOL	B	229	6/6	0.78	0.17	32,34,35,36	0
3	ACT	A	229	4/4	0.83	0.19	32,32,32,32	0
5	IPH	B	228	7/7	0.88	0.15	29,30,30,31	0
4	MN	B	1002	1/1	0.99	0.09	16,16,16,16	0
4	MN	A	1001	1/1	1.00	0.07	20,20,20,20	0

6.5 Other polymers (i)

There are no such residues in this entry.