



Full wwPDB X-ray Structure Validation Report ⓘ

Apr 10, 2017 – 05:45 PM EDT

PDB ID : 5UOQ
Title : CRYSTAL STRUCTURE OF THE PROTOTYPE FOAMY VIRUS INTA-SOME WITH A 2- PYRIDINONE AMINAL INHIBITOR (COMPOUND 31)
Authors : Klein, D.J.
Deposited on : 2017-02-01
Resolution : 2.61 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

<http://wwpdb.org/validation/2016/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 : 1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20029077
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20029077

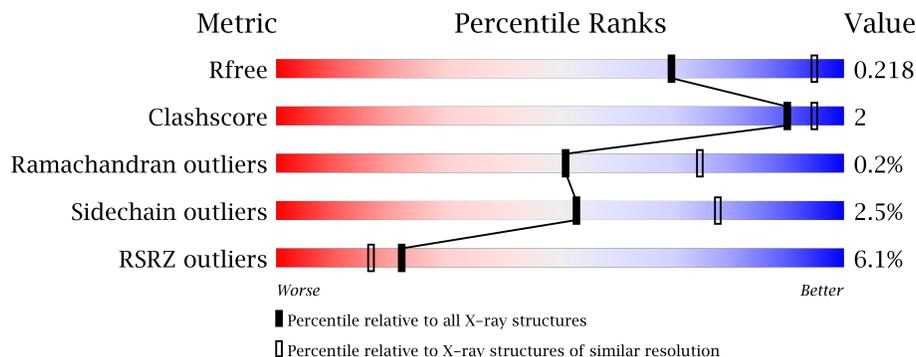
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.61 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}	100719	2983 (2.64-2.60)
Clashscore	112137	3351 (2.64-2.60)
Ramachandran outliers	110173	3298 (2.64-2.60)
Sidechain outliers	110143	3298 (2.64-2.60)
RSRZ outliers	101464	2992 (2.64-2.60)

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. 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	395	 4% 85% 8% 7%
1	B	395	 6% 43% 53%
2	C	19	 84% 16%
3	D	17	 88% 12%

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 crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
5	GOL	A	408	-	-	-	X
5	GOL	B	402	-	-	-	X
7	SO4	A	406	-	-	-	X

2 Entry composition [i](#)

There are 9 unique types of molecules in this entry. The entry contains 5329 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 INTEGRASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	368	2908	1864	510	530	4	0	0	0
1	B	184	1424	925	231	267	1	0	0	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	expression tag	UNP P14350
A	-1	PRO	-	expression tag	UNP P14350
A	0	GLY	-	expression tag	UNP P14350
A	217	SER	GLY	engineered mutation	UNP P14350
A	218	GLY	SER	engineered mutation	UNP P14350
B	-2	GLY	-	expression tag	UNP P14350
B	-1	PRO	-	expression tag	UNP P14350
B	0	GLY	-	expression tag	UNP P14350
B	217	SER	GLY	engineered mutation	UNP P14350
B	218	GLY	SER	engineered mutation	UNP P14350

- Molecule 2 is a DNA chain called NUCLEOTIDE PREPROCESSED PFV DONOR DNA (NON-TRANSFERRED STRAND).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	C	19	387	187	68	114	18	0	0	0

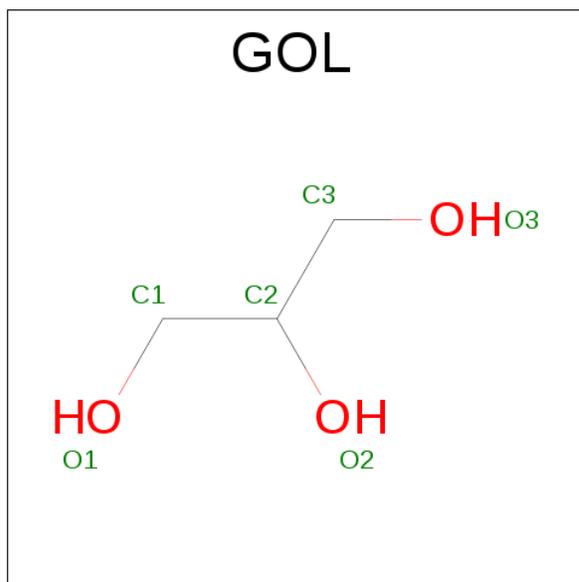
- Molecule 3 is a DNA chain called NUCLEOTIDE PREPROCESSED PFV DONOR DNA (TRANSFERRED STRAND).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	D	17	345	166	65	98	16	0	0	0

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total Zn 1 1	0	0

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).

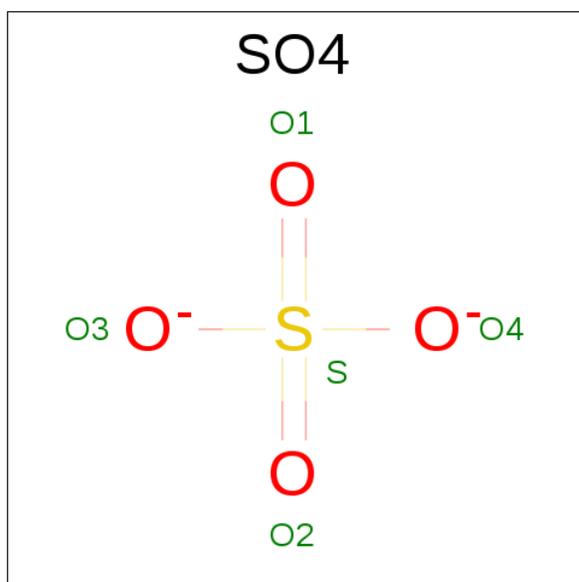


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total C O 6 3 3	0	0
5	A	1	Total C O 6 3 3	0	0
5	A	1	Total C O 6 3 3	0	0
5	B	1	Total C O 6 3 3	0	0
5	C	1	Total C O 6 3 3	0	0

- Molecule 6 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

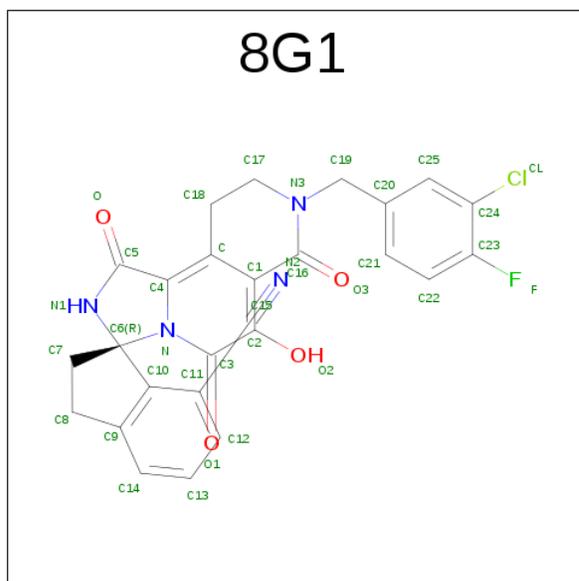
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	B	1	Total Mg 1 1	0	0
6	A	2	Total Mg 2 2	0	0

- Molecule 7 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



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

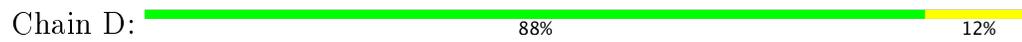
- Molecule 8 is (3R)-8-[(3-chloro-4-fluorophenyl)methyl]-6-hydroxy-1,5,7-trioxo-1,2',3',5,7,8,9,10-octahydro-2H-spiro[imidazo[5,1-a][2,6]naphthyridine-3,1'-indene]-7'-carbonitrile (three-letter code: 8G1) (formula: C₂₆H₁₈ClFN₄O₄).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	
			Total	C	Cl	F	N			O
8	A	1	36	26	1	1	4	4	0	0

- Molecule 9 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	A	114	Total 114	O 114	0	0
9	B	31	Total 31	O 31	0	0
9	C	18	Total 18	O 18	0	0
9	D	17	Total 17	O 17	0	0



4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α , β , γ	159.51Å 159.51Å 123.82Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	37.96 – 2.61 37.96 – 2.61	Depositor EDS
% Data completeness (in resolution range)	100.0 (37.96-2.61) 100.0 (37.96-2.61)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.42 (at 2.61Å)	Xtrriage
Refinement program	BUSTER 2.11.6	Depositor
R, R_{free}	0.199 , 0.217 0.196 , 0.218	Depositor DCC
R_{free} test set	2460 reflections (5.28%)	DCC
Wilson B-factor (Å ²)	67.8	Xtrriage
Anisotropy	0.257	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 48.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	5329	wwPDB-VP
Average B, all atoms (Å ²)	70.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.51% 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: GOL, MG, 8G1, ZN, SO4

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	0.48	0/2987	0.69	0/4081
1	B	0.51	0/1465	0.65	0/2008
2	C	1.20	0/433	0.99	0/667
3	D	1.22	0/387	1.00	0/595
All	All	0.66	0/5272	0.74	0/7351

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	A	2908	0	2919	15	0
1	B	1424	0	1385	5	0
2	C	387	0	218	2	0
3	D	345	0	193	1	0
4	A	1	0	0	0	0
5	A	18	0	24	1	0
5	B	6	0	8	0	0
5	C	6	0	8	0	0
6	A	2	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	B	1	0	0	0	0
7	A	10	0	0	0	0
7	B	5	0	0	0	0
8	A	36	0	0	0	0
9	A	114	0	0	0	0
9	B	31	0	0	0	0
9	C	18	0	0	0	0
9	D	17	0	0	0	0
All	All	5329	0	4755	21	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 2.

All (21) 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)
2:C:1:DA:H4'	2:C:2:DT:H5'	1.46	0.95
3:D:1:DT:H2''	3:D:2:DG:OP2	1.96	0.64
1:B:134:PRO:HG2	1:B:239:PRO:HB2	1.90	0.52
1:A:73:LEU:HD13	1:A:86:ARG:HG3	1.91	0.52
1:A:69:ARG:HD2	1:A:70:GLU:OE2	2.10	0.51
1:A:97:GLN:HE21	1:A:339:LYS:HG2	1.77	0.50
1:B:283:THR:C	1:B:285:ASP:H	2.15	0.48
1:B:127:ILE:HA	1:B:144:VAL:O	2.13	0.47
1:A:219:LYS:HE3	2:C:6:DC:OP1	2.15	0.47
1:A:161:PRO:O	1:A:189:ALA:HB2	2.16	0.46
1:B:163:THR:HG21	1:B:194:THR:HB	2.00	0.43
1:A:113:LEU:HG	1:A:350:ARG:HG3	2.00	0.43
1:A:149:MET:HE3	1:A:150:THR:HG23	2.00	0.43
1:A:145:VAL:HG11	1:A:173:LEU:HD21	2.00	0.43
1:A:292:LEU:HD22	1:B:271:GLY:HA3	2.01	0.43
1:A:12:LEU:HB3	1:A:37:VAL:HG21	2.02	0.42
1:A:127:ILE:HG22	1:A:145:VAL:HG22	2.02	0.42
1:A:156:TYR:OH	1:A:173:LEU:HB2	2.20	0.41
1:A:108:ALA:O	1:A:314:SER:HA	2.21	0.41
1:A:131:GLY:O	5:A:408:GOL:H11	2.21	0.41
1:A:122:PHE:O	1:A:179:PRO:HA	2.21	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	366/395 (93%)	356 (97%)	10 (3%)	0	100	100
1	B	182/395 (46%)	171 (94%)	10 (6%)	1 (0%)	32	56
All	All	548/790 (69%)	527 (96%)	20 (4%)	1 (0%)	51	74

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	236	VAL

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	324/354 (92%)	317 (98%)	7 (2%)	57	80
1	B	155/354 (44%)	150 (97%)	5 (3%)	44	71
All	All	479/708 (68%)	467 (98%)	12 (2%)	53	78

All (12) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	13	ASP
1	A	19	HIS
1	A	32	LEU
1	A	73	LEU
1	A	137	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	217	SER
1	A	312	SER
1	B	149	MET
1	B	164	SER
1	B	274	SER
1	B	288	ARG
1	B	289	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	97	GLN
1	A	281	GLN
1	B	266	HIS

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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 13 ligands modelled in this entry, 4 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
5	GOL	A	402	-	5,5,5	0.15	0	5,5,5	0.26	0
5	GOL	A	403	-	5,5,5	0.20	0	5,5,5	0.49	0
7	SO4	A	406	-	4,4,4	0.28	0	6,6,6	0.15	0
7	SO4	A	407	-	4,4,4	0.11	0	6,6,6	0.21	0
5	GOL	A	408	-	5,5,5	0.20	0	5,5,5	0.38	0
8	8G1	A	409	6	35,41,41	0.93	3 (8%)	42,64,64	1.38	2 (4%)
7	SO4	B	401	-	4,4,4	0.07	0	6,6,6	0.34	0
5	GOL	B	402	-	5,5,5	0.15	0	5,5,5	0.25	0
5	GOL	C	101	-	5,5,5	0.13	0	5,5,5	0.41	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	GOL	A	402	-	-	0/4/4/4	0/0/0/0
5	GOL	A	403	-	-	0/4/4/4	0/0/0/0
7	SO4	A	406	-	-	0/0/0/0	0/0/0/0
7	SO4	A	407	-	-	0/0/0/0	0/0/0/0
5	GOL	A	408	-	-	0/4/4/4	0/0/0/0
8	8G1	A	409	6	-	0/6/46/46	0/5/6/6
7	SO4	B	401	-	-	0/0/0/0	0/0/0/0
5	GOL	B	402	-	-	0/4/4/4	0/0/0/0
5	GOL	C	101	-	-	0/4/4/4	0/0/0/0

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
8	A	409	8G1	C4-C5	-2.04	1.48	1.51
8	A	409	8G1	C1-C2	2.75	1.43	1.40
8	A	409	8G1	C1-C	3.63	1.47	1.40

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
8	A	409	8G1	C3-N-C4	4.28	124.88	118.68
8	A	409	8G1	O-C5-C4	6.02	128.11	119.83

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	408	GOL	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, 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	368/395 (93%)	0.17	14 (3%) 41 34	51, 64, 99, 139	0
1	B	184/395 (46%)	0.23	22 (11%) 5 3	55, 70, 132, 149	0
2	C	19/19 (100%)	-0.15	0 100 100	53, 67, 90, 110	0
3	D	17/17 (100%)	-0.23	0 100 100	57, 62, 92, 108	0
All	All	588/826 (71%)	0.16	36 (6%) 22 17	51, 66, 115, 149	0

All (36) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	280	ASN	5.0
1	A	8	LEU	4.3
1	A	16	LEU	3.6
1	B	279	ALA	3.3
1	B	260	VAL	3.3
1	B	213	HIS	3.2
1	B	296	GLN	3.2
1	B	237	GLY	3.1
1	A	17	GLN	2.9
1	B	297	GLU	2.9
1	B	294	LEU	2.7
1	A	15	LEU	2.7
1	A	358	LEU	2.6
1	A	357	HIS	2.6
1	A	34	ASP	2.6
1	A	153	THR	2.5
1	B	292	LEU	2.5
1	A	146	VAL	2.5
1	B	257	TYR	2.4
1	B	293	SER	2.4
1	A	359	GLY	2.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	289	GLU	2.4
1	B	295	LEU	2.3
1	B	116	ASP	2.3
1	B	281	GLN	2.2
1	B	263	TYR	2.2
1	B	261	LEU	2.2
1	B	262	LYS	2.2
1	A	20	TYR	2.1
1	A	21	ILE	2.1
1	A	253	LEU	2.1
1	B	298	ILE	2.1
1	B	278	PHE	2.1
1	A	220	VAL	2.0
1	B	215	GLN	2.0
1	B	259	PRO	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 carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å ²)	Q<0.9
5	GOL	A	408	6/6	0.70	0.39	3.83	81,86,88,91	0
5	GOL	B	402	6/6	0.91	0.44	3.37	81,85,86,91	0
7	SO4	A	406	5/5	0.73	0.29	2.57	152,152,153,154	0
5	GOL	C	101	6/6	0.97	0.21	1.05	69,73,78,83	0
4	ZN	A	401	1/1	1.00	0.17	0.80	61,61,61,61	0
8	8G1	A	409	36/36	0.94	0.18	0.02	42,76,89,90	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(\AA^2)	Q<0.9
6	MG	A	405	1/1	0.99	0.10	-2.13	59,59,59,59	0
6	MG	B	403	1/1	0.69	0.43	-	100,100,100,100	0
5	GOL	A	402	6/6	0.80	0.33	-	93,95,96,96	0
7	SO4	A	407	5/5	0.97	0.13	-	88,91,94,95	0
5	GOL	A	403	6/6	0.81	0.22	-	83,94,99,100	0
7	SO4	B	401	5/5	0.91	0.23	-	107,109,110,111	0
6	MG	A	404	1/1	0.98	0.08	-	60,60,60,60	0

6.5 Other polymers [i](#)

There are no such residues in this entry.