Full wwPDB X-ray Structure Validation Report (i)

Feb 1, 2016 – 09:06 AM GMT

PDB ID	:	3H8R
Title	:	Structure determination of DNA methylation lesions N1-meA and N3-meC in duplex DNA using a cross-linked host-guest system
Authors	:	Lu, L.; Yi, C.; Jian, X.; Zheng, G; He, C.
Deposited on	:	2009-04-29
Resolution	:	1.77 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

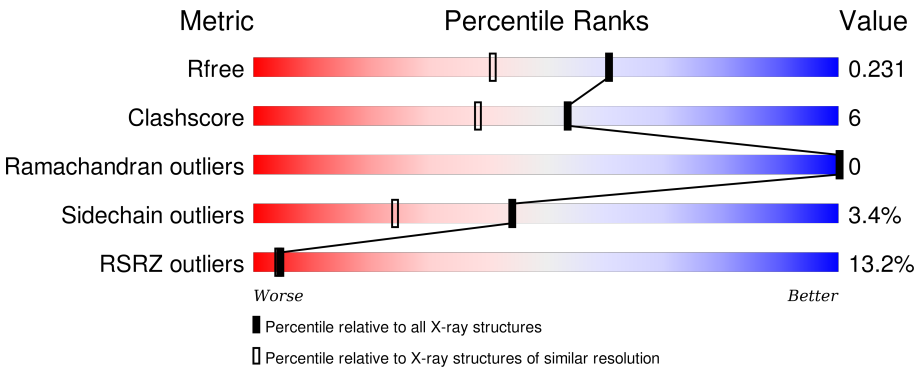
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 1.77 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}	91344	6655 (1.80-1.76)
Clashscore	102246	7658 (1.80-1.76)
Ramachandran outliers	100387	7570 (1.80-1.76)
Sidechain outliers	100360	7569 (1.80-1.76)
RSRZ outliers	91569	6671 (1.80-1.76)

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 <=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	209	14% 82% 15% • •
2	B	13	77% 23%
3	C	13	38% 46% 15%

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	GOL	A	301	-	-	-	X

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 2411 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 Alpha-ketoglutarate-dependent dioxygenase alkB homolog 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	204	1651	1054	305	289	3	0	2	0

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	53	GLY	-	EXPRESSION TAG	UNP Q6NS38
A	54	SER	-	EXPRESSION TAG	UNP Q6NS38
A	55	HIS	-	EXPRESSION TAG	UNP Q6NS38
A	67	SER	CYS	ENGINEERED	UNP Q6NS38
A	165	SER	CYS	ENGINEERED	UNP Q6NS38
A	175	CYS	GLU	ENGINEERED	UNP Q6NS38
A	192	SER	CYS	ENGINEERED	UNP Q6NS38

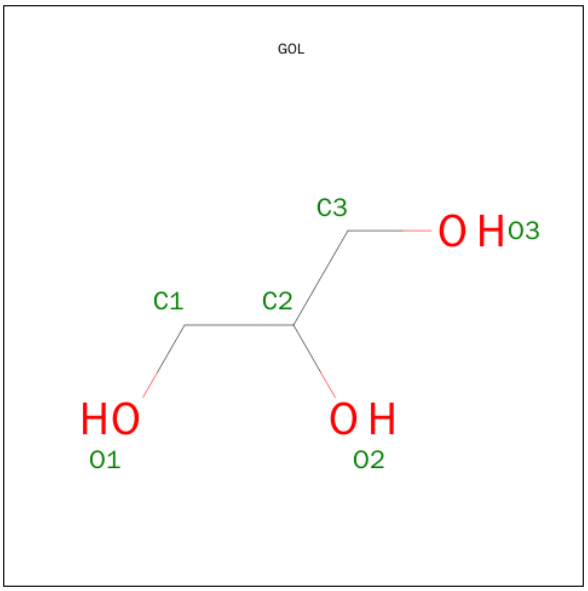
- Molecule 2 is a DNA chain called 5'-D(*CP*TP*GP*TP*AP*TP*(2YR)P*AP*TP*(6MA)P*GP*CP*G)-3'.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	P	S			
2	B	13	267	130	47	77	12	1	0	0	0

- Molecule 3 is a DNA chain called 5'-D(*TP*CP*GP*CP*TP*AP*TP*AP*AP*TP*AP*CP*A)-3'.

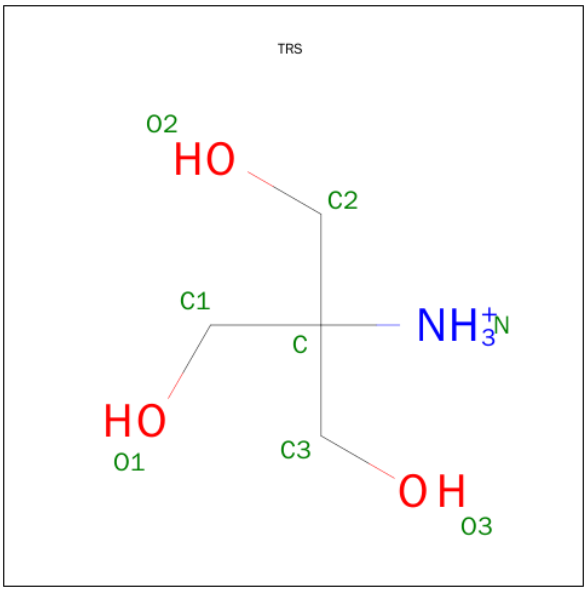
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	C	13	261	127	47	75	12	0	0	0

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



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

- Molecule 5 is 2-AMINO-2-HYDROXYMETHYL-PROPANE-1,3-DIOL (three-letter code: TRS) (formula: $C_4H_{12}NO_3$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
5	C	1	8	4	1	3	0	0

- Molecule 6 is water.

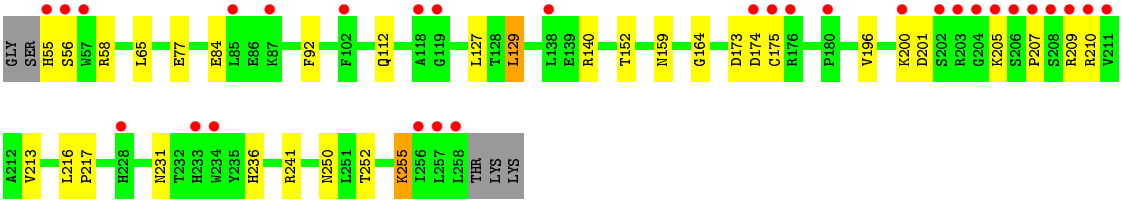
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
6	A	142	142	142	0	0
			Total	O		
6	B	36	36	36	0	0
			Total	O		
6	C	40	40	40	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

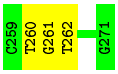
- Molecule 1: Alpha-ketoglutarate-dependent dioxygenase alkB homolog 2

Chain A:



- Molecule 2: 5'-D(*CP*TP*GP*TP*AP*TP*(2YR)P*AP*TP*(6MA)P*GP*CP*G)-3'

Chain B:



- Molecule 3: 5'-D(*TP*CP*GP*CP*TP*AP*TP*AP*AP*TP*AP*CP*A)-3'

Chain C:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants a, b, c, α, β, γ	78.08Å 78.08Å 228.71Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	15.00 – 1.77 14.97 – 1.77	Depositor EDS
% Data completeness (in resolution range)	99.9 (15.00-1.77) 98.1 (14.97-1.77)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.00 (at 1.77Å)	Xtriage
Refinement program	REFMAC 5.4.0073	Depositor
R, R_{free}	0.210 , 0.236 0.205 , 0.231	Depositor DCC
R_{free} test set	1993 reflections (5.19%)	DCC
Wilson B-factor (Å^2)	31.8	Xtriage
Anisotropy	0.208	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 46.3	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Outliers	0 of 40364 reflections	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	2411	wwPDB-VP
Average B, all atoms (Å^2)	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.81% 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, 0.375 respectively for untwinned datasets, and 0.333, 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, TRS, 2YR, 6MA

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	0/1700	0.62	1/2303 (0.0%)
2	B	1.01	1/247 (0.4%)	1.63	2/375 (0.5%)
3	C	0.91	0/292	1.73	12/448 (2.7%)
All	All	0.63	1/2239 (0.0%)	1.02	15/3126 (0.5%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	260	DT	O3'-P	-6.56	1.53	1.61

All (15) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	278	DT	C1'-O4'-C4'	-7.75	102.35	110.10
2	B	262	DT	O4'-C4'-C3'	-7.13	101.65	104.50
3	C	278	DT	O4'-C4'-C3'	6.99	110.19	106.00
3	C	276	DT	C4-C5-C7	6.57	122.94	119.00
3	C	277	DA	O4'-C1'-N9	-5.99	103.81	108.00
3	C	284	DA	O4'-C1'-N9	-5.89	103.88	108.00
3	C	276	DT	O4'-C1'-N1	-5.84	103.91	108.00
3	C	278	DT	O4'-C1'-N1	5.80	112.06	108.00
3	C	284	DA	O4'-C1'-C2'	-5.71	101.33	105.90
3	C	273	DC	O4'-C1'-N1	-5.70	104.01	108.00
3	C	272	DT	O4'-C1'-N1	-5.63	104.06	108.00
1	A	175	CYS	CB-CA-C	-5.39	99.61	110.40
3	C	276	DT	C6-C5-C7	-5.22	119.77	122.90
2	B	261	DG	P-O5'-C5'	-5.09	112.75	120.90
3	C	281	DT	N3-C4-O4	5.05	122.93	119.90

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	1651	0	1645	24	0
2	B	267	0	155	0	0
3	C	261	0	149	5	0
4	A	6	0	8	3	0
5	C	8	0	12	1	0
6	A	142	0	0	0	0
6	B	36	0	0	0	0
6	C	40	0	0	0	0
All	All	2411	0	1969	27	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 6.

All (27) 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:207:PRO:HD2	1:A:210:ARG:NH2	1.95	0.82
1:A:205:LYS:HE3	3:C:277:DA:OP1	1.90	0.72
1:A:216[B]:LEU:HD12	1:A:217:PRO:HD2	1.75	0.69
1:A:250:ASN:HD21	1:A:252:THR:HG23	1.61	0.65
3:C:277:DA:N3	5:C:301:TRS:O2	2.26	0.64
1:A:207:PRO:HD2	1:A:210:ARG:HH22	1.63	0.63
1:A:92:PHE:CE2	1:A:129:LEU:HG	2.38	0.58
1:A:200:LYS:HG3	1:A:201:ASP:OD1	2.04	0.57
1:A:205:LYS:HB2	3:C:276:DT:H3'	1.90	0.53
1:A:140:ARG:HB3	4:A:301:GOL:O2	2.09	0.53
1:A:216[B]:LEU:HD12	1:A:217:PRO:CD	2.39	0.53
1:A:127:LEU:HG	1:A:129:LEU:HD13	1.91	0.52
3:C:275:DC:H2"	3:C:276:DT:H5'	1.93	0.50
1:A:173:ASP:H	1:A:231:ASN:ND2	2.11	0.49
1:A:196:VAL:HG13	1:A:213:VAL:CG1	2.44	0.47
3:C:275:DC:H2'	3:C:276:DT:C6	2.50	0.46

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:112:GLN:HG2	1:A:159:ASN:OD1	2.14	0.46
1:A:207:PRO:HG2	1:A:210:ARG:NH1	2.32	0.45
1:A:140:ARG:NE	4:A:301:GOL:O2	2.47	0.44
1:A:174:ASP:OD1	1:A:174:ASP:O	2.37	0.43
1:A:84:GLU:OE1	1:A:140:ARG:NH2	2.47	0.42
1:A:77:GLU:OE2	4:A:301:GOL:H11	2.19	0.42
1:A:210:ARG:HG2	1:A:210:ARG:HH11	1.84	0.42
1:A:164:GLY:HA3	1:A:241:ARG:O	2.19	0.42
1:A:152:THR:OG1	1:A:255:LYS:HE3	2.20	0.41
1:A:58:ARG:NH1	1:A:216[B]:LEU:HD13	2.36	0.40
1:A:196:VAL:O	1:A:236:HIS:HA	2.22	0.40

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	204/209 (98%)	200 (98%)	4 (2%)	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	178/180 (99%)	172 (97%)	6 (3%)	44 24

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	55	HIS
1	A	56	SER
1	A	65	LEU
1	A	129	LEU
1	A	209	ARG
1	A	255	LYS

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	112	GLN
1	A	231	ASN
1	A	250	ASN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

1 non-standard protein/DNA/RNA residue is 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	$\#\|Z\| > 2$
2	6MA	B	268	3,2	16,24,25	0.93	0	18,34,37	3.10	4 (22%)

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
2	6MA	B	268	3,2	-	0/5/23/24	0/3/3/3

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	268	6MA	N3-C2-N1	-10.31	121.00	128.89
2	B	268	6MA	C4-C5-N7	-2.98	106.74	109.48
2	B	268	6MA	C1'-N9-C4	-2.14	123.53	127.16
2	B	268	6MA	C2-N1-C6	6.55	121.20	116.48

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

2 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	$\#\|Z\| > 2$
4	GOL	A	301	-	5,5,5	0.55	0	5,5,5	0.72	0
5	TRS	C	301	-	7,7,7	0.71	0	9,9,9	2.22	5 (55%)

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	A	301	-	-	0/4/4/4	0/0/0/0
5	TRS	C	301	-	-	0/9/9/9	0/0/0/0

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	C	301	TRS	C2-C-C1	-3.44	103.34	110.78
5	C	301	TRS	C3-C-C1	-2.78	104.77	110.78
5	C	301	TRS	O1-C1-C	-2.52	106.08	111.18
5	C	301	TRS	C3-C-N	2.26	112.19	108.09
5	C	301	TRS	C3-C-C2	3.08	117.45	110.78

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	301	GOL	3	0
5	C	301	TRS	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	204/209 (97%)	0.75	30 (14%) 3 3	19, 25, 44, 57	0
2	B	11/13 (84%)	0.26	0 100 100	18, 26, 34, 36	0
3	C	13/13 (100%)	0.05	0 100 100	25, 29, 33, 33	1 (7%)
All	All	228/235 (97%)	0.68	30 (13%) 4 4	18, 26, 43, 57	1 (0%)

All (30) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	204	GLY	9.8
1	A	206	SER	9.1
1	A	258	LEU	6.5
1	A	205	LYS	6.2
1	A	208	SER	5.9
1	A	209	ARG	5.8
1	A	55	HIS	5.6
1	A	207	PRO	5.6
1	A	176	ARG	5.2
1	A	180	PRO	4.2
1	A	119	GLY	4.2
1	A	56	SER	3.9
1	A	257	LEU	3.3
1	A	203	ARG	3.3
1	A	228	HIS	3.3
1	A	211	VAL	3.3
1	A	210	ARG	3.1
1	A	234	TRP	2.9
1	A	200	LYS	2.8
1	A	175	CYS	2.7
1	A	233	HIS	2.7
1	A	118	ALA	2.4
1	A	57	TRP	2.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	174	ASP	2.3
1	A	202	SER	2.3
1	A	138	LEU	2.3
1	A	85	LEU	2.2
1	A	102	PHE	2.1
1	A	256	ILE	2.1
1	A	87	LYS	2.0

6.2 Non-standard residues in protein, DNA, RNA chains (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
2	6MA	B	268	22/23	0.92	0.12	-	18,23,27,28	0

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
4	GOL	A	301	6/6	0.74	0.19	2.16	36,45,48,50	0
5	TRS	C	301	8/8	0.76	0.23	-	34,36,37,39	0

6.5 Other polymers (i)

There are no such residues in this entry.