Full wwPDB X-ray Structure Validation Report (i)

Aug 9, 2020 – 09:03 AM BST

PDB ID	:	3EPZ
Title	:	Structure of the replication foci-targeting sequence of human DNA cytosine methyltransferase DNMT1
Authors	:	Walker, J.R.; Avvakumov, G.V.; Xue, S.; Li, Y.; Bountra, C.; Weigelt, J.; Arrowsmith, C.H.; Edwards, A.M.; Bochkarev, A.; Dhe-Paganon, S.; Structural Genomics Consortium (SGC)
Deposited on	:	2008-09-30
Resolution	:	2.31 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 (i) symbol.

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

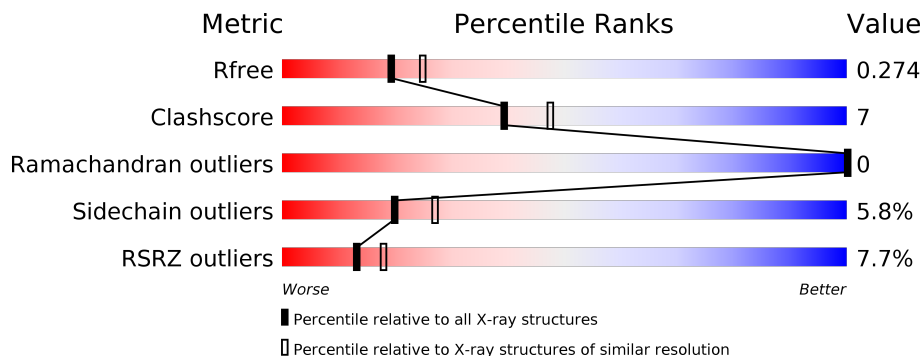|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.13.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.13.1

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 2.31 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	Percentile Ranks	Value
Rfree		0.274
Clashscore		7
Ramachandran outliers		0
Sidechain outliers		5.8%
RSRZ outliers		7.7%

Worse   *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	5974 (2.34-2.30)
Clashscore	141614	6604 (2.34-2.30)
Ramachandran outliers	138981	6523 (2.34-2.30)
Sidechain outliers	138945	6523 (2.34-2.30)
RSRZ outliers	127900	5855 (2.34-2.30)

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 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	268	3% / 76% 8% · 15%
1	B	268	10% / 65% 16% · 17%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 3560 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 DNA (cytosine-5)-methyltransferase 1.

Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
1	A	227	Total	C	N	O	S	Se		0	1	0
			1741	1114	282	335	6	4				
1	B	223	Total	C	N	O	S	Se		0	0	0
			1705	1098	276	321	6	4				

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	333	MSE	-	expression tag	UNP P26358
A	334	HIS	-	expression tag	UNP P26358
A	335	HIS	-	expression tag	UNP P26358
A	336	HIS	-	expression tag	UNP P26358
A	337	HIS	-	expression tag	UNP P26358
A	338	HIS	-	expression tag	UNP P26358
A	339	HIS	-	expression tag	UNP P26358
A	340	SER	-	expression tag	UNP P26358
A	341	SER	-	expression tag	UNP P26358
A	342	GLY	-	expression tag	UNP P26358
A	343	ARG	-	expression tag	UNP P26358
A	344	GLU	-	expression tag	UNP P26358
A	345	ASN	-	expression tag	UNP P26358
A	346	LEU	-	expression tag	UNP P26358
A	347	TYR	-	expression tag	UNP P26358
A	348	PHE	-	expression tag	UNP P26358
A	349	GLN	-	expression tag	UNP P26358
A	350	GLY	-	expression tag	UNP P26358
B	333	MSE	-	expression tag	UNP P26358
B	334	HIS	-	expression tag	UNP P26358
B	335	HIS	-	expression tag	UNP P26358
B	336	HIS	-	expression tag	UNP P26358
B	337	HIS	-	expression tag	UNP P26358
B	338	HIS	-	expression tag	UNP P26358
B	339	HIS	-	expression tag	UNP P26358

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	340	SER	-	expression tag	UNP P26358
B	341	SER	-	expression tag	UNP P26358
B	342	GLY	-	expression tag	UNP P26358
B	343	ARG	-	expression tag	UNP P26358
B	344	GLU	-	expression tag	UNP P26358
B	345	ASN	-	expression tag	UNP P26358
B	346	LEU	-	expression tag	UNP P26358
B	347	TYR	-	expression tag	UNP P26358
B	348	PHE	-	expression tag	UNP P26358
B	349	GLN	-	expression tag	UNP P26358
B	350	GLY	-	expression tag	UNP P26358

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

- Molecule 3 is beta-D-glucopyranose (three-letter code: BGC) (formula: $C_6H_{12}O_6$).
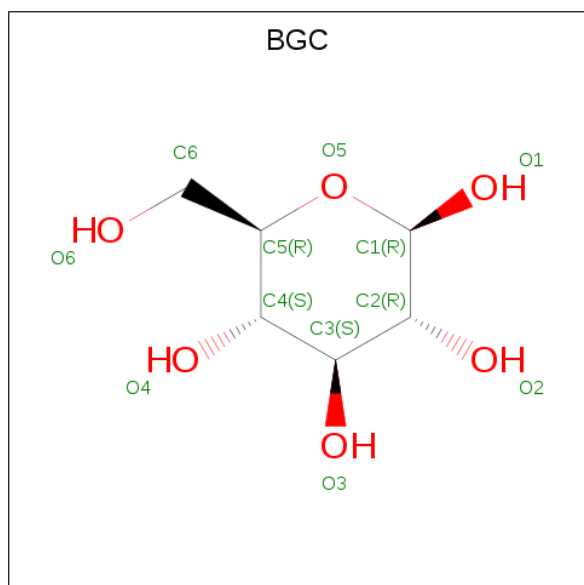
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			12	6	6		

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

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

- Molecule 5 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	1	Total 1	Na 1	0	0

- Molecule 6 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	66	Total 67	O 67	0	1
6	B	25	Total 26	O 26	0	1

3 Residue-property plots ⓘ

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: DNA (cytosine-5)-methyltransferase 1
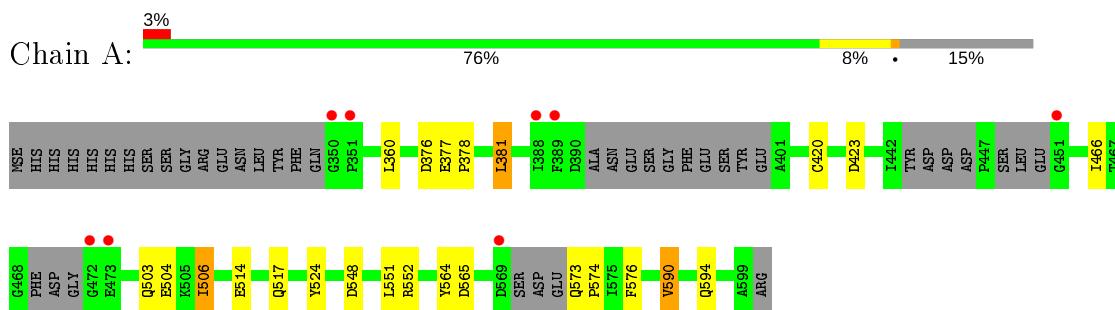
• Molecule 1: DNA (cytosine-5)-methyltransferase 1
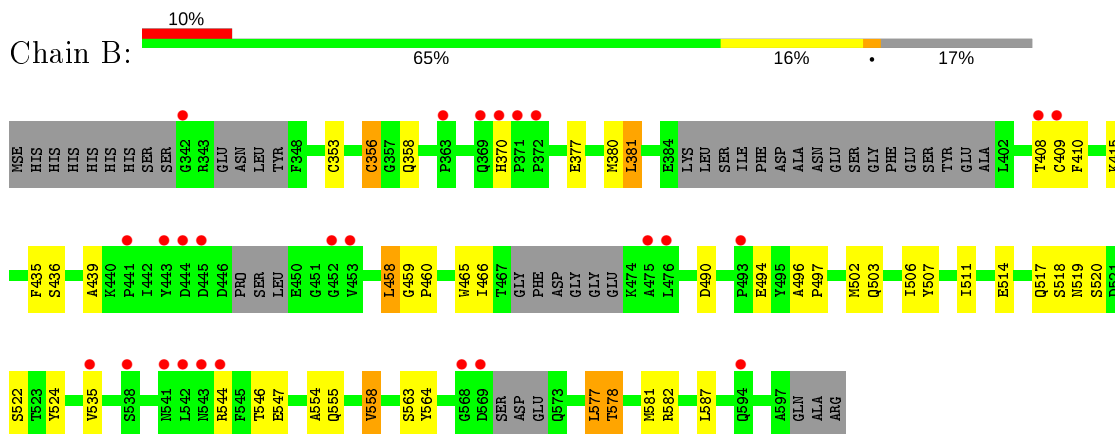
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	57.54Å 59.77Å 96.29Å 90.00° 92.31° 90.00°	Depositor
Resolution (Å)	48.11 – 2.31 41.44 – 2.31	Depositor EDS
% Data completeness (in resolution range)	99.5 (48.11-2.31) 99.5 (41.44-2.31)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.09	Depositor
$< I/\sigma(I) >$ [1]	1.53 (at 2.32Å)	Xtriage
Refinement program	REFMAC 5.5.0044	Depositor
R, R_{free}	0.213 , 0.264 0.224 , 0.274	Depositor DCC
R_{free} test set	1473 reflections (5.11%)	wwPDB-VP
Wilson B-factor (Å^2)	47.7	Xtriage
Anisotropy	0.082	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 60.7	EDS
L-test for twinning[2]	$< \lvert L \rvert > = 0.50, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.014 for -k,-h,-l 0.008 for k,h,-l 0.026 for h,-k,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	3560	wwPDB-VP
Average B, all atoms (Å^2)	58.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.25% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< \lvert L \rvert >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, ZN, BGC, NA

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	0.49	1/1775 (0.1%)	0.57	0/2403
1	B	0.42	0/1740	0.54	0/2356
All	All	0.46	1/3515 (0.0%)	0.56	0/4759

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	420	CYS	CB-SG	-5.53	1.72	1.81

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	1741	0	1656	20	0
1	B	1705	0	1608	31	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	12	0	12	0	0
4	A	6	0	8	0	0
5	A	1	0	0	0	0
6	A	67	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	26	0	0	0	0
All	All	3560	0	3284	50	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 7.

All (50) 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:423:ASP:HB2	1:A:506:ILE:HG13	1.49	0.95
1:B:506:ILE:HD12	1:B:507:TYR:N	1.96	0.81
1:B:506:ILE:HD12	1:B:507:TYR:H	1.47	0.79
1:A:503:GLN:HB2	1:A:506:ILE:HD12	1.68	0.76
1:B:356:CYS:HB3	1:B:358:GLN:H	1.52	0.75
1:A:590:VAL:HG13	1:A:594:GLN:HB2	1.68	0.72
1:B:408:THR:HG22	1:B:409:CYS:H	1.55	0.71
1:B:502:MSE:SE	1:B:506:ILE:HD13	2.44	0.68
1:B:546:THR:HG22	1:B:547:GLU:N	2.11	0.65
1:B:554:ALA:O	1:B:558:VAL:HG13	1.99	0.61
1:B:502:MSE:SE	1:B:506:ILE:CD1	2.99	0.61
1:B:353:CYS:SG	1:B:356:CYS:HB2	2.41	0.60
1:A:506:ILE:HD13	1:A:506:ILE:H	1.68	0.58
1:B:353:CYS:HB3	1:B:356:CYS:HB2	1.86	0.58
1:A:503:GLN:HB2	1:A:506:ILE:CD1	2.36	0.56
1:B:408:THR:HG23	1:B:490:ASP:H	1.70	0.56
1:A:376:ASP:HB3	1:A:378:PRO:HD2	1.87	0.55
1:B:578:THR:HG22	1:B:581:MSE:H	1.71	0.55
1:A:503:GLN:O	1:A:506:ILE:HD13	2.07	0.55
1:A:565:ASP:OD2	1:A:576:PHE:HB2	2.06	0.54
1:A:423:ASP:CB	1:A:506:ILE:HG13	2.32	0.54
1:B:546:THR:HG22	1:B:547:GLU:H	1.70	0.54
1:B:507:TYR:CE2	1:B:511:ILE:HD11	2.43	0.54
1:B:535:VAL:HG22	1:B:544:ARG:HH12	1.73	0.53
1:B:503:GLN:HB2	1:B:506:ILE:HG13	1.89	0.53
1:B:517:GLN:NE2	1:B:564:TYR:OH	2.41	0.53
1:B:496:ALA:HB3	1:B:497:PRO:HD3	1.92	0.52
1:B:408:THR:HG22	1:B:409:CYS:N	2.26	0.50
1:B:380:MSE:HE2	1:B:439:ALA:H	1.76	0.49
1:A:590:VAL:CG1	1:A:594:GLN:HB2	2.41	0.48
1:A:506:ILE:N	1:A:506:ILE:HD13	2.29	0.48
1:A:517:GLN:NE2	1:A:564:TYR:OH	2.47	0.47

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:410:PHE:HA	1:B:436:SER:O	2.15	0.47
1:A:378:PRO:HB3	1:B:577:LEU:HB3	1.98	0.46
1:B:520:SER:O	1:B:578:THR:HG23	2.16	0.45
1:B:435:PHE:O	1:B:458:LEU:HB2	2.16	0.45
1:B:459:GLY:HA3	1:B:460:PRO:HA	1.75	0.45
1:B:377:GLU:HG2	1:B:381:LEU:HD22	1.99	0.44
1:A:573:GLN:HA	1:A:574:PRO:HD3	1.89	0.44
1:A:377:GLU:HG2	1:A:381:LEU:HD22	2.00	0.44
1:B:524:TYR:CD2	1:B:587:LEU:HD22	2.53	0.44
1:A:503:GLN:O	1:A:506:ILE:CD1	2.66	0.43
1:A:466:ILE:HG22	1:A:514:GLU:HG2	2.00	0.43
1:B:546:THR:CG2	1:B:547:GLU:N	2.79	0.42
1:A:504:GLU:OE1	1:A:552:ARG:NH2	2.52	0.42
1:A:524:TYR:OH	1:A:551:LEU:HD21	2.20	0.42
1:B:380:MSE:HE2	1:B:439:ALA:N	2.35	0.41
1:B:380:MSE:HE2	1:B:439:ALA:CB	2.51	0.41
1:B:519:ASN:HB3	1:B:522:SER:OG	2.21	0.40
1:A:423:ASP:OD1	1:A:506:ILE:CD1	2.69	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	217/268 (81%)	214 (99%)	3 (1%)	0	100	100
1	B	211/268 (79%)	204 (97%)	7 (3%)	0	100	100
All	All	428/536 (80%)	418 (98%)	10 (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	186/230 (81%)	181 (97%)	5 (3%)	44	60
1	B	178/230 (77%)	162 (91%)	16 (9%)	9	11
All	All	364/460 (79%)	343 (94%)	21 (6%)	20	27

All (21) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	360	LEU
1	A	381	LEU
1	A	506	ILE
1	A	548	ASP
1	A	590	VAL
1	B	356	CYS
1	B	370	HIS
1	B	381	LEU
1	B	415	LYS
1	B	458	LEU
1	B	465	TRP
1	B	466	ILE
1	B	494	GLU
1	B	514	GLU
1	B	518	SER
1	B	555	GLN
1	B	558	VAL
1	B	563	SER
1	B	577	LEU
1	B	578	THR
1	B	582	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (10) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	383	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	457	ASN
1	A	517	GLN
1	A	553	HIS
1	B	358	GLN
1	B	379	GLN
1	B	383	ASN
1	B	503	GLN
1	B	517	GLN
1	B	519	ASN

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 5 ligands modelled in this entry, 3 are monoatomic - leaving 2 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	GOL	A	703	-	5,5,5	0.31	0	5,5,5	0.82	0
3	BGC	A	702	-	12,12,12	0.53	0	17,17,17	0.68	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

Full wwPDB X-ray Structure Validation Report

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	703	-	-	3/4/4/4	-
3	BGC	A	702	-	-	0/2/22/22	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	703	GOL	C1-C2-C3-O3
4	A	703	GOL	O2-C2-C3-O3
4	A	703	GOL	O1-C1-C2-O2

There are no ring outliers.

No monomer is involved in short contacts.

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(Å²)	Q<0.9
1	A	223/268 (83%)	0.41	8 (3%) 42	49	33, 46, 74, 105	0
1	B	219/268 (81%)	0.64	26 (11%) 4	6	41, 66, 105, 120	0
All	All	442/536 (82%)	0.52	34 (7%) 13	18	33, 58, 99, 120	0

All (34) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	388	ILE	7.4
1	B	342	GLY	5.0
1	B	443	TYR	4.8
1	A	389	PHE	4.6
1	B	568	GLY	4.3
1	B	542	LEU	3.8
1	B	538	SER	3.8
1	B	543	ASN	3.8
1	A	472	GLY	3.7
1	B	569	ASP	3.6
1	B	493	PRO	3.4
1	B	476	LEU	3.4
1	B	372	PRO	3.2
1	B	409	CYS	3.2
1	B	541	ASN	3.2
1	B	370	HIS	3.2
1	B	544	ARG	2.9
1	B	441	PRO	2.8
1	A	350	GLY	2.7
1	B	408	THR	2.6
1	B	369	GLN	2.6
1	B	535	VAL	2.4
1	B	363	PRO	2.4
1	A	569	ASP	2.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	371	PRO	2.3
1	A	351	PRO	2.3
1	B	475	ALA	2.2
1	B	445	ASP	2.2
1	B	594	GLN	2.2
1	B	444	ASP	2.1
1	B	453	VAL	2.1
1	A	473	GLU	2.1
1	A	451	GLY	2.1
1	B	452	GLY	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, 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
4	GOL	A	703	6/6	0.74	0.26	71,72,73,73	0
3	BGC	A	702	12/12	0.79	0.32	103,104,105,105	0
5	NA	A	1	1/1	0.90	0.10	50,50,50,50	0
2	ZN	B	701	1/1	0.99	0.15	60,60,60,60	0
2	ZN	A	701	1/1	0.99	0.15	35,35,35,35	0

6.5 Other polymers (i)

There are no such residues in this entry.