WORLDWIDE PDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

May 25, 2020 – 02:40 am BST

PDB ID	:	4BPU
Title	:	Crystal structure of human primase in heterodimeric form, comprising PriS and truncated PriL lacking the C-terminal Fe-S domain.
Authors	:	Kilkenny, M.L.; Perera, R.L.; Pellegrini, L.
Deposited on	:	2013-05-28
Resolution	:	2.70 Å(reported)

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

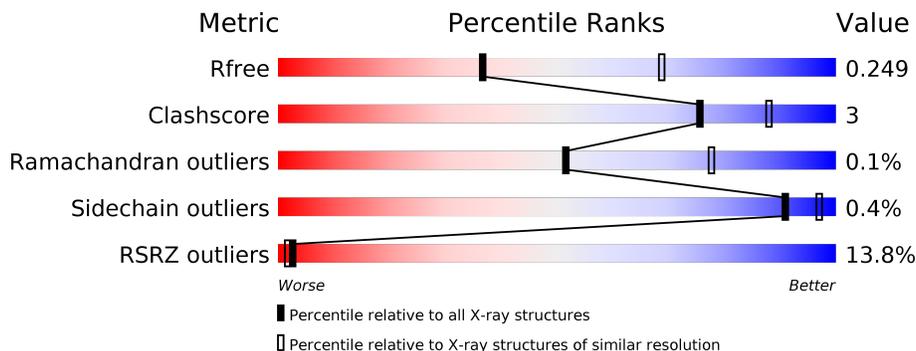|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 2.70 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		0.249
Clashscore		3
Ramachandran outliers		0.1%
Sidechain outliers		0.4%
RSRZ outliers		13.8%

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	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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	423	6% / 82% / 7% / 10%
1	C	423	16% / 80% / 6% / 13%
2	B	253	12% / 73% / 12% / 15%
2	D	253	15% / 76% / 6% / 17%

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 20014 atoms, of which 9940 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 PRIMASE SMALL SUBUNIT.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	380	Total 6450	C 2077	H 3221	N 563	O 576	S 13	0	10	0
1	C	367	Total 6171	C 1994	H 3079	N 536	O 549	S 13	0	3	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	expression tag	UNP P49642
A	-1	THR	-	expression tag	UNP P49642
A	0	SER	-	expression tag	UNP P49642
A	72	ALA	LYS	engineered mutation	UNP P49642
A	73	ALA	MET	engineered mutation	UNP P49642
C	-2	GLY	-	expression tag	UNP P49642
C	-1	THR	-	expression tag	UNP P49642
C	0	SER	-	expression tag	UNP P49642
C	72	ALA	LYS	engineered mutation	UNP P49642
C	73	ALA	MET	engineered mutation	UNP P49642

- Molecule 2 is a protein called DNA PRIMASE LARGE SUBUNIT.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	216	Total 3616	C 1148	H 1835	N 307	O 324	S 2	0	0	0
2	D	209	Total 3505	C 1116	H 1773	N 297	O 317	S 2	0	0	0

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	C	1	Total	Zn	0	0
			1	1		

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).
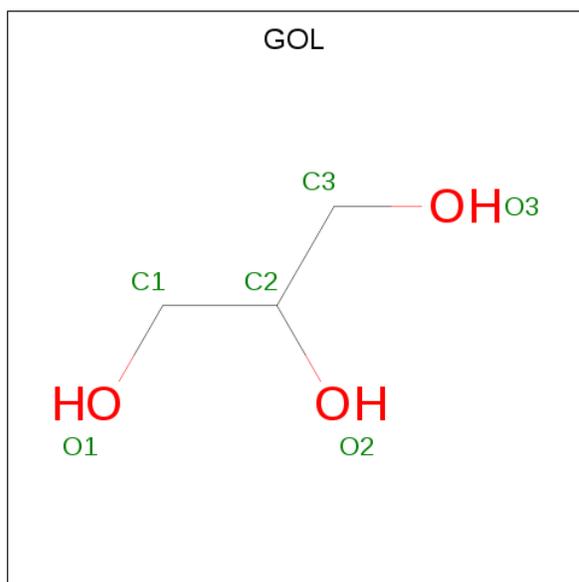
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total	C	H	O	0	0
			14	3	8	3		
4	A	1	Total	C	H	O	0	0
			14	3	8	3		
4	A	1	Total	C	H	O	0	0
			14	3	8	3		
4	A	1	Total	C	H	O	0	0
			14	3	8	3		

- Molecule 5 is water.

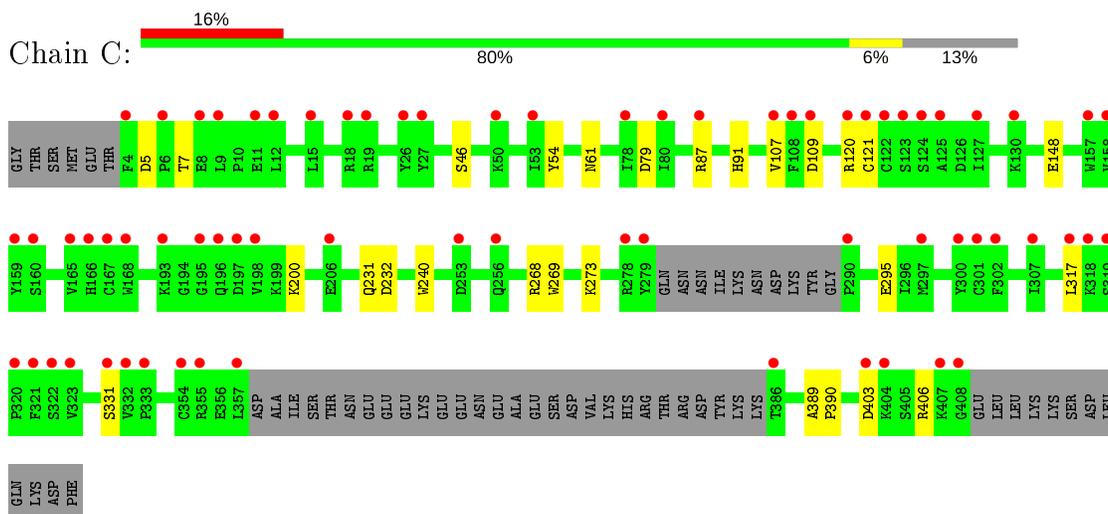| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	143	Total	O	0	0
			143	143		
5	B	13	Total	O	0	0
			13	13		
5	C	56	Total	O	0	0
			56	56		
5	D	2	Total	O	0	0
			2	2		

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA and DNA 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.
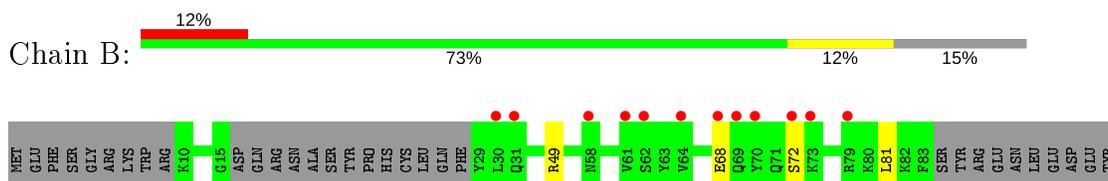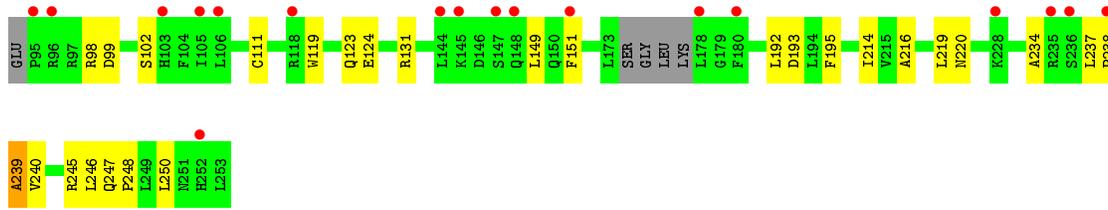
• Molecule 1: DNA PRIMASE SMALL SUBUNIT
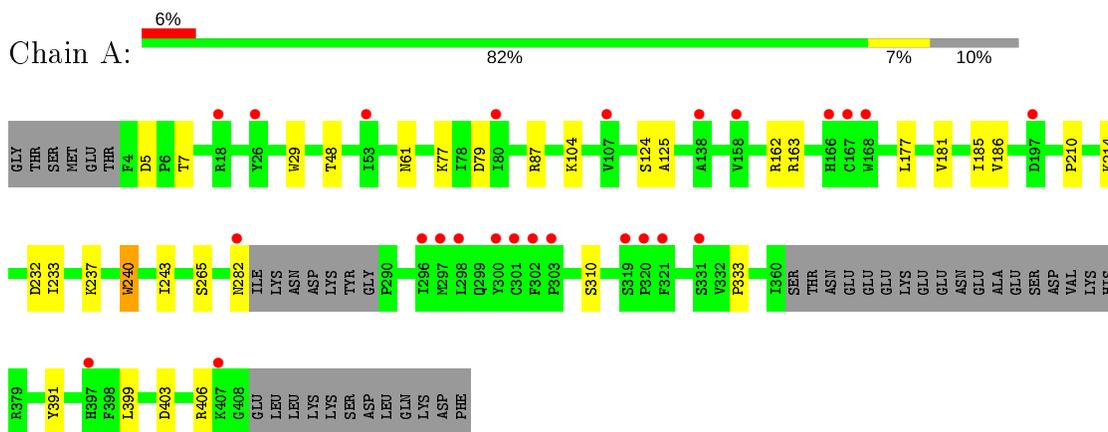
• Molecule 1: DNA PRIMASE SMALL SUBUNIT

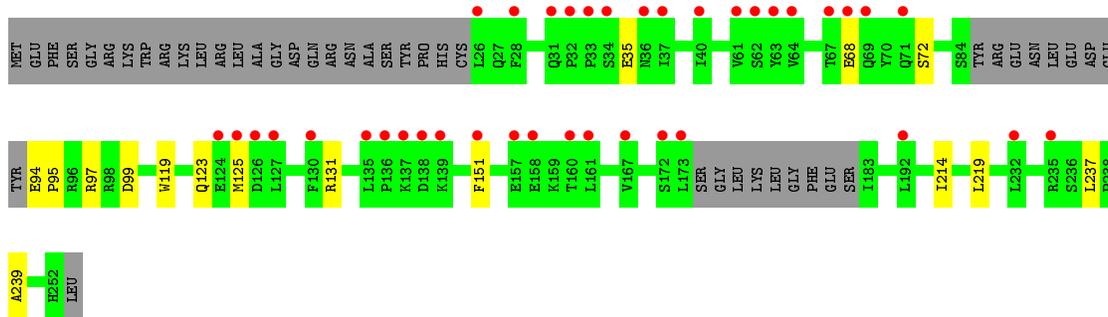• Molecule 2: DNA PRIMASE LARGE SUBUNIT

GLU P95 R96 R97 R98 D99 S102 H103 F104 I105 L106 C111 R118 W119 Q123 E124 R131 L144 K145 D146 S147 Q148 L149 Q150 F151 L173 SER GLY LEU LYS L178 G179 F180 L192 D193 L194 F195 I214 V215 A216 L219 N220 K228 A234 R235 S236 L237 P238

A239 V240 R245 L246 Q247 P248 L249 L250 N251 H252 L253

• Molecule 2: DNA PRIMASE LARGE SUBUNIT

Chain D:

15%
76%   6%   17%

MET GLU PHE SER GLY ARG LYS TRP ARG LYS LEU ARG LEU ALA GLY ASP GLN GLN ARG ASN ALA SER TYR PRO HIS CYS L26 Q27 F28 Q31 P32 P33 S34 E35 N36 I37 I40 V61 S62 Y63 V64 T67 E68 Q69 Y70 Q71 S72 S84 TYR ARG GLU ASN LEU GLU ASP GLU

TYR E94 P95 R96 R97 R98 D99 W119 Q123 E124 M125 D126 L127 F130 R131 L135 P136 K137 D138 K139 F151 E157 E158 K159 T160 L161 V167 S172 L173 SER GLY LEU LYS LEU GLY PHE GLU SER I183 L192 I214 L219 L232 R235 S236 L237 P238

A239 H252 LEU

4　Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	114.55Å　68.64Å　126.76Å 90.00°　104.36°　90.00°	Depositor
Resolution (Å)	29.62　–　2.70 29.90　–　2.70	Depositor EDS
% Data completeness (in resolution range)	98.5 (29.62-2.70) 98.5 (29.90-2.70)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.58 (at 2.68Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.203　,　0.245 0.205　,　0.249	Depositor DCC
R_{free} test set	2658 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å^2)	81.6	Xtriage
Anisotropy	0.254	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.40 , 70.7	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	20014	wwPDB-VP
Average B, all atoms (Å^2)	100.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.97% 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 (i)

5.1 Standard geometry (i)

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

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	0.37	0/3360	0.52	0/4535
1	C	0.32	0/3182	0.47	0/4296
2	B	0.33	0/1811	0.52	0/2430
2	D	0.32	0/1763	0.52	0/2371
All	All	0.34	0/10116	0.50	0/13632

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	3229	3221	3167	17	0
1	C	3092	3079	3058	18	0
2	B	1781	1835	1832	19	0
2	D	1732	1773	1770	9	0
3	A	1	0	0	0	0
3	C	1	0	0	0	0
4	A	24	32	32	0	0
5	A	143	0	0	3	0
5	B	13	0	0	0	0

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	C	56	0	0	6	0
5	D	2	0	0	0	0
All	All	10074	9940	9859	62	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 3.

The worst 5 of 62 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:61:ASN:OD1	1:A:87:ARG:NH2	2.20	0.74
1:C:403:ASP:OD1	1:C:406:ARG:NH2	2.26	0.69
1:A:403:ASP:OD1	1:A:406:ARG:NH2	2.26	0.68
2:B:193:ASP:OD2	2:B:220:ASN:ND2	2.27	0.66
1:A:79:ASP:OD1	5:A:2030:HOH:O	2.14	0.65

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	384/423 (91%)	368 (96%)	16 (4%)	0	100	100
1	C	364/423 (86%)	353 (97%)	11 (3%)	0	100	100
2	B	208/253 (82%)	198 (95%)	9 (4%)	1 (0%)	29	54
2	D	203/253 (80%)	197 (97%)	6 (3%)	0	100	100
All	All	1159/1352 (86%)	1116 (96%)	42 (4%)	1 (0%)	51	78

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	239	ALA

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	363/393 (92%)	360 (99%)	3 (1%)	81	93
1	C	344/393 (88%)	343 (100%)	1 (0%)	92	98
2	B	196/230 (85%)	196 (100%)	0	100	100
2	D	192/230 (84%)	192 (100%)	0	100	100
All	All	1095/1246 (88%)	1091 (100%)	4 (0%)	91	97

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	124	SER
1	A	240	TRP
1	A	282	ASN
1	C	240	TRP

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

Mol	Chain	Res	Type
1	A	166	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 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	$\#|Z| > 2$
4	GOL	A	1410	-	5,5,5	0.40	0	5,5,5	0.31	0
4	GOL	A	1413	-	5,5,5	0.41	0	5,5,5	0.44	0
4	GOL	A	1411	-	5,5,5	0.34	0	5,5,5	0.17	0
4	GOL	A	1412	-	5,5,5	0.44	0	5,5,5	0.28	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	A	1410	-	-	3/4/4/4	-
4	GOL	A	1413	-	-	4/4/4/4	-
4	GOL	A	1411	-	-	2/4/4/4	-
4	GOL	A	1412	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 11 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	1413	GOL	C1-C2-C3-O3
4	A	1413	GOL	O2-C2-C3-O3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
4	A	1410	GOL	C1-C2-C3-O3
4	A	1411	GOL	C1-C2-C3-O3
4	A	1412	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 ⓘ

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	380/423 (89%)	0.45	25 (6%)	18	16	43, 64, 97, 125	0
1	C	367/423 (86%)	0.99	69 (18%)	1	0	58, 95, 135, 161	0
2	B	216/253 (85%)	0.84	30 (13%)	2	2	48, 106, 149, 158	0
2	D	209/253 (82%)	0.98	38 (18%)	1	1	75, 110, 151, 167	0
All	All	1172/1352 (86%)	0.79	162 (13%)	2	2	43, 91, 143, 167	0

The worst 5 of 162 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	195	GLY	7.1
1	C	197	ASP	6.9
2	B	148	GLN	6.9
1	C	322	SER	6.9
1	C	198	VAL	6.6

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

6.4 Ligands ⓘ

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(Å²)	Q<0.9
4	GOL	A	1410	6/6	0.79	0.37	62,89,107,108	0
4	GOL	A	1412	6/6	0.84	0.42	53,75,97,102	0
4	GOL	A	1413	6/6	0.91	0.46	76,91,98,101	0
4	GOL	A	1411	6/6	0.93	0.19	71,85,102,106	0
3	ZN	C	1409	1/1	0.98	0.06	104,104,104,104	0
3	ZN	A	1409	1/1	1.00	0.16	61,61,61,61	0

6.5 Other polymers (i)

There are no such residues in this entry.