W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

May 26, 2020 – 10:45 pm BST

PDB ID	:	5C73
Title	:	ATP-driven lipid-linked oligosaccharide flippase PglK in outward-occluded conformation
Authors	:	Perez, C.; Locher, K.P.
Deposited on	:	2015-06-24
Resolution	:	5.90 Å(reported)

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

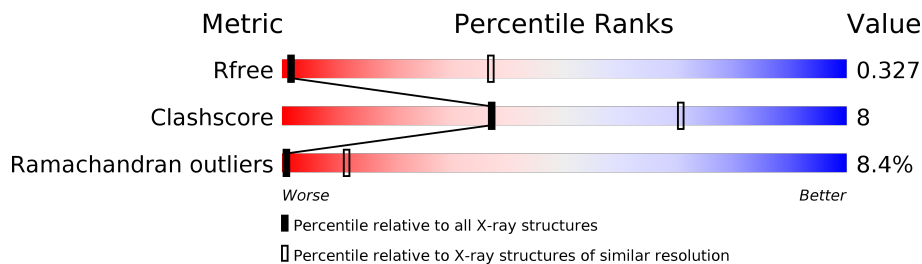|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 5.90 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		0.327
Clashscore		8
Ramachandran outliers		8.4%

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	1016 (7.94-3.86)
Clashscore	141614	1042 (7.88-3.90)
Ramachandran outliers	138981	1011 (7.94-3.86)

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%

Mol	Chain	Length	Quality of chain
1	A	564	82% 15% •
1	B	564	82% 16% •
1	C	564	83% 15% •
1	F	564	85% 14% •
1	G	564	82% 16% •
1	K	564	85% 13% •

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 16752 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 Protein glycosylation K.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
1	C	564	Total 2792	C 1664	N 564	O 564	0	0	0
1	K	564	Total 2792	C 1664	N 564	O 564	0	0	0
1	A	564	Total 2792	C 1664	N 564	O 564	0	0	0
1	B	564	Total 2792	C 1664	N 564	O 564	0	0	0
1	G	564	Total 2792	C 1664	N 564	O 564	0	0	0
1	F	564	Total 2792	C 1664	N 564	O 564	0	0	0

There are 72 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	2	VAL	LEU	conflict	UNP Q0P9C4
C	27	VAL	ILE	conflict	UNP Q0P9C4
C	105	LYS	TYR	conflict	UNP Q0P9C4
C	202	LEU	VAL	conflict	UNP Q0P9C4
C	325	ASN	ASP	conflict	UNP Q0P9C4
C	341	GLY	GLU	conflict	UNP Q0P9C4
C	410	LYS	GLU	conflict	UNP Q0P9C4
C	416	SER	ASN	conflict	UNP Q0P9C4
C	418	ALA	THR	conflict	UNP Q0P9C4
C	456	LYS	ARG	conflict	UNP Q0P9C4
C	505	ILE	MET	conflict	UNP Q0P9C4
C	510	GLN	GLU	conflict	UNP Q0P9C4
K	2	VAL	LEU	conflict	UNP Q0P9C4
K	27	VAL	ILE	conflict	UNP Q0P9C4
K	105	LYS	TYR	conflict	UNP Q0P9C4
K	202	LEU	VAL	conflict	UNP Q0P9C4
K	325	ASN	ASP	conflict	UNP Q0P9C4

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
K	341	GLY	GLU	conflict	UNP Q0P9C4
K	410	LYS	GLU	conflict	UNP Q0P9C4
K	416	SER	ASN	conflict	UNP Q0P9C4
K	418	ALA	THR	conflict	UNP Q0P9C4
K	456	LYS	ARG	conflict	UNP Q0P9C4
K	505	ILE	MET	conflict	UNP Q0P9C4
K	510	GLN	GLU	conflict	UNP Q0P9C4
A	2	VAL	LEU	conflict	UNP Q0P9C4
A	27	VAL	ILE	conflict	UNP Q0P9C4
A	105	LYS	TYR	conflict	UNP Q0P9C4
A	202	LEU	VAL	conflict	UNP Q0P9C4
A	325	ASN	ASP	conflict	UNP Q0P9C4
A	341	GLY	GLU	conflict	UNP Q0P9C4
A	410	LYS	GLU	conflict	UNP Q0P9C4
A	416	SER	ASN	conflict	UNP Q0P9C4
A	418	ALA	THR	conflict	UNP Q0P9C4
A	456	LYS	ARG	conflict	UNP Q0P9C4
A	505	ILE	MET	conflict	UNP Q0P9C4
A	510	GLN	GLU	conflict	UNP Q0P9C4
B	2	VAL	LEU	conflict	UNP Q0P9C4
B	27	VAL	ILE	conflict	UNP Q0P9C4
B	105	LYS	TYR	conflict	UNP Q0P9C4
B	202	LEU	VAL	conflict	UNP Q0P9C4
B	325	ASN	ASP	conflict	UNP Q0P9C4
B	341	GLY	GLU	conflict	UNP Q0P9C4
B	410	LYS	GLU	conflict	UNP Q0P9C4
B	416	SER	ASN	conflict	UNP Q0P9C4
B	418	ALA	THR	conflict	UNP Q0P9C4
B	456	LYS	ARG	conflict	UNP Q0P9C4
B	505	ILE	MET	conflict	UNP Q0P9C4
B	510	GLN	GLU	conflict	UNP Q0P9C4
G	2	VAL	LEU	conflict	UNP Q0P9C4
G	27	VAL	ILE	conflict	UNP Q0P9C4
G	105	LYS	TYR	conflict	UNP Q0P9C4
G	202	LEU	VAL	conflict	UNP Q0P9C4
G	325	ASN	ASP	conflict	UNP Q0P9C4
G	341	GLY	GLU	conflict	UNP Q0P9C4
G	410	LYS	GLU	conflict	UNP Q0P9C4
G	416	SER	ASN	conflict	UNP Q0P9C4
G	418	ALA	THR	conflict	UNP Q0P9C4
G	456	LYS	ARG	conflict	UNP Q0P9C4
G	505	ILE	MET	conflict	UNP Q0P9C4

Continued from previous page...

| Chain	Residue	Modelled	Actual	Comment	Reference
G	510	GLN	GLU	conflict	UNP Q0P9C4
F	2	VAL	LEU	conflict	UNP Q0P9C4
F	27	VAL	ILE	conflict	UNP Q0P9C4
F	105	LYS	TYR	conflict	UNP Q0P9C4
F	202	LEU	VAL	conflict	UNP Q0P9C4
F	325	ASN	ASP	conflict	UNP Q0P9C4
F	341	GLY	GLU	conflict	UNP Q0P9C4
F	410	LYS	GLU	conflict	UNP Q0P9C4
F	416	SER	ASN	conflict	UNP Q0P9C4
F	418	ALA	THR	conflict	UNP Q0P9C4
F	456	LYS	ARG	conflict	UNP Q0P9C4
F	505	ILE	MET	conflict	UNP Q0P9C4
F	510	GLN	GLU	conflict	UNP Q0P9C4

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. 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. 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.
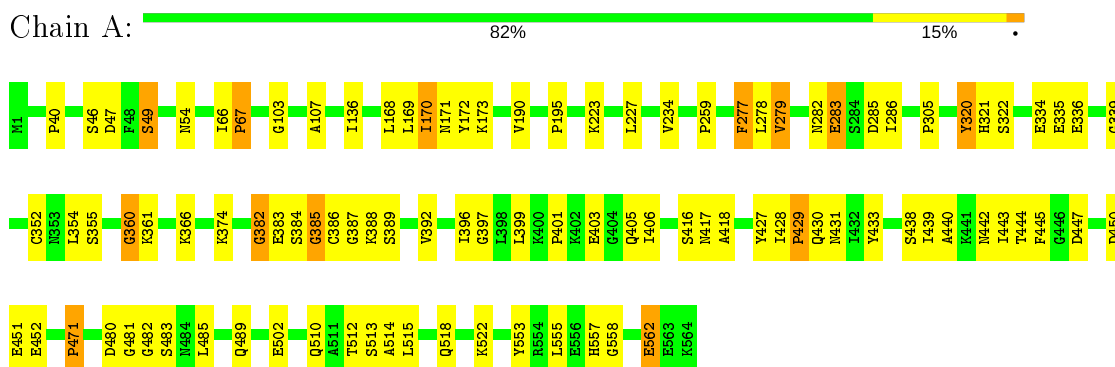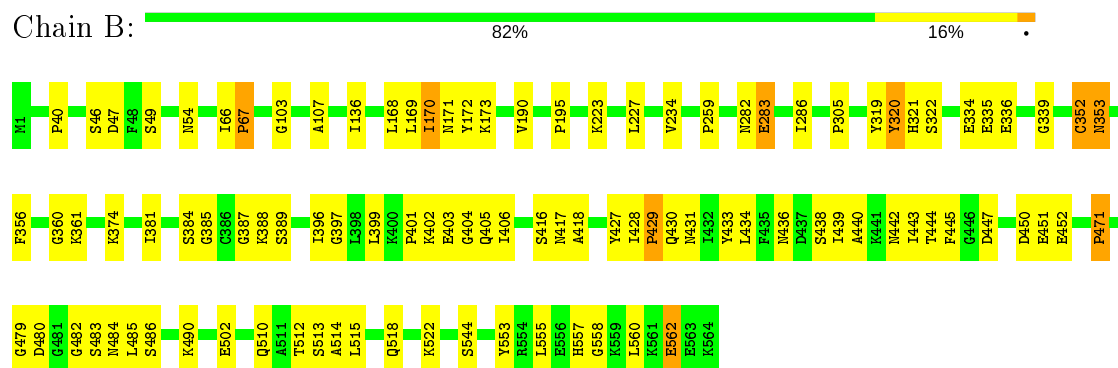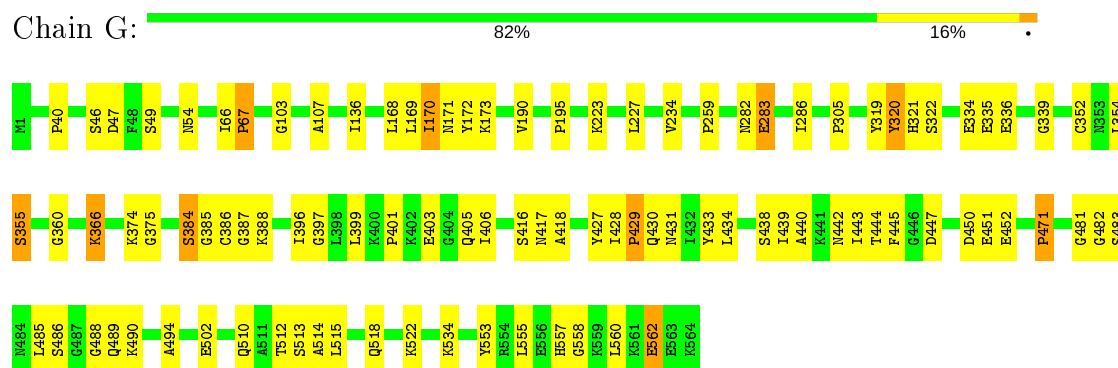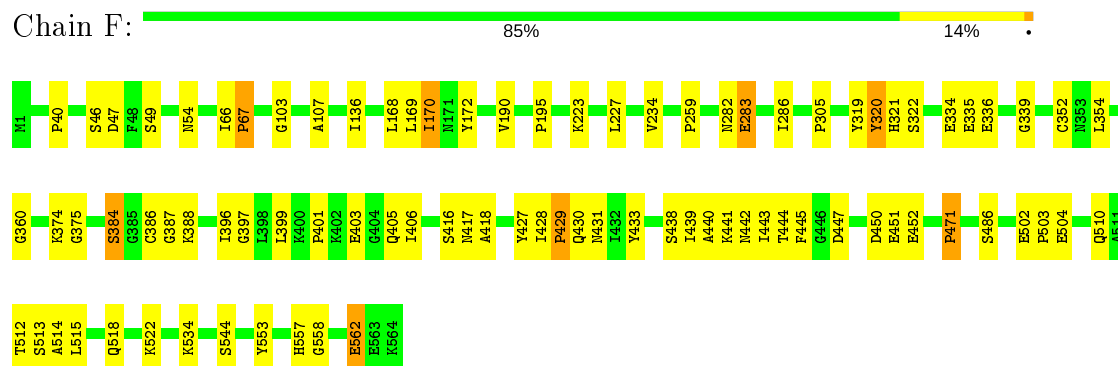
- Molecule 1: Protein glycosylation K
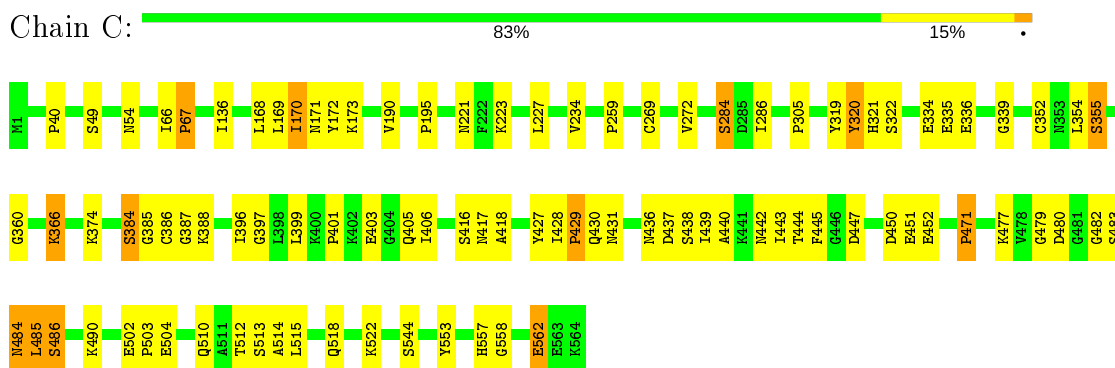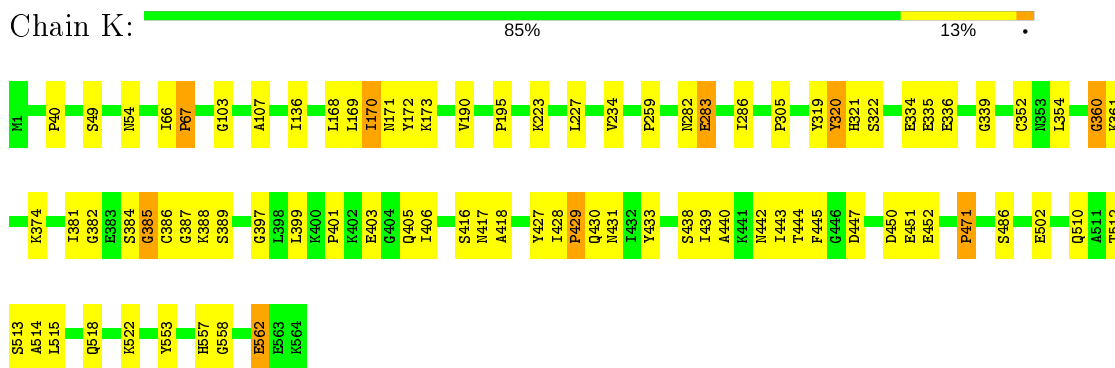
- Molecule 1: Protein glycosylation K

- Molecule 1: Protein glycosylation K

- Molecule 1: Protein glycosylation K

Chain B:

- Molecule 1: Protein glycosylation K

Chain G:

- Molecule 1: Protein glycosylation K

Chain F:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	200.03Å 200.03Å 693.95Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.94 – 5.90 29.94 – 5.50	Depositor EDS
% Data completeness (in resolution range)	84.5 (29.94-5.90) 70.7 (29.94-5.50)	Depositor EDS
R_{merge}	0.20	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.58 (at 5.67Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.284 , 0.317 0.300 , 0.327	Depositor DCC
R_{free} test set	1669 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å^2)	319.2	Xtriage
Anisotropy	0.122	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.56 , 999.0	EDS
L-test for twinning[2]	$<|L|> = 0.42, <L^2> = 0.25$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.72	EDS
Total number of atoms	16752	wwPDB-VP
Average B, all atoms (Å^2)	368.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.78% 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)

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	0.39	0/2791	0.49	8/3889 (0.2%)
1	B	0.38	0/2791	0.49	8/3889 (0.2%)
1	C	0.38	0/2791	0.50	8/3889 (0.2%)
1	F	0.39	0/2791	0.51	8/3889 (0.2%)
1	G	0.39	0/2791	0.50	8/3889 (0.2%)
1	K	0.37	0/2791	0.49	8/3889 (0.2%)
All	All	0.38	0/16746	0.50	48/23334 (0.2%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	C	0	1

There are no bond length outliers.

The worst 5 of 48 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	305	PRO	N-CA-CB	6.10	110.62	103.30
1	C	305	PRO	N-CA-CB	6.04	110.55	103.30
1	K	305	PRO	N-CA-CB	6.01	110.51	103.30
1	K	40	PRO	N-CA-CB	5.99	110.48	103.30
1	B	401	PRO	N-CA-CB	5.96	110.45	103.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	284	SER	Peptide

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	2792	0	1242	34	0
1	B	2792	0	1242	36	0
1	C	2792	0	1242	36	0
1	F	2792	0	1242	24	0
1	G	2792	0	1242	35	0
1	K	2792	0	1242	25	0
All	All	16752	0	7452	188	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 8.

The worst 5 of 188 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:G:434:LEU:CB	1:G:481:GLY:HA2	1.79	1.13
1:C:486:SER:O	1:C:490:LYS:CB	2.07	1.01
1:B:356:PHE:HA	1:B:402:LYS:CB	1.97	0.94
1:C:482:GLY:HA2	1:C:483:SER:O	1.76	0.86
1:C:486:SER:O	1:C:490:LYS:N	2.14	0.81

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	562/564 (100%)	423 (75%)	89 (16%)	50 (9%)	1	11

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles		
1	B	562/564 (100%)	421 (75%)	96 (17%)	45 (8%)	1	12	
1	C	562/564 (100%)	424 (75%)	86 (15%)	52 (9%)	0	10	
1	F	562/564 (100%)	425 (76%)	92 (16%)	45 (8%)	1	12	
1	G	562/564 (100%)	427 (76%)	89 (16%)	46 (8%)	1	12	
1	K	562/564 (100%)	424 (75%)	93 (16%)	45 (8%)	1	12	
All	All	3372/3384 (100%)	2544 (75%)	545 (16%)	283 (8%)	1	11	

5 of 283 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	66	ILE
1	C	67	PRO
1	C	284	SER
1	C	321	HIS
1	C	335	GLU

5.3.2   Protein sidechains (i)

There are no protein residues with a non-rotameric sidechain to report in this entry.

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)

There are no ligands in this entry.

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)

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.