wwPDB X-ray Structure Validation Summary Report (i)

Nov 15, 2023 – 01:45 PM JST

PDB ID	:	6J9X
Title	:	Crystal structure of Trypanosoma brucei gambiense glycerol kinase phosphorylated at Thr12(pyrophosphatase reaction)
Authors	:	Balogun, E.O.; Chishima, T.; Ichinose, M.; Inaoka, D.K.; Kido, Y.; Ibrahim, B.; Bringaud, F.; de Koning, H.; McKerrow, J.H.; Watanabe, Y.; Nozaki, T.; Michels, P.A.M.; Harada, S.; Kita, K.; Shiba, T.
Deposited on	:	2019-01-24
Resolution	:	2.70 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

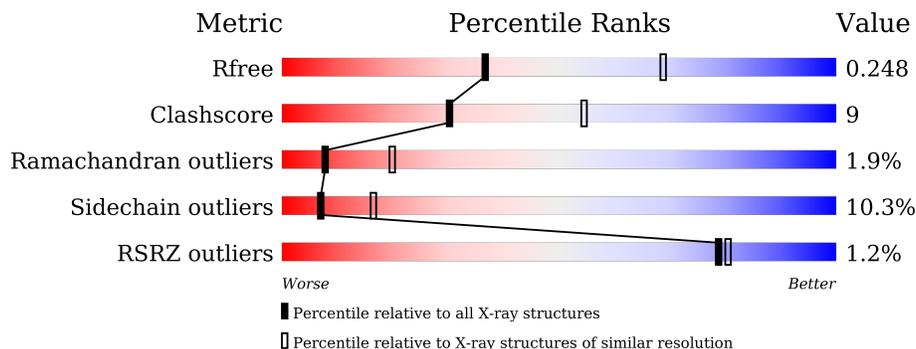|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.70 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		0.248
Clashscore		9
Ramachandran outliers		1.9%
Sidechain outliers		10.3%
RSRZ outliers		1.2%

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 of 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	518	3% 67% 27% · ··
1	C	518	.% 73% 23% ··
2	B	518	77% 20% ··
2	D	518	.% 73% 23% ··

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
3	GOL	A	601	-	-	X	-

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 15910 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 Glycerol kinase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	513	Total 3961	C 2499	N 694	O 734	P 1	S 33	0	0	0
1	C	513	Total 3961	C 2499	N 694	O 734	P 1	S 33	0	0	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-5	GLY	-	expression tag	UNP D3KVM3
A	-4	ILE	-	expression tag	UNP D3KVM3
A	-3	ASP	-	expression tag	UNP D3KVM3
A	-2	PRO	-	expression tag	UNP D3KVM3
A	-1	PHE	-	expression tag	UNP D3KVM3
A	0	THR	-	expression tag	UNP D3KVM3
C	-5	GLY	-	expression tag	UNP D3KVM3
C	-4	ILE	-	expression tag	UNP D3KVM3
C	-3	ASP	-	expression tag	UNP D3KVM3
C	-2	PRO	-	expression tag	UNP D3KVM3
C	-1	PHE	-	expression tag	UNP D3KVM3
C	0	THR	-	expression tag	UNP D3KVM3

- Molecule 2 is a protein called Glycerol kinase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	513	Total 3957	C 2499	N 694	O 731	S 33	0	0	0
2	D	513	Total 3957	C 2499	N 694	O 731	S 33	0	0	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-5	GLY	-	expression tag	UNP D3KVM3
B	-4	ILE	-	expression tag	UNP D3KVM3
B	-3	ASP	-	expression tag	UNP D3KVM3
B	-2	PRO	-	expression tag	UNP D3KVM3
B	-1	PHE	-	expression tag	UNP D3KVM3
B	0	THR	-	expression tag	UNP D3KVM3
D	-5	GLY	-	expression tag	UNP D3KVM3
D	-4	ILE	-	expression tag	UNP D3KVM3
D	-3	ASP	-	expression tag	UNP D3KVM3
D	-2	PRO	-	expression tag	UNP D3KVM3
D	-1	PHE	-	expression tag	UNP D3KVM3
D	0	THR	-	expression tag	UNP D3KVM3

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).
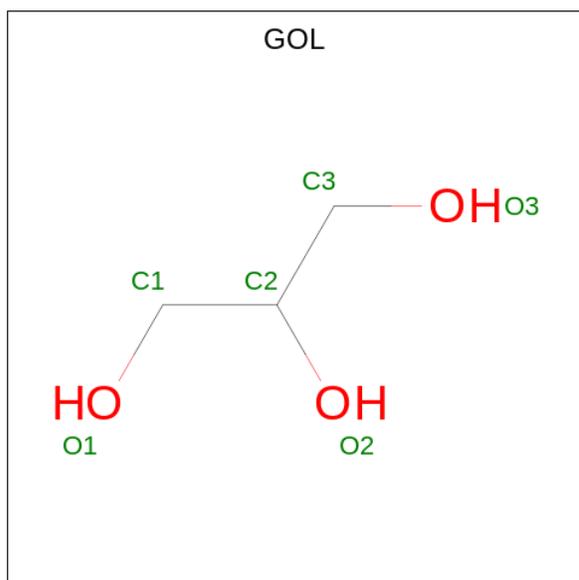
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total 6	C 3	O 3	0	0
3	B	1	Total 6	C 3	O 3	0	0
3	C	1	Total 6	C 3	O 3	0	0
3	D	1	Total 6	C 3	O 3	0	0

- Molecule 4 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	10	Total 10	O 10	0	0
4	B	11	Total 11	O 11	0	0
4	C	12	Total 12	O 12	0	0
4	D	17	Total 17	O 17	0	0

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.
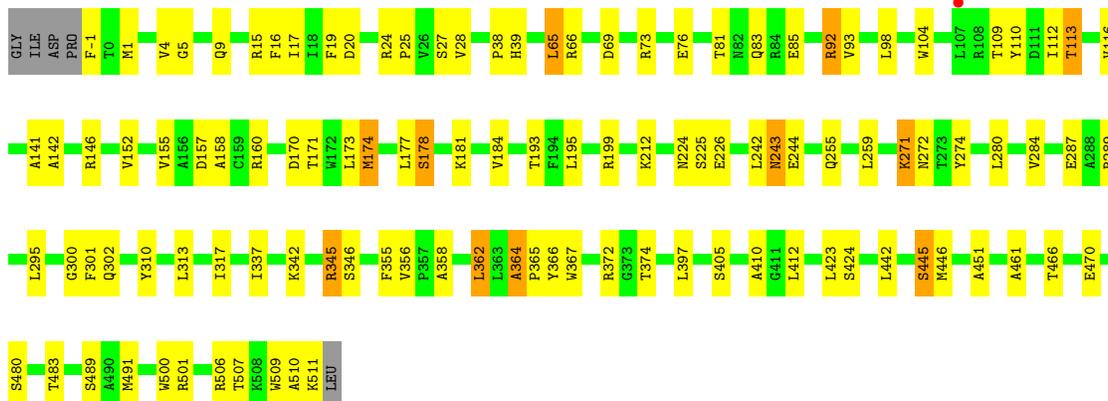
• Molecule 1: Glycerol kinase
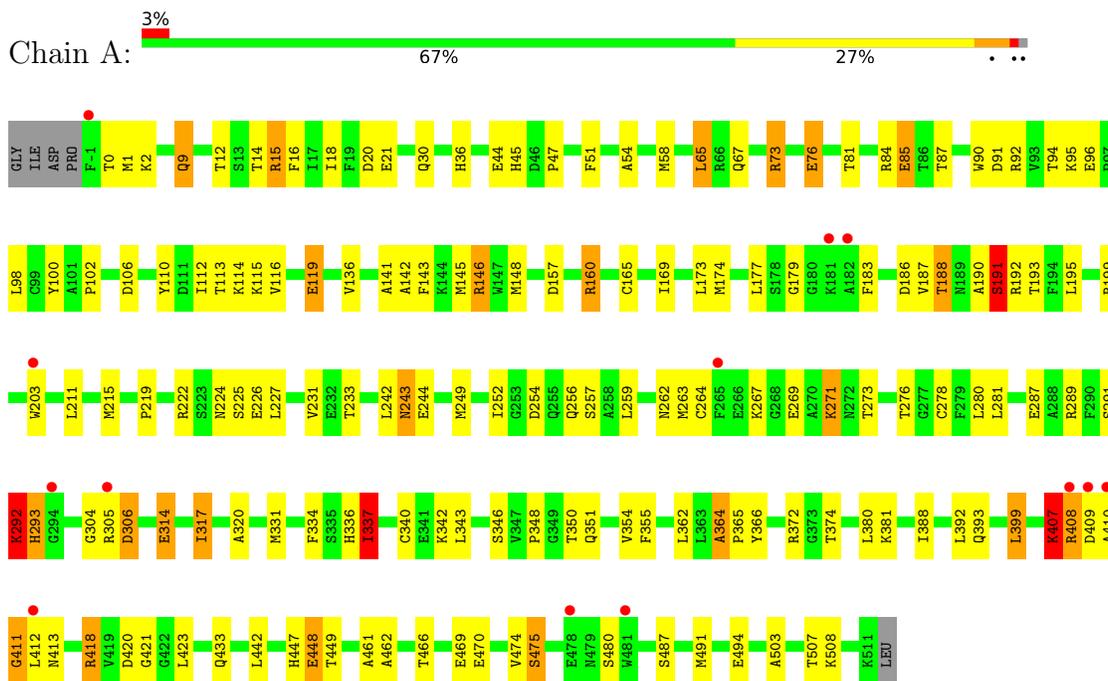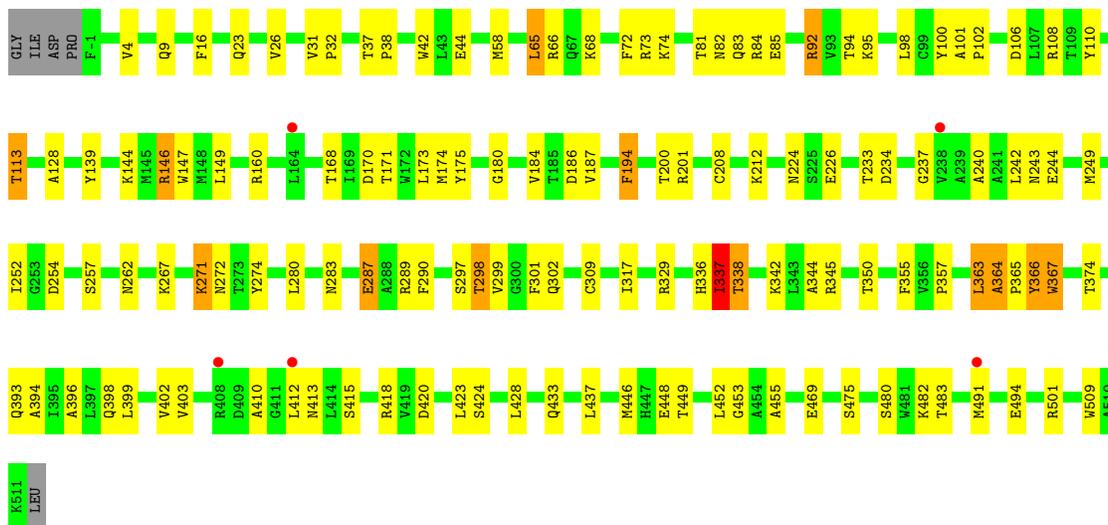

• Molecule 1: Glycerol kinase
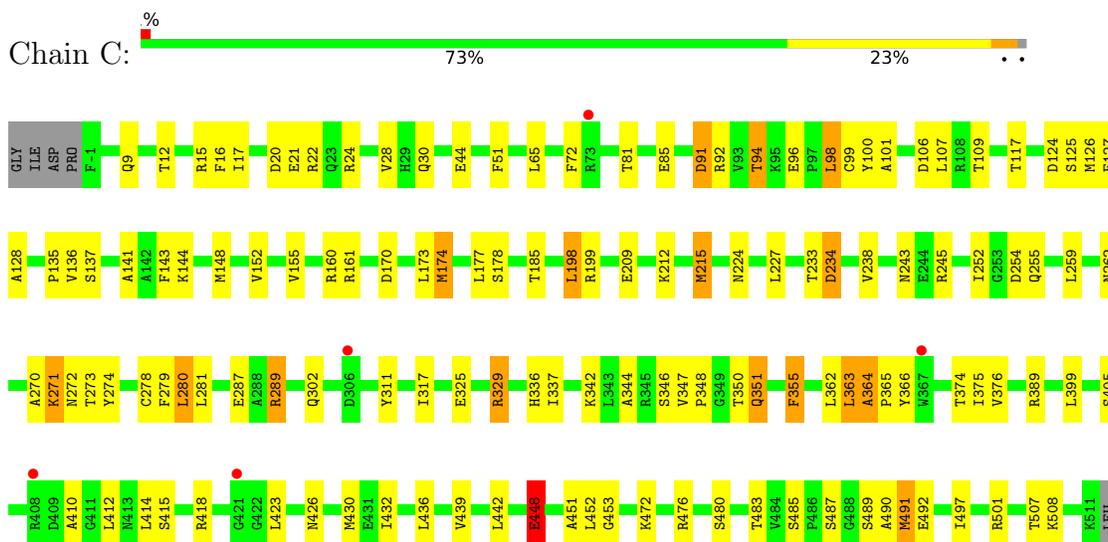
• Molecule 2: Glycerol kinase

Chain B:

• Molecule 2: Glycerol kinase

Chain D:

4　Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	62.38Å　119.34Å　154.73Å 90.00°　90.82°　90.00°	Depositor
Resolution (Å)	19.89　–　2.70 19.89　–　2.70	Depositor EDS
% Data completeness (in resolution range)	99.1 (19.89-2.70) 99.4 (19.89-2.70)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.64 (at 2.71Å)	Xtriage
Refinement program	REFMAC 5.7.0032	Depositor
R, R_{free}	0.184　,　0.247 0.185　,　0.248	Depositor DCC
R_{free} test set	3151 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å^2)	88.9	Xtriage
Anisotropy	0.105	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 58.2	EDS
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.129 for h,-k,-l	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	15910	wwPDB-VP
Average B, all atoms (Å^2)	96.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.08% 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 ⓘ

5.1 Standard geometry ⓘ

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

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	0.56	0/4031	0.80	1/5452 (0.0%)
1	C	0.55	0/4031	0.79	0/5452
2	B	0.58	0/4039	0.81	0/5465
2	D	0.56	0/4039	0.78	0/5465
All	All	0.56	0/16140	0.79	1/21834 (0.0%)

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	A	0	4

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	331	MET	N-CA-C	-5.83	95.25	111.00

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	291	SER	Peptide
1	A	306	ASP	Peptide
1	A	407	LYS	Peptide
1	A	411	GLY	Peptide

WORLDWIDE wwPDB PROTEIN DATA BANK

5.2 Too-close contacts ⓘ

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	3961	0	3968	98	0
1	C	3961	0	3968	74	0
2	B	3957	0	3970	55	0
2	D	3957	0	3970	71	0
3	A	6	0	8	4	0
3	B	6	0	8	0	0
3	C	6	0	8	0	0
3	D	6	0	8	3	0
4	A	10	0	0	1	0
4	B	11	0	0	2	0
4	C	12	0	0	2	0
4	D	17	0	0	3	0
All	All	15910	0	15908	296	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 9.

The worst 5 of 296 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:D:84:ARG:HD3	3:D:601:GOL:O3	1.56	1.03
2:B:364:ALA:HB1	2:B:365:PRO:HD2	1.53	0.90
1:A:190:ALA:O	1:A:191:SER:HB3	1.77	0.84
1:A:410:ALA:HB3	1:A:411:GLY:HA2	1.62	0.82
1:A:364:ALA:HB1	1:A:365:PRO:HD3	1.62	0.81

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	510/518 (98%)	443 (87%)	51 (10%)	16 (3%)	4	9
1	C	510/518 (98%)	450 (88%)	50 (10%)	10 (2%)	7	19
2	B	511/518 (99%)	476 (93%)	30 (6%)	5 (1%)	15	37
2	D	511/518 (99%)	465 (91%)	38 (7%)	8 (2%)	9	24
All	All	2042/2072 (99%)	1834 (90%)	169 (8%)	39 (2%)	8	20

5 of 39 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	188	THR
1	A	203	TRP
1	A	293	HIS
1	A	337	ILE
1	A	364	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	420/424 (99%)	370 (88%)	50 (12%)	5	12
1	C	420/424 (99%)	378 (90%)	42 (10%)	7	18
2	B	421/425 (99%)	380 (90%)	41 (10%)	8	19
2	D	421/425 (99%)	381 (90%)	40 (10%)	8	20
All	All	1682/1698 (99%)	1509 (90%)	173 (10%)	7	16

5 of 173 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	317	ILE
2	D	146	ARG
1	C	342	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	485	SER
2	D	287	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	224	ASN
2	D	23	GLN
1	C	293	HIS
2	D	224	ASN
1	A	433	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

2 non-standard protein/DNA/RNA residues 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	#$|Z| > 2$
1	TPO	A	12	1	8,10,11	1.48	1 (12%)	10,14,16	1.86	2 (20%)
1	TPO	C	12	1	8,10,11	1.52	1 (12%)	10,14,16	1.61	2 (20%)

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
1	TPO	A	12	1	-	5/9/11/13	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
1	TPO	C	12	1	-	5/9/11/13	-

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	12	TPO	P-OG1	3.92	1.66	1.59
1	C	12	TPO	P-OG1	3.86	1.66	1.59

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	12	TPO	CG2-CB-CA	-4.58	104.12	113.16
1	C	12	TPO	CG2-CB-CA	-3.07	107.11	113.16
1	A	12	TPO	P-OG1-CB	2.66	131.25	123.21
1	C	12	TPO	CB-CA-N	-2.07	100.80	114.41

There are no chirality outliers.

5 of 10 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	12	TPO	N-CA-CB-CG2
1	A	12	TPO	N-CA-CB-OG1
1	A	12	TPO	C-CA-CB-CG2
1	A	12	TPO	O-C-CA-CB
1	C	12	TPO	N-CA-CB-CG2

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

4 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	#$|Z| > 2$
3	GOL	D	601	-	5,5,5	0.39	0	5,5,5	1.73	1 (20%)
3	GOL	B	601	-	5,5,5	0.32	0	5,5,5	0.41	0
3	GOL	A	601	-	5,5,5	0.53	0	5,5,5	1.60	1 (20%)
3	GOL	C	601	-	5,5,5	0.41	0	5,5,5	0.49	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
3	GOL	D	601	-	-	1/4/4/4	-
3	GOL	B	601	-	-	2/4/4/4	-
3	GOL	A	601	-	-	2/4/4/4	-
3	GOL	C	601	-	-	2/4/4/4	-

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	601	GOL	C3-C2-C1	-2.54	101.81	111.70
3	D	601	GOL	C3-C2-C1	-2.49	102.02	111.70

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	601	GOL	O1-C1-C2-C3
3	C	601	GOL	O1-C1-C2-C3
3	C	601	GOL	O1-C1-C2-O2
3	B	601	GOL	C1-C2-C3-O3
3	B	601	GOL	O2-C2-C3-O3

There are no ring outliers.

2 monomers are involved in 7 short contacts:

wwPDB X-ray Structure Validation Summary Report

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	D	601	GOL	3	0
3	A	601	GOL	4	0

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

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	512/518 (98%)	0.09	13 (2%)	57	59	65, 99, 136, 174	1 (0%)
1	C	512/518 (98%)	-0.07	5 (0%)	82	83	67, 96, 122, 160	1 (0%)
2	B	513/518 (99%)	-0.14	1 (0%)	95	96	64, 86, 110, 128	1 (0%)
2	D	513/518 (99%)	-0.11	5 (0%)	82	83	66, 93, 117, 144	1 (0%)
All	All	2050/2072 (98%)	-0.06	24 (1%)	79	80	64, 93, 125, 174	4 (0%)

The worst 5 of 24 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	409	ASP	8.9
1	A	410	ALA	6.5
1	C	367	TRP	5.4
1	A	182	ALA	4.8
1	A	408	ARG	4.3

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

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
1	TPO	A	12	11/12	0.89	0.16	64,83,113,122	0
1	TPO	C	12	11/12	0.91	0.13	74,88,131,136	0

6.3 Carbohydrates ⓘ

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
3	GOL	A	601	6/6	0.87	0.19	78,87,102,117	0
3	GOL	B	601	6/6	0.93	0.18	75,89,99,101	0
3	GOL	D	601	6/6	0.94	0.18	77,82,84,88	0
3	GOL	C	601	6/6	0.96	0.18	70,76,82,84	0

6.5 Other polymers (i)

There are no such residues in this entry.