wwPDB X-ray Structure Validation Summary Report (i)

Oct 30, 2023 – 01:19 PM JST

PDB ID	:	5AZI
Title	:	Crystal structure of glycerol kinase from Trypanosoma brucei gambiense complexed with 4NP
Authors	:	Balogun, E.O.; Inaoka, D.K.; Shiba, T.; Tokuoka, S.M.; Tokumasu, F.; Sakamoto, K.; Michels, P.A.M.; Harada, S.; Kita, K.
Deposited on	:	2015-10-08
Resolution	:	2.45 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 (1)) were used in the production of this report:

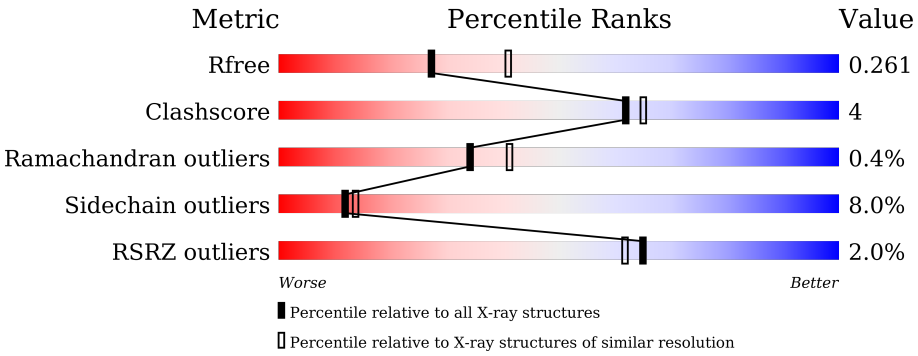
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA,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

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.45 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}	130704	1544 (2.48-2.44)
Clashscore	141614	1613 (2.48-2.44)
Ramachandran outliers	138981	1598 (2.48-2.44)
Sidechain outliers	138945	1598 (2.48-2.44)
RSRZ outliers	127900	1523 (2.48-2.44)

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	81% 16%
1	C	518	87% 11%
2	B	518	3% 82% 16%
2	D	518	4% 81% 16%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 16014 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	C	N	O	S	0	0	0
			3951	2493	694	731	33			
1	C	513	Total	C	N	O	S	0	0	0
			3951	2493	694	731	33			

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

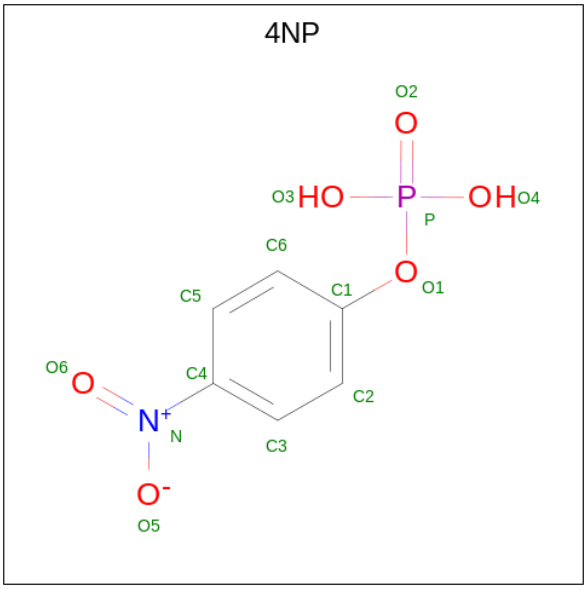
- Molecule 2 is a protein called Glycerol kinase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	513	Total	C	N	O	P	S	0	0	0
			3955	2493	694	734	1	33			
2	D	513	Total	C	N	O	P	S	0	0	0
			3955	2493	694	734	1	33			

There are 12 discrepancies between the modelled and reference sequences:

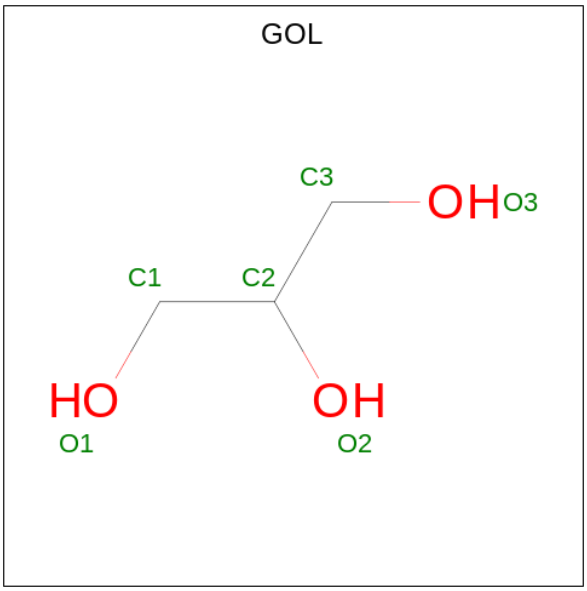
Chain	Residue	Modelled	Actual	Comment	Reference
B	-5	GLY	-	expression tag	UNP D3KVM3
B	-4	ILE	-	expression tag	UNP D3KVM3
B	-3	ASP	-	expression tag	UNP D3KVM3
B	-2	PRO	-	expression tag	UNP D3KVM3
B	-1	PHE	-	expression tag	UNP D3KVM3
B	0	THR	-	expression tag	UNP D3KVM3
D	-5	GLY	-	expression tag	UNP D3KVM3
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 4-NITROPHENYL PHOSPHATE (three-letter code: 4NP) (formula: $C_6H_6NO_6P$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	A	1	Total	C	N	O	P	0	0
			14	6	1	6	1		
3	C	1	Total	C	N	O	P	0	0
			14	6	1	6	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 O 6 3 3	0	0
4	B	1	Total C O 6 3 3	0	0
4	C	1	Total C O 6 3 3	0	0
4	D	1	Total C O 6 3 3	0	0

- Molecule 5 is water.

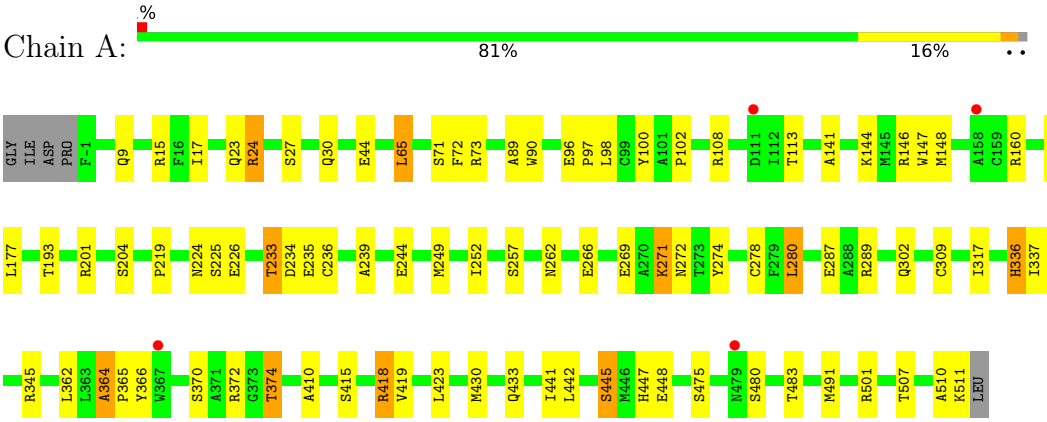
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	47	Total O 47 47	0	0
5	B	24	Total O 24 24	0	0
5	C	49	Total O 49 49	0	0
5	D	30	Total O 30 30	0	0

3 Residue-property plots (i)

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.

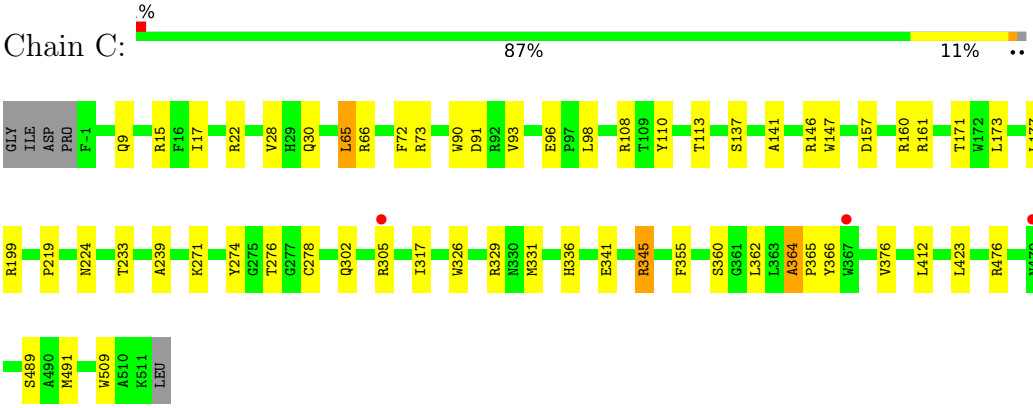
• Molecule 1: Glycerol kinase

Chain A:



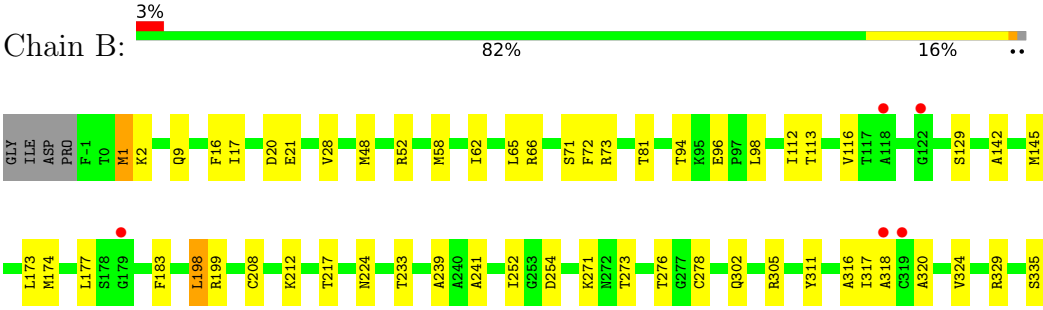
• Molecule 1: Glycerol kinase

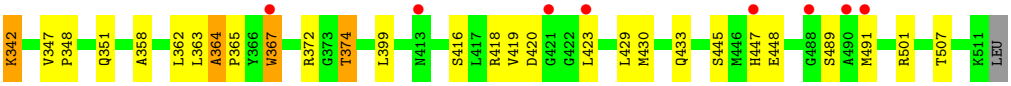
Chain C:



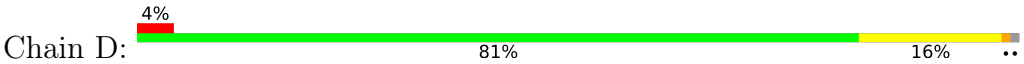
• Molecule 2: Glycerol kinase

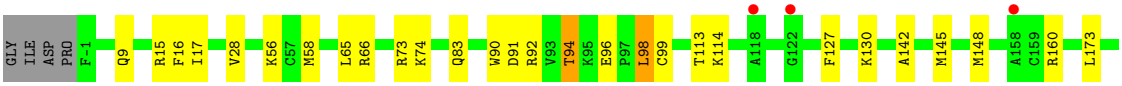
Chain B:

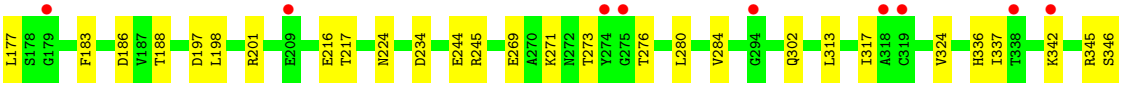


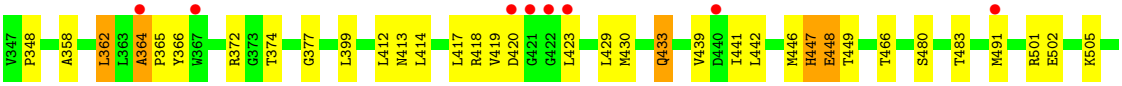


• Molecule 2: Glycerol kinase









4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	63.67Å 120.25Å 154.59Å 90.00° 89.96° 90.00°	Depositor
Resolution (Å)	20.00 – 2.45 19.85 – 2.45	Depositor EDS
% Data completeness (in resolution range)	84.8 (20.00-2.45) 84.7 (19.85-2.45)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	6.08 (at 2.44Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, R_{free}	0.196 , 0.260 0.201 , 0.261	Depositor DCC
R_{free} test set	3627 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	37.6	Xtriage
Anisotropy	0.180	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 24.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.480 for h,-k,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	16014	wwPDB-VP
Average B, all atoms (Å^2)	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.23% 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: GOL, TPO, 4NP

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	0.50	0/4032	0.69	1/5456 (0.0%)
1	C	0.50	0/4032	0.68	0/5456
2	B	0.47	0/4024	0.68	0/5443
2	D	0.47	0/4024	0.67	0/5443
All	All	0.49	0/16112	0.68	1/21798 (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	C	0	1
2	B	0	1
All	All	0	2

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	24	ARG	NE-CZ-NH1	5.75	123.17	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	316	ALA	Peptide
1	C	345	ARG	Sidechain

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	3951	0	3963	47	0
1	C	3951	0	3963	22	0
2	B	3955	0	3962	39	0
2	D	3955	0	3962	33	0
3	A	14	0	4	0	0
3	C	14	0	4	0	0
4	A	6	0	8	0	0
4	B	6	0	8	0	0
4	C	6	0	8	0	0
4	D	6	0	8	0	0
5	A	47	0	0	1	0
5	B	24	0	0	0	0
5	C	49	0	0	0	0
5	D	30	0	0	0	0
All	All	16014	0	15890	139	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 4.

The worst 5 of 139 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:224:ASN:HD22	1:A:302:GLN:H	1.26	0.84
2:D:364:ALA:HB1	2:D:365:PRO:CD	2.15	0.77
1:C:364:ALA:HB1	1:C:365:PRO:CD	2.15	0.77
2:B:364:ALA:HB1	2:B:365:PRO:CD	2.16	0.76
1:A:364:ALA:HB1	1:A:365:PRO:CD	2.19	0.72

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	511/518 (99%)	480 (94%)	30 (6%)	1 (0%)	47	57
1	C	511/518 (99%)	482 (94%)	28 (6%)	1 (0%)	47	57
2	B	510/518 (98%)	482 (94%)	24 (5%)	4 (1%)	19	22
2	D	510/518 (98%)	476 (93%)	31 (6%)	3 (1%)	25	29
All	All	2042/2072 (99%)	1920 (94%)	113 (6%)	9 (0%)	34	41

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	448	GLU
2	D	448	GLU
1	A	364	ALA
2	B	364	ALA
1	C	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	420/425 (99%)	389 (93%)	31 (7%)	13	16
1	C	420/425 (99%)	391 (93%)	29 (7%)	15	18
2	B	419/424 (99%)	386 (92%)	33 (8%)	12	14
2	D	419/424 (99%)	377 (90%)	42 (10%)	7	7
All	All	1678/1698 (99%)	1543 (92%)	135 (8%)	12	14

5 of 135 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	271	LYS
2	D	342	LYS
2	D	466	THR
2	B	317	ILE
2	B	305	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 24 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	224	ASN
2	D	23	GLN
1	C	336	HIS
2	D	151	ASN
2	B	23	GLN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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	$\#\|Z\| > 2$
2	TPO	B	276	2	8,10,11	1.25	1 (12%)	10,14,16	1.52	2 (20%)
2	TPO	D	276	2	8,10,11	1.46	1 (12%)	10,14,16	1.48	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
2	TPO	B	276	2	-	4/9/11/13	-
2	TPO	D	276	2	-	4/9/11/13	-

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	276	TPO	P-OG1	3.63	1.66	1.59
2	B	276	TPO	P-OG1	3.09	1.65	1.59

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	276	TPO	CG2-CB-CA	-3.48	106.30	113.16
2	D	276	TPO	CG2-CB-CA	-3.19	106.87	113.16
2	D	276	TPO	P-OG1-CB	2.34	130.28	123.21
2	B	276	TPO	P-OG1-CB	2.01	129.28	123.21

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	276	TPO	N-CA-CB-OG1
2	B	276	TPO	C-CA-CB-CG2
2	D	276	TPO	N-CA-CB-OG1
2	D	276	TPO	CB-OG1-P-O1P
2	B	276	TPO	N-CA-CB-CG2

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

6 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	602	-	5,5,5	0.47	0	5,5,5	0.27	0
3	4NP	A	601	-	13,14,14	1.09	2 (15%)	18,20,20	1.00	0
4	GOL	D	601	-	5,5,5	0.36	0	5,5,5	0.68	0
4	GOL	C	602	-	5,5,5	0.49	0	5,5,5	0.35	0
4	GOL	B	601	-	5,5,5	0.36	0	5,5,5	0.56	0
3	4NP	C	601	-	13,14,14	1.13	2 (15%)	18,20,20	0.74	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	602	-	-	2/4/4/4	-
3	4NP	A	601	-	-	2/7/9/9	0/1/1/1
4	GOL	D	601	-	-	3/4/4/4	-
4	GOL	C	602	-	-	0/4/4/4	-
4	GOL	B	601	-	-	2/4/4/4	-
3	4NP	C	601	-	-	2/7/9/9	0/1/1/1

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	601	4NP	P-O1	2.54	1.63	1.59
3	A	601	4NP	C4-N	-2.43	1.39	1.45
3	C	601	4NP	C4-N	-2.10	1.40	1.45
3	A	601	4NP	P-O1	2.08	1.62	1.59

There are no bond angle outliers.

There are no chirality outliers.

5 of 11 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	601	4NP	C3-C4-N-O6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
3	A	601	4NP	C5-C4-N-O6
3	C	601	4NP	C3-C4-N-O6
3	C	601	4NP	C5-C4-N-O6
4	A	602	GOL	C1-C2-C3-O3

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, 95^{th} 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	513/518 (99%)	-0.09	4 (0%)	86	86	19, 37, 63, 79	1 (0%)
1	C	513/518 (99%)	-0.11	4 (0%)	86	86	18, 37, 61, 80	1 (0%)
2	B	512/518 (98%)	0.09	13 (2%)	57	53	19, 48, 73, 100	1 (0%)
2	D	512/518 (98%)	0.07	20 (3%)	39	36	21, 48, 75, 96	1 (0%)
All	All	2050/2072 (98%)	-0.01	41 (2%)	65	62	18, 41, 70, 100	4 (0%)

The worst 5 of 41 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	318	ALA	6.8
2	D	318	ALA	5.5
2	B	488	GLY	5.2
2	B	319	CYS	4.6
1	A	367	TRP	4.0

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, 95^{th} 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
2	TPO	B	276	11/12	0.91	0.12	44,51,55,57	0
2	TPO	D	276	11/12	0.91	0.12	45,50,51,51	0

6.3 Carbohydrates

There are no monosaccharides 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, 95^{th} 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	4NP	C	601	14/14	0.78	0.24	90,95,101,101	0
3	4NP	A	601	14/14	0.86	0.21	73,88,99,100	0
4	GOL	C	602	6/6	0.89	0.26	44,45,46,46	0
4	GOL	A	602	6/6	0.96	0.16	40,42,45,48	0
4	GOL	B	601	6/6	0.97	0.13	24,26,28,30	0
4	GOL	D	601	6/6	0.99	0.12	24,26,27,27	0

6.5 Other polymers

There are no such residues in this entry.