WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Sep 25, 2023 – 07:36 AM EDT

PDB ID	:	5VHV
Title	:	Pseudomonas fluorescens alkylpurine DNA glycosylase AlkC bound to DNA containing an oxocarbenium-intermediate analog
Authors	:	Shi, R.; Eichman, B.F.
Deposited on	:	2017-04-13
Resolution	:	1.80 Å(reported)

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

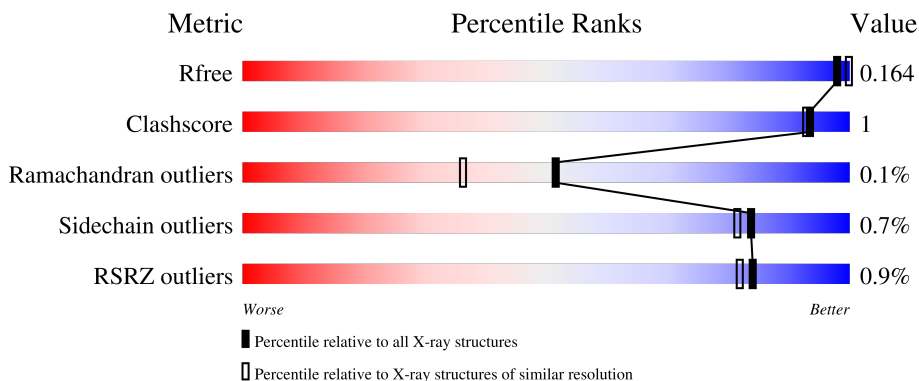|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 1.80 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		0.164
Clashscore		1
Ramachandran outliers		0.1%
Sidechain outliers		0.7%
RSRZ outliers		0.9%

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	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)
RSRZ outliers	127900	5850 (1.80-1.80)

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	361	% 98% •
1	B	361	96% •
2	C	11	9% 82% 18%
2	E	11	9% 45% 55%
3	D	11	9% 73% 18% 9%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	11	9% / 82% / 18%

2 Entry composition ⓘ

There are 8 unique types of molecules in this entry. The entry contains 14539 atoms, of which 6451 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 alkylpurine DNA glycosylase AlkC.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	361	Total	C	H	N	O	S	0	3	0
			5736	1834	2863	522	507	10			
1	B	361	Total	C	H	N	O	S	0	5	0
			5789	1843	2905	524	507	10			

- Molecule 2 is a DNA chain called DNA (5'-D(*TP*GP*TP*CP*CP*AP*(NRI)P*GP*TP* CP*T)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	C	11	Total	C	H	N	O	P	0	0	0
			336	102	126	33	65	10			
2	E	11	Total	C	H	N	O	P	0	0	0
			336	102	126	33	65	10			

- Molecule 3 is a DNA chain called DNA (5'-D(*AP*AP*GP*AP*CP*TP*TP*GP*GP*AP* C)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	D	11	Total	C	H	N	O	P	0	2	0
			412	128	148	55	70	11			
3	F	11	Total	C	H	N	O	P	0	2	0
			412	128	148	55	70	11			

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

GOL

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total 14	C 3	H 8	O 3	0	1
4	F	1	Total 14	C 3	H 8	O 3	0	1

- Molecule 5 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (three-letter code: MES) (formula: $C_6H_{13}NO_4S$).

MES

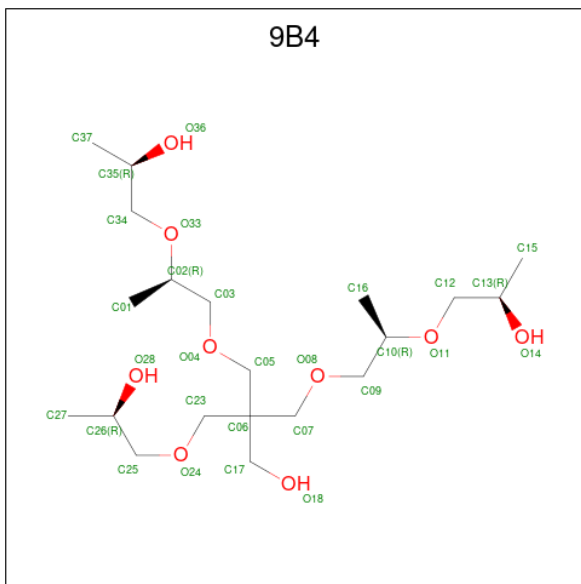| Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
5	A	1	Total 25	C 6	H 13	N 1	O 4	S 1	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
5	B	1	Total	C	H	N	O	S	0	1
			25	6	13	1	4	1		
5	B	1	Total	C	H	N	O	S	0	0
			25	6	13	1	4	1		

- Molecule 6 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	C	1	Total	Na	0	0
			1	1		
6	E	1	Total	Na	0	0
			1	1		

- Molecule 7 is (2R,5R,13R,16R)-9-(hydroxymethyl)-9-{[(2R)-2-hydroxypropoxy]methyl}-5,13-dimethyl-4,7,11,14-tetraoxaheptadecane-2,16-diol (three-letter code: 9B4) (formula: $C_{20}H_{42}O_9$).

| Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
7	D	1	Total	C	H	O	0	1
			69	20	40	9		
7	F	1	Total	C	H	O	0	1
			69	20	40	9		

- Molecule 8 is water.

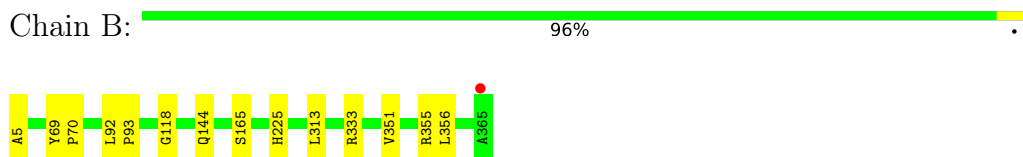| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	526	Total 526	O 526	0	0
8	B	544	Total 544	O 544	0	0
8	C	44	Total 44	O 44	0	0
8	D	57	Total 57	O 57	0	0
8	E	47	Total 47	O 47	0	0
8	F	57	Total 57	O 57	0	0

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: alkylpurine DNA glycosylase AlkC

Chain A:
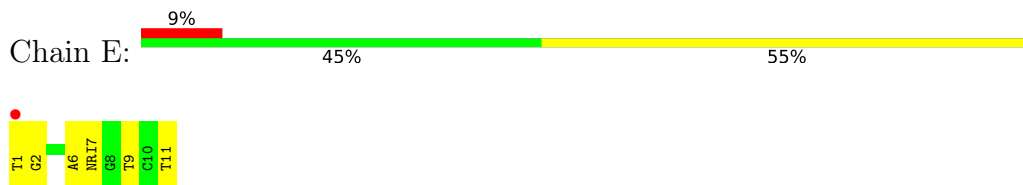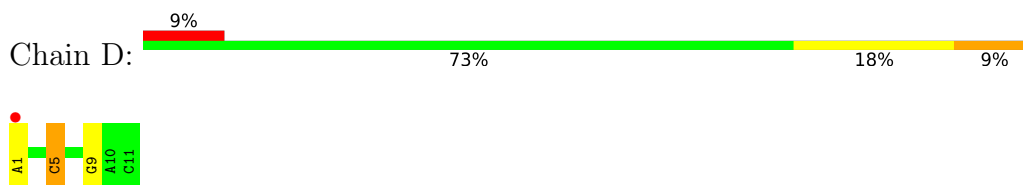
• Molecule 1: alkylpurine DNA glycosylase AlkC

Chain B:
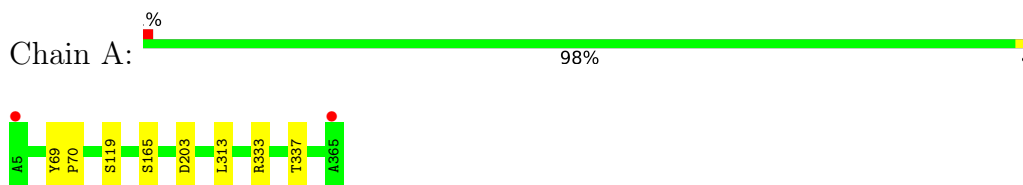
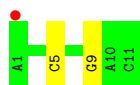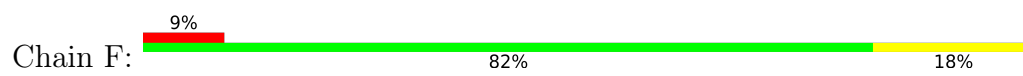• Molecule 2: DNA (5'-D(*TP*GP*TP*CP*CP*AP*(NRI)P*GP*TP*CP*T)-3')

Chain C:

• Molecule 2: DNA (5'-D(*TP*GP*TP*CP*CP*AP*(NRI)P*GP*TP*CP*T)-3')

Chain E:

• Molecule 3: DNA (5'-D(*AP*AP*GP*AP*CP*TP*TP*GP*GP*AP*C)-3')

Chain D:

• Molecule 3: DNA (5'-D(*AP*AP*GP*AP*CP*TP*TP*GP*GP*AP*C)-3')

Chain F:

9%

82%

18%

A1 C5 G9 A10 C11
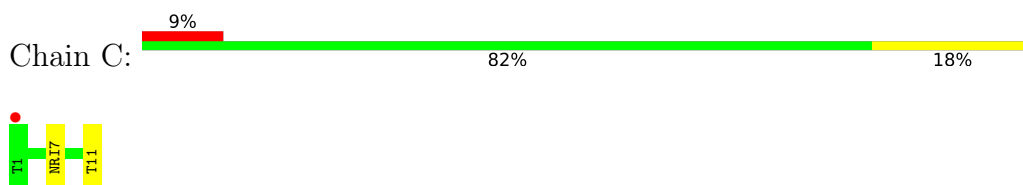
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 61	Depositor
Cell constants a, b, c, α, β, γ	198.37Å 198.37Å 60.19Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	49.29 – 1.80 49.29 – 1.80	Depositor EDS
% Data completeness (in resolution range)	99.0 (49.29-1.80) 95.7 (49.29-1.80)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.54 (at 1.79Å)	Xtriage
Refinement program	PHENIX 1.11.1_2575	Depositor
R, R_{free}	0.141 , 0.164 0.142 , 0.164	Depositor DCC
R_{free} test set	6250 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	19.0	Xtriage
Anisotropy	0.049	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 54.8	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.018 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	14539	wwPDB-VP
Average B, all atoms (Å^2)	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.60% 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: NRI, 9B4, GOL, NA, MES

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	0.41	0/2952	0.58	1/3999 (0.0%)
1	B	0.41	0/2970	0.58	0/4024
2	C	0.97	0/220	1.21	1/335 (0.3%)
2	E	1.13	0/220	1.31	3/335 (0.9%)
3	D	1.04	0/298	1.10	1/459 (0.2%)
3	F	1.00	0/298	1.10	1/459 (0.2%)
All	All	0.55	0/6958	0.71	7/9611 (0.1%)

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	D	5	DC	O5'-P-OP2	-9.92	96.77	105.70
3	F	5	DC	O5'-P-OP2	-7.53	98.92	105.70
2	C	11	DT	O4'-C1'-N1	6.20	112.34	108.00
2	E	11	DT	O4'-C1'-N1	5.73	112.01	108.00
1	A	203	ASP	CB-CG-OD1	5.58	123.33	118.30

There are no chirality outliers.

There are no planarity outliers.

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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2873	2863	2891	4	0
1	B	2884	2905	2907	9	0
2	C	210	126	125	0	0
2	E	210	126	125	3	0
3	D	264	148	148	3	0
3	F	264	148	148	1	0
4	A	6	8	0	0	0
4	F	6	8	0	0	0
5	A	12	13	13	0	0
5	B	24	26	13	0	0
6	C	1	0	0	0	0
6	E	1	0	0	0	0
7	D	29	40	0	0	0
7	F	29	40	0	0	0
8	A	526	0	0	1	0
8	B	544	0	0	4	0
8	C	44	0	0	0	0
8	D	57	0	0	0	0
8	E	47	0	0	0	0
8	F	57	0	0	1	0
All	All	8088	6451	6370	18	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 1.

The worst 5 of 18 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:B:225:HIS:ND1	8:B:502:HOH:O	2.16	0.78
1:B:355:ARG:NH1	8:B:501:HOH:O	2.16	0.75
1:A:119:SER:N	8:A:501:HOH:O	2.19	0.75
1:B:5:ALA:N	8:B:503:HOH:O	2.22	0.72
2:E:1:DT:H2'	2:E:2:DG:C8	2.37	0.58

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	362/361 (100%)	354 (98%)	8 (2%)	0	100	100
1	B	364/361 (101%)	355 (98%)	8 (2%)	1 (0%)	41	27
All	All	726/722 (101%)	709 (98%)	16 (2%)	1 (0%)	51	36

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	118	GLY

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	301/298 (101%)	299 (99%)	2 (1%)	84	81
1	B	303/298 (102%)	301 (99%)	2 (1%)	84	81
All	All	604/596 (101%)	600 (99%)	4 (1%)	84	81

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

Mol	Chain	Res	Type
1	A	313	LEU
1	A	333	ARG
1	B	313	LEU
1	B	333	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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	$\#|Z| > 2$
2	NRI	E	7	2	6,11,12	0.56	0	4,14,17	1.76	1 (25%)
2	NRI	C	7	2	6,11,12	0.45	0	4,14,17	1.69	1 (25%)

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	NRI	E	7	2	-	0/3/15/16	0/1/1/1
2	NRI	C	7	2	-	0/3/15/16	0/1/1/1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	E	7	NRI	C5'-C4'-C3'	-3.39	108.05	114.66
2	C	7	NRI	C5'-C4'-C3'	-3.34	108.15	114.66

There are no chirality outliers.

There are no torsion outliers.

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 ⓘ

Of 9 ligands modelled in this entry, 2 are monoatomic - leaving 7 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	#$|Z| > 2$
5	MES	B	402	-	12,12,12	2.17	1 (8%)	14,16,16	1.68	3 (21%)
5	MES	A	402	-	12,12,12	2.17	1 (8%)	14,16,16	1.44	3 (21%)

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
5	MES	B	402	-	-	5/6/14/14	0/1/1/1
5	MES	A	402	-	-	4/6/14/14	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	B	402	MES	C8-S	-7.26	1.67	1.77
5	A	402	MES	C8-S	-7.19	1.67	1.77

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	B	402	MES	C5-N4-C3	4.16	118.20	108.83
5	A	402	MES	O1S-S-C8	3.01	110.55	106.92
5	B	402	MES	O3S-S-C8	2.84	110.36	105.77
5	A	402	MES	C5-N4-C3	2.33	114.08	108.83
5	B	402	MES	O2S-S-C8	2.17	109.53	106.92

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	A	402	MES	C7-C8-S-O1S
5	B	402	MES	C8-C7-N4-C5
5	B	402	MES	C7-C8-S-O2S
5	B	402	MES	C7-C8-S-O3S
5	A	402	MES	C7-C8-S-O3S

There are no ring outliers.

No monomer is involved in short contacts.

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(Å²)	Q<0.9
1	A	361/361 (100%)	-0.47	2 (0%)	89	87	13, 22, 41, 59	2 (0%)
1	B	361/361 (100%)	-0.56	1 (0%)	94	92	12, 20, 35, 60	5 (1%)
2	C	10/11 (90%)	0.11	1 (10%)	7	5	15, 19, 48, 90	0
2	E	10/11 (90%)	-0.01	1 (10%)	7	5	15, 19, 39, 84	0
3	D	11/11 (100%)	-0.16	1 (9%)	9	7	17, 21, 34, 48	0
3	F	11/11 (100%)	-0.03	1 (9%)	9	7	15, 18, 30, 42	0
All	All	764/766 (99%)	-0.49	7 (0%)	84	82	12, 21, 39, 90	7 (0%)

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	F	1[A]	DA	3.8
1	A	365	ALA	3.5
1	B	365	ALA	3.4
3	D	1[A]	DA	3.4
2	E	1	DT	3.3

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(Å²)	Q<0.9
2	NRI	C	7	11/12	0.99	0.10	12,17,22,26	0
2	NRI	E	7	11/12	0.99	0.09	11,15,19,20	0

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
5	MES	B	402	12/12	0.80	0.23	54,106,125,132	0
7	9B4	D	101[A]	29/29	0.91	0.18	22,48,115,122	0
7	9B4	F	102[A]	29/29	0.92	0.11	23,50,86,97	0
5	MES	A	402	12/12	0.95	0.11	31,63,80,81	0
5	MES	B	401[A]	12/12	0.96	0.15	31,85,112,117	1
4	GOL	F	101[A]	6/6	0.96	0.10	16,23,28,28	0
4	GOL	A	401[A]	6/6	0.98	0.06	20,24,35,35	0
6	NA	C	101	1/1	0.99	0.06	22,22,22,22	0
6	NA	E	101	1/1	0.99	0.07	20,20,20,20	0

6.5 Other polymers (i)

There are no such residues in this entry.