Full wwPDB X-ray Structure Validation Report (i)

Sep 19, 2023 – 03:41 AM EDT

PDB ID	:	5CG9
Title	:	NgTET1 in complex with 5mC DNA in space group P3221
Authors	:	Hashimoto, H.; Pais, J.E.; Dai, N.; Zhang, X.; Zheng, Y.; Cheng, X.
Deposited on	:	2015-07-09
Resolution	:	2.69 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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:

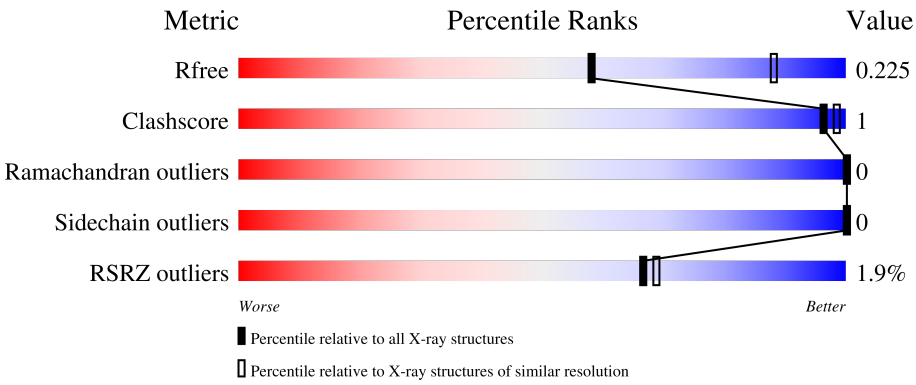
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry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

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.69 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	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	267	96%
1	D	267	97%
2	B	11	9% 82% 18%
2	E	11	100%
3	C	13	15% 92% 8%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	13	8% 92% 8%

2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 5218 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 Tet-like dioxygenase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	267	2121	1362	344	408	7	0	0	0
1	D	264	2085	1334	339	406	6	0	1	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	55	HIS	-	expression tag	UNP D2W6T1
A	56	MET	-	expression tag	UNP D2W6T1
D	55	HIS	-	expression tag	UNP D2W6T1
D	56	MET	-	expression tag	UNP D2W6T1

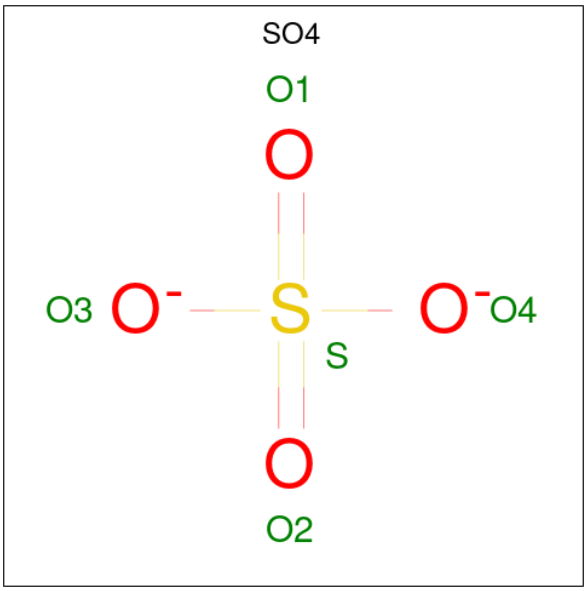
- Molecule 2 is a DNA chain called DNA (5'-D(P*CP*AP*TP*GP*CP*GP*CP*TP*GP*AP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	B	11	224	106	41	66	11	0	0	0
2	E	11	224	106	41	66	11	0	0	0

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

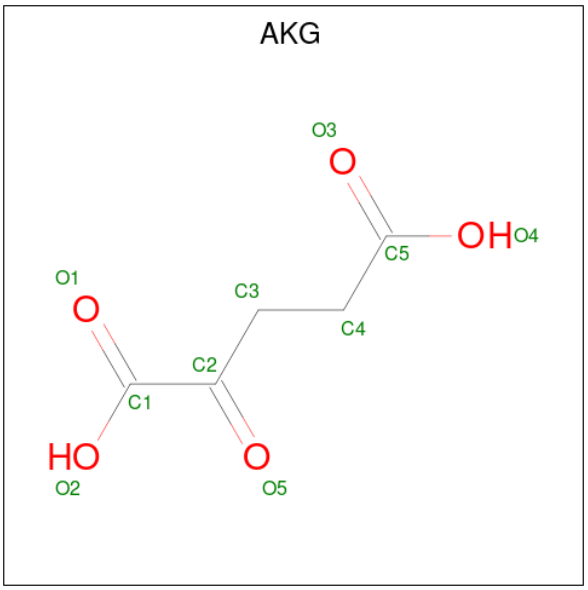
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	C	13	239	113	43	71	12	0	0	0
3	F	13	266	128	50	76	12	0	0	0

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total O S 5 4 1	0	0
4	D	1	Total O S 5 4 1	0	0
4	D	1	Total O S 5 4 1	0	0

- Molecule 5 is 2-OXOGLUTARIC ACID (three-letter code: AKG) (formula: $C_5H_6O_5$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total C O 10 5 5	0	0

Continued on next page...

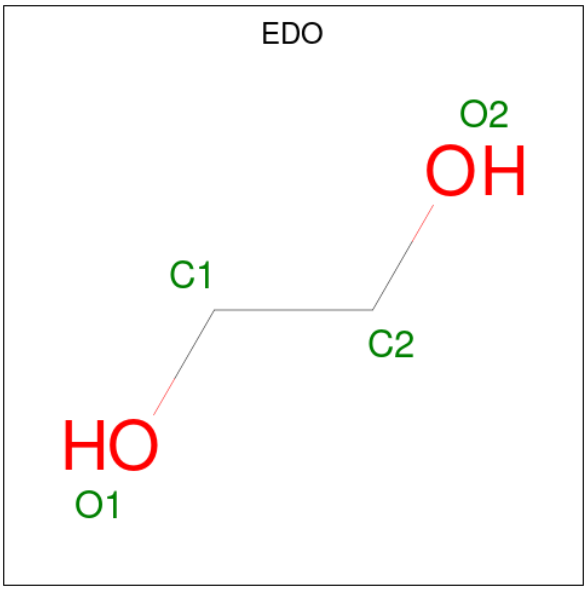
Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
5	D	1	10	5	5	0	0

- Molecule 6 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mn		
6	A	1	1	1	0	0
			Total	Mn		
6	D	1	1	1	0	0

- Molecule 7 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
7	D	1	4	2	2	0	0
			Total	C	O		
7	F	1	4	2	2	0	0

- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
8	A	10	10	10	0	0
			Total	O		
8	C	1	1	1	0	0

Continued on next page...

Continued from previous page...

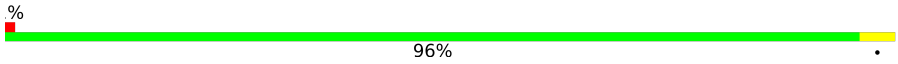
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
8	D	3	3	3	0	0

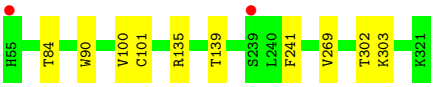
3 Residue-property plots (i)

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: Tet-like dioxygenase

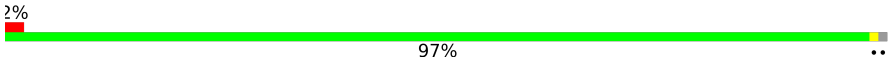
Chain A:

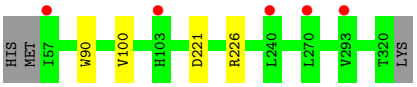




• Molecule 1: Tet-like dioxygenase

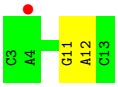
Chain D:





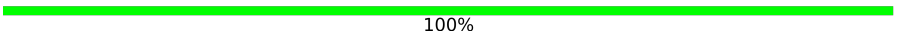
• Molecule 2: DNA (5'-D(P*CP*AP*TP*GP*CP*GP*CP*TP*GP*AP*C)-3')

Chain B:



• Molecule 2: DNA (5'-D(P*CP*AP*TP*GP*CP*GP*CP*TP*GP*AP*C)-3')

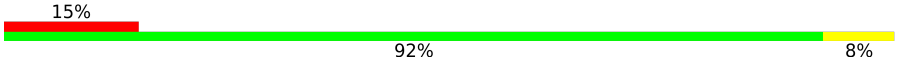
Chain E:

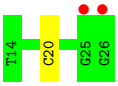


There are no outlier residues recorded for this chain.

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

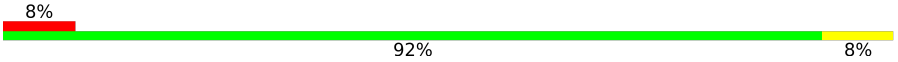
Chain C:





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

Chain F:



T14
C20
G26

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, α, β, γ	191.22Å 191.22Å 51.25Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	29.74 – 2.69 29.74 – 2.69	Depositor EDS
% Data completeness (in resolution range)	98.5 (29.74-2.69) 92.4 (29.74-2.69)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.05 (at 2.68Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.193 , 0.222 0.199 , 0.225	Depositor DCC
R_{free} test set	1492 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	60.0	Xtriage
Anisotropy	0.417	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 41.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.029 for -h,-k,l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	5218	wwPDB-VP
Average B, all atoms (Å^2)	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.94% 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: SO4, EDO, 5CM, MN, AKG

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	0.26	0/2171	0.41	0/2948
1	D	0.25	0/2138	0.40	0/2913
2	B	0.57	0/250	0.91	0/383
2	E	0.56	0/250	0.87	0/383
3	C	0.56	0/244	0.89	0/373
3	F	0.62	0/275	0.90	0/422
All	All	0.34	0/5328	0.55	0/7422

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	2121	0	2016	7	0
1	D	2085	0	1957	2	0
2	B	224	0	124	1	0
2	E	224	0	124	0	0
3	C	239	0	128	0	0
3	F	266	0	147	0	0
4	A	5	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	D	10	0	0	0	0
5	A	10	0	4	0	0
5	D	10	0	4	0	0
6	A	1	0	0	0	0
6	D	1	0	0	0	0
7	D	4	0	6	0	0
7	F	4	0	6	0	0
8	A	10	0	0	0	0
8	C	1	0	0	0	0
8	D	3	0	0	0	0
All	All	5218	0	4516	10	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.

All (10) 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:135:ARG:NH2	1:A:139:THR:O	2.31	0.64
1:D:90:TRP:CZ2	1:D:100:VAL:HG11	2.48	0.48
1:A:90:TRP:CZ2	1:A:100:VAL:HG11	2.49	0.47
1:D:221:ASP:OD1	1:D:226:ARG:NH1	2.48	0.47
1:A:100:VAL:HG12	1:A:101:CYS:N	2.29	0.46
2:B:11:DG:H2"	2:B:12:DA:O5'	2.17	0.45
1:A:302:THR:OG1	1:A:303:LYS:N	2.52	0.42
1:A:241:PHE:CD2	1:A:269:VAL:HG12	2.54	0.42
1:A:84:THR:O	1:A:100:VAL:HG13	2.20	0.42
1:A:100:VAL:CG1	1:A:101:CYS:N	2.85	0.40

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	265/267 (99%)	260 (98%)	5 (2%)	0	100	100
1	D	263/267 (98%)	258 (98%)	5 (2%)	0	100	100
All	All	528/534 (99%)	518 (98%)	10 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	228/246 (93%)	228 (100%)	0	100	100
1	D	226/246 (92%)	226 (100%)	0	100	100
All	All	454/492 (92%)	454 (100%)	0	100	100

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

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

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	$\#\|Z\| > 2$
3	5CM	C	20	3	17,21,22	0.62	0	24,30,33	1.33	3 (12%)
3	5CM	F	20	3	17,21,22	0.57	0	24,30,33	1.28	3 (12%)

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	5CM	C	20	3	-	4/7/21/22	0/2/2/2
3	5CM	F	20	3	-	4/7/21/22	0/2/2/2

There are no bond length outliers.

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	20	5CM	C2'-C1'-N1	3.26	121.29	113.77
3	F	20	5CM	C2'-C1'-N1	2.96	120.58	113.77
3	F	20	5CM	C1'-N1-C2	2.74	122.54	117.74
3	C	20	5CM	C1'-N1-C2	2.64	122.37	117.74
3	C	20	5CM	C5A-C5-C6	-2.47	119.56	122.85
3	F	20	5CM	C5A-C5-C6	-2.34	119.73	122.85

There are no chirality outliers.

All (8) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	C	20	5CM	C2'-C1'-N1-C2
3	F	20	5CM	C2'-C1'-N1-C2
3	C	20	5CM	C2'-C1'-N1-C6
3	F	20	5CM	O4'-C1'-N1-C6
3	C	20	5CM	O4'-C1'-N1-C6
3	F	20	5CM	C2'-C1'-N1-C6
3	F	20	5CM	O4'-C1'-N1-C2
3	C	20	5CM	O4'-C1'-N1-C2

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)

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	$\#\|Z\| > 2$
7	EDO	D	403	-	3,3,3	0.46	0	2,2,2	0.30	0
4	SO4	D	401	-	4,4,4	0.14	0	6,6,6	0.05	0
4	SO4	A	401	-	4,4,4	0.12	0	6,6,6	0.09	0
5	AKG	A	402	6	9,9,9	1.35	1 (11%)	11,11,11	1.05	0
7	EDO	F	101	-	3,3,3	0.46	0	2,2,2	0.34	0
5	AKG	D	404	6	9,9,9	1.33	1 (11%)	11,11,11	1.11	0
4	SO4	D	402	-	4,4,4	0.14	0	6,6,6	0.07	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
5	AKG	D	404	6	-	0/9/9/9	-
5	AKG	A	402	6	-	0/9/9/9	-
7	EDO	F	101	-	-	0/1/1/1	-
7	EDO	D	403	-	-	0/1/1/1	-

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	D	404	AKG	C2-C1	-3.34	1.49	1.53
5	A	402	AKG	C2-C1	-3.30	1.49	1.53

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

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 (i)

6.1 Protein, DNA and RNA chains (i)

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	267/267 (100%)	-0.25	2 (0%) 87 89	40, 58, 84, 121	0
1	D	264/267 (98%)	-0.09	5 (1%) 66 69	51, 71, 91, 124	0
2	B	11/11 (100%)	-0.03	1 (9%) 9 7	57, 63, 144, 162	0
2	E	11/11 (100%)	0.34	0 100 100	73, 87, 157, 160	0
3	C	12/13 (92%)	0.48	2 (16%) 1 1	60, 75, 143, 147	0
3	F	12/13 (92%)	0.27	1 (8%) 11 9	74, 87, 167, 176	0
All	All	577/582 (99%)	-0.14	11 (1%) 66 69	40, 65, 96, 176	0

All (11) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	103	HIS	3.8
1	A	55	HIS	3.4
3	C	26	DG	3.1
1	D	57	ILE	2.9
1	D	293	VAL	2.7
3	F	26	DG	2.7
3	C	25	DG	2.2
2	B	4	DA	2.2
1	A	239	SER	2.1
1	D	240	LEU	2.1
1	D	270	LEU	2.0

6.2 Non-standard residues in protein, DNA, RNA chains (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	5CM	F	20	20/21	0.94	0.15	51,67,82,92	0
3	5CM	C	20	20/21	0.97	0.18	42,51,67,70	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
7	EDO	F	101	4/4	0.93	0.12	70,76,83,87	0
7	EDO	D	403	4/4	0.94	0.14	58,60,65,73	0
4	SO4	A	401	5/5	0.94	0.09	86,91,104,124	0
4	SO4	D	401	5/5	0.96	0.13	81,87,100,118	0
4	SO4	D	402	5/5	0.96	0.07	76,83,100,119	0
5	AKG	A	402	10/10	0.98	0.21	44,54,70,73	0
5	AKG	D	404	10/10	0.99	0.22	56,61,74,82	0
6	MN	D	405	1/1	0.99	0.12	66,66,66,66	0
6	MN	A	403	1/1	1.00	0.17	54,54,54,54	0

6.5 Other polymers (i)

There are no such residues in this entry.