W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 18, 2024 – 05:38 PM JST

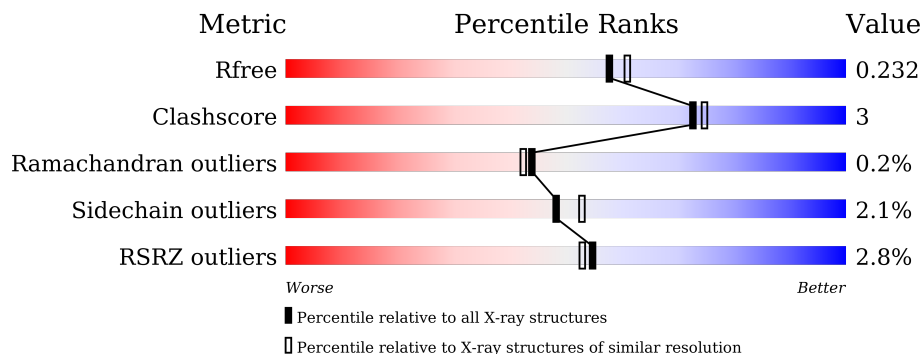|		
PDB ID	:	8XK7
Title	:	binary complex of DNA polymerase SFM4-3 recognizing C2 methyoxy nucleotide
Authors	:	Wen, C.; Liu, H.; Yang, L.; Gong, W.
Deposited on	:	2023-12-22
Resolution	:	2.00 Å(reported)

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 2.00 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		0.232
Clashscore		3
Ramachandran outliers		0.2%
Sidechain outliers		2.1%
RSRZ outliers		2.8%

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$_{free}$	164625	9409 (2.00-2.00)
Clashscore	180529	10737 (2.00-2.00)
Ramachandran outliers	177936	10628 (2.00-2.00)
Sidechain outliers	177891	10627 (2.00-2.00)
RSRZ outliers	164620	9409 (2.00-2.00)

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	554	3% / 88% / 10% ·
2	B	17	41% / 59%
3	C	13	69% / 23% / 8%

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 5330 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 DNA polymerase I, thermostable.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	546	4333	2751	789	781	12	0	2	0

There are 24 discrepancies between the modelled and reference sequences:

| Chain	Residue	Modelled	Actual	Comment	Reference
A	294	MET	LEU	conflict	UNP P19821
A	518	ALA	VAL	conflict	UNP P19821
A	583	SER	ASN	conflict	UNP P19821
A	614	GLU	ILE	conflict	UNP P19821
A	615	GLY	GLU	conflict	UNP P19821
A	655	ASN	ASP	conflict	UNP P19821
A	681	LYS	GLU	conflict	UNP P19821
A	742	GLN	GLU	conflict	UNP P19821
A	747	ARG	MET	conflict	UNP P19821
A	833	ALA	-	expression tag	UNP P19821
A	834	ALA	-	expression tag	UNP P19821
A	835	LYS	-	expression tag	UNP P19821
A	836	LEU	-	expression tag	UNP P19821
A	837	ALA	-	expression tag	UNP P19821
A	838	ALA	-	expression tag	UNP P19821
A	839	ALA	-	expression tag	UNP P19821
A	840	LEU	-	expression tag	UNP P19821
A	841	GLU	-	expression tag	UNP P19821
A	842	HIS	-	expression tag	UNP P19821
A	843	HIS	-	expression tag	UNP P19821
A	844	HIS	-	expression tag	UNP P19821
A	845	HIS	-	expression tag	UNP P19821
A	846	HIS	-	expression tag	UNP P19821
A	847	HIS	-	expression tag	UNP P19821

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	B	17	354	166	69	102	17	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	C	13	265	125	53	75	12	0	0	0

- Molecule 4 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).
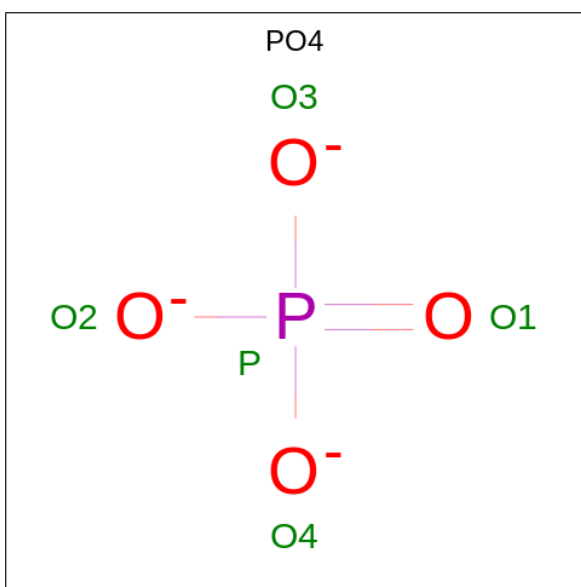
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	P		
4	A	1	5	4	1	0	0
4	A	1	5	4	1	0	0

- Molecule 5 is SULFATE ION (three-letter code: SO4) (formula: O_4S).

SO4

O1

O

O3 O⁻ — S — O⁻ O4

S

O

O2

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

- Molecule 6 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).

PEG

C4

OH O4

O2

O

C3

C1

C2

HO

O1
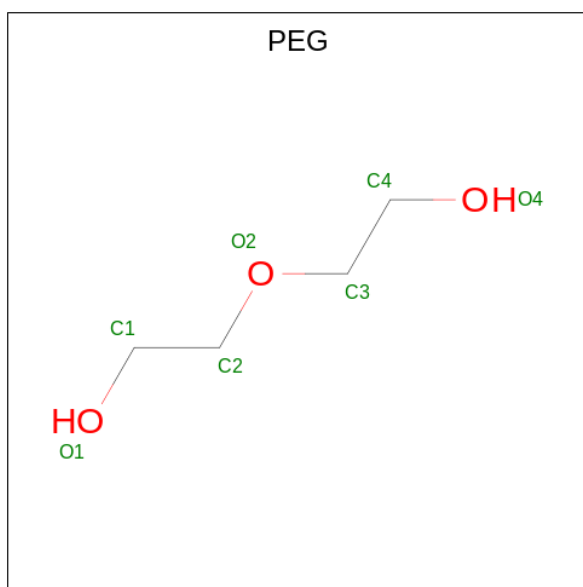
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	A	1	Total 7	C 4	O 3	0	0

WORLDWIDE

PDB

PROTEIN DATA BANK

• Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	272	Total 272	O 272	0	0
7	B	42	Total 42	O 42	0	0
7	C	32	Total 32	O 32	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.
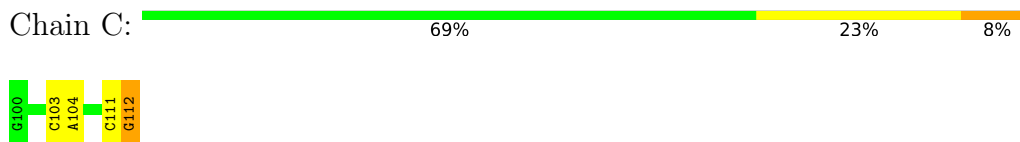
• Molecule 1: DNA polymerase I, thermostable

Chain A:
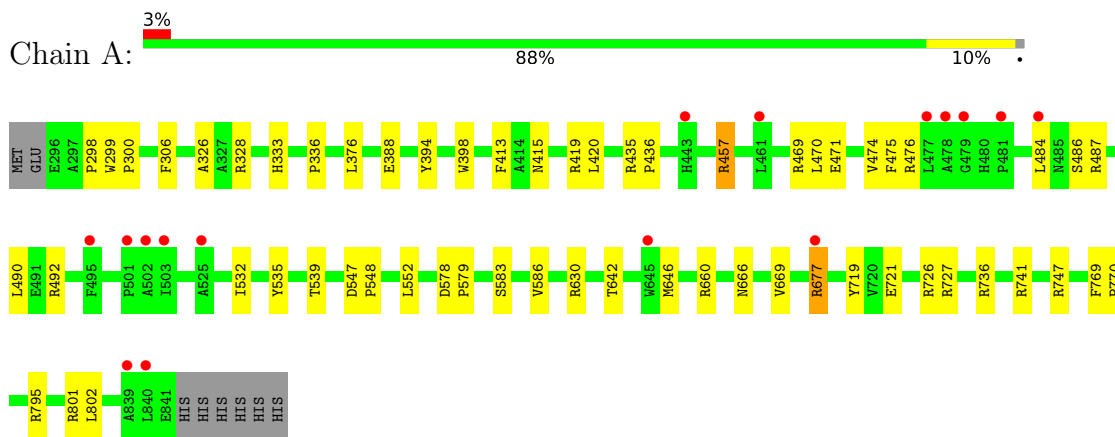
• Molecule 2: DNA(5'-D(*AP*AP*AP*CP*GP*GP*CP*GP*CP*CP*GP*TP*GP*GP*TP*CP*OMG)-3')

Chain B:
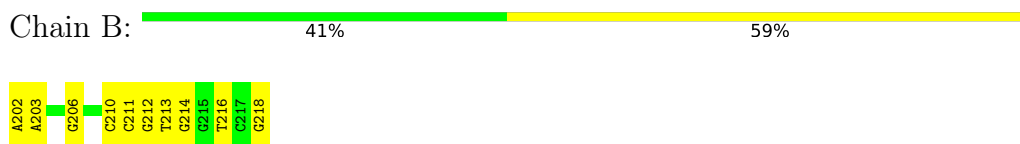
• Molecule 3: DNA(5'-D(*GP*AP*CP*CP*AP*CP*GP*GP*CP*GP*CP*CP*OMG)-3')

Chain C:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	61.61Å 65.63Å 107.38Å 90.00° 103.51° 90.00°	Depositor
Resolution (Å)	24.41 – 2.00 24.41 – 2.00	Depositor EDS
% Data completeness (in resolution range)	98.8 (24.41-2.00) 98.8 (24.41-2.00)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.19 (at 1.99Å)	Xtriage
Refinement program	REFMAC 5.8.0352	Depositor
R, R_{free}	0.192 , 0.231 0.201 , 0.232	Depositor DCC
R_{free} test set	2864 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å^2)	32.3	Xtriage
Anisotropy	0.083	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 30.4	EDS
L-test for twinning[2]	$< |L| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	5330	wwPDB-VP
Average B, all atoms (Å^2)	39.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.82% 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: PEG, SO4, PO4, OMG

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	0.33	0/4424	0.59	0/5991
2	B	0.73	1/370 (0.3%)	1.36	6/569 (1.1%)
3	C	0.70	0/270	1.18	0/414
All	All	0.40	1/5064 (0.0%)	0.73	6/6974 (0.1%)

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	9

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	202	DA	P-O5'	5.23	1.65	1.59

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	203	DA	OP1-P-OP2	6.71	129.66	119.60
2	B	203	DA	O5'-P-OP2	-6.23	100.09	105.70
2	B	206	DG	O4'-C1'-N9	5.91	112.14	108.00
2	B	216	DT	P-O3'-C3'	-5.45	113.16	119.70
2	B	206	DG	O5'-P-OP2	-5.41	100.83	105.70

There are no chirality outliers.

5 of 9 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	457	ARG	Sidechain
1	A	469	ARG	Sidechain
1	A	476	ARG	Sidechain
1	A	630	ARG	Sidechain
1	A	677	ARG	Sidechain

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	4333	0	4390	26	0
2	B	354	0	192	3	0
3	C	265	0	147	3	0
4	A	10	0	0	0	0
5	A	15	0	0	0	0
6	A	7	0	10	1	0
7	A	272	0	0	1	0
7	B	42	0	0	0	0
7	C	32	0	0	0	0
All	All	5330	0	4739	30	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 3.

The worst 5 of 30 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:415[A]:ASN:HD21	1:A:419:ARG:HH21	1.46	0.62
1:A:376:LEU:HD22	1:A:420:LEU:CD1	2.30	0.61
1:A:388:GLU:HG3	1:A:398:TRP:CD1	2.42	0.54
1:A:719:TYR:CD1	1:A:727:ARG:HG3	2.43	0.54
1:A:470:LEU:O	1:A:474:VAL:HG23	2.10	0.52

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	546/554 (99%)	532 (97%)	13 (2%)	1 (0%)	44	42

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	586	VAL

5.3.2   Protein sidechains ⓘ

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	441/449 (98%)	432 (98%)	9 (2%)	50	55

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	677	ARG
1	A	802	LEU
1	A	487	ARG
1	A	492	ARG
1	A	579	PRO

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	333	HIS
1	A	485	ASN
1	A	666	ASN
1	A	750	ASN
1	A	754	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	#$|Z| > 2$
3	OMG	C	112	2,3	18,26,27	0.94	1 (5%)	19,38,41	0.98	2 (10%)
2	OMG	B	218	2	18,26,27	0.98	1 (5%)	19,38,41	0.75	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	OMG	C	112	2,3	-	1/5/27/28	0/3/3/3
2	OMG	B	218	2	-	0/5/27/28	0/3/3/3

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	218	OMG	C5-C6	-2.55	1.42	1.47
3	C	112	OMG	C5-C6	-2.49	1.42	1.47

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	112	OMG	O6-C6-C5	2.57	129.40	124.37
3	C	112	OMG	O2'-C2'-C1'	2.08	113.22	109.09

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	C	112	OMG	C3'-C2'-O2'-CM2

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	C	112	OMG	1	0

5.5 Carbohydrates ⓘ

There are no oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

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	PO4	A	901	-	4,4,4	0.97	0	6,6,6	0.46	0
5	SO4	A	904	-	4,4,4	0.35	0	6,6,6	0.06	0
5	SO4	A	905	-	4,4,4	0.30	0	6,6,6	0.07	0
4	PO4	A	902	-	4,4,4	0.77	0	6,6,6	0.50	0
6	PEG	A	906	-	6,6,6	0.18	0	5,5,5	0.12	0
5	SO4	A	903	-	4,4,4	0.36	0	6,6,6	0.21	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
6	PEG	A	906	-	-	1/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	A	906	PEG	O1-C1-C2-O2

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	A	906	PEG	1	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	546/554 (98%)	0.08	16 (2%)	54	52	14, 36, 60, 82	2 (0%)
2	B	16/17 (94%)	-0.39	0	100	100	26, 37, 56, 59	0
3	C	12/13 (92%)	-0.14	0	100	100	27, 39, 59, 60	0
All	All	574/584 (98%)	0.06	16 (2%)	55	53	14, 36, 60, 82	2 (0%)

The worst 5 of 16 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	645	TRP	3.0
1	A	461	LEU	2.8
1	A	501	PRO	2.7
1	A	479	GLY	2.5
1	A	503	ILE	2.5

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
2	OMG	B	218	24/25	0.89	0.10	45,55,67,70	0
3	OMG	C	112	24/25	0.96	0.06	26,28,33,36	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, 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	SO4	A	904	5/5	0.63	0.13	83,90,94,94	0
5	SO4	A	905	5/5	0.72	0.11	66,74,76,83	0
6	PEG	A	906	7/7	0.85	0.15	52,60,65,69	0
5	SO4	A	903	5/5	0.87	0.17	51,58,60,60	0
4	PO4	A	901	5/5	0.89	0.17	52,53,59,59	0
4	PO4	A	902	5/5	0.95	0.07	47,49,52,53	0

6.5   Other polymers ⓘ

There are no such residues in this entry.