



wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 12, 2024 – 10:35 PM EDT

PDB ID : 4B9V
Title : Structure of the high fidelity DNA polymerase I with extending from an oxidative formamidopyrimidine-dG DNA lesion -dA basepair.
Authors : Gehrke, T.H.; Lischke, U.; Arnold, S.; Schneider, S.; Carell, T.
Deposited on : 2012-09-06
Resolution : 2.00 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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.36.2
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)	:	2.36.2

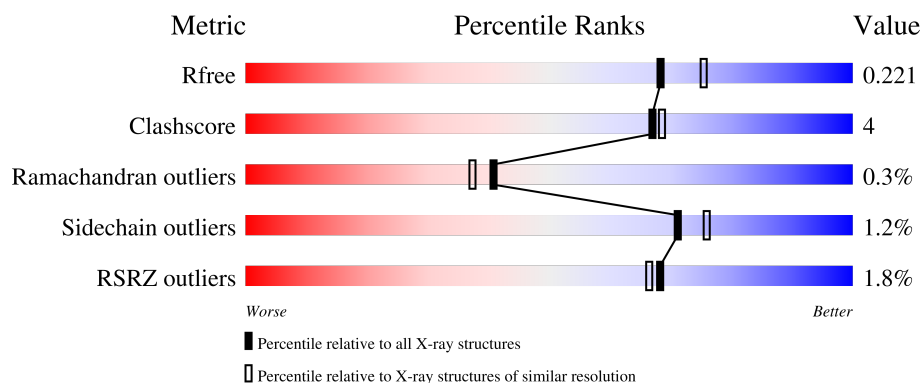
1 Overall quality at a glance

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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (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 $\leq 5\%$. 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	619	<div> <div style="width: 100%; height: 10px; background-color: red; position: relative;"> % </div> <div style="display: flex; justify-content: space-between; margin-top: 5px;"> 87% 6% 6% </div> </div>
2	B	11	<div> <div style="width: 100%; height: 10px; background-color: green; position: relative;"> % </div> <div style="display: flex; justify-content: space-between; margin-top: 5px;"> 73% 27% </div> </div>
3	C	15	<div> <div style="width: 100%; height: 10px; background-color: red; position: relative;"> 13% </div> <div style="display: flex; justify-content: space-between; margin-top: 5px;"> 67% 7% 7% 20% </div> </div>

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	FOX	C	5	-	-	X	-

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 5354 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.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	579	Total	C	N	O	S	0	6	0
			4675	2978	809	870	18			

There are 39 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	258	MET	-	expression tag	UNP E1C9K5
A	259	ALA	-	expression tag	UNP E1C9K5
A	260	SER	-	expression tag	UNP E1C9K5
A	261	TRP	-	expression tag	UNP E1C9K5
A	262	SER	-	expression tag	UNP E1C9K5
A	263	HIS	-	expression tag	UNP E1C9K5
A	264	PRO	-	expression tag	UNP E1C9K5
A	265	GLN	-	expression tag	UNP E1C9K5
A	266	PHE	-	expression tag	UNP E1C9K5
A	267	GLU	-	expression tag	UNP E1C9K5
A	268	LYS	-	expression tag	UNP E1C9K5
A	269	GLY	-	expression tag	UNP E1C9K5
A	270	ALA	-	expression tag	UNP E1C9K5
A	271	SER	-	expression tag	UNP E1C9K5
A	272	THR	-	expression tag	UNP E1C9K5
A	273	SER	-	expression tag	UNP E1C9K5
A	274	LEU	-	expression tag	UNP E1C9K5
A	275	TYR	-	expression tag	UNP E1C9K5
A	276	LYS	-	expression tag	UNP E1C9K5
A	277	LYS	-	expression tag	UNP E1C9K5
A	278	ALA	-	expression tag	UNP E1C9K5
A	279	GLY	-	expression tag	UNP E1C9K5
A	280	SER	-	expression tag	UNP E1C9K5
A	281	ALA	-	expression tag	UNP E1C9K5
A	282	ALA	-	expression tag	UNP E1C9K5
A	283	ALA	-	expression tag	UNP E1C9K5
A	284	VAL	-	expression tag	UNP E1C9K5

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	285	LEU	-	expression tag	UNP E1C9K5
A	286	GLU	-	expression tag	UNP E1C9K5
A	287	GLU	-	expression tag	UNP E1C9K5
A	288	ASN	-	expression tag	UNP E1C9K5
A	289	LEU	-	expression tag	UNP E1C9K5
A	290	TYR	-	expression tag	UNP E1C9K5
A	291	PHE	-	expression tag	UNP E1C9K5
A	292	GLN	-	expression tag	UNP E1C9K5
A	293	GLY	-	expression tag	UNP E1C9K5
A	294	SER	-	expression tag	UNP E1C9K5
A	295	PHE	-	expression tag	UNP E1C9K5
A	296	THR	-	expression tag	UNP E1C9K5

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	11	Total	C	N	O	P	0	0	0
			221	106	38	66	11			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	12	Total	C	N	O	P	0	0	0
			252	119	50	71	12			

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



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

- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	1	Total	Mg	0	0
			1	1		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	173	Total	O	0	0
			173	173		
6	B	1	Total	O	0	0
			1	1		

Continued on next page...

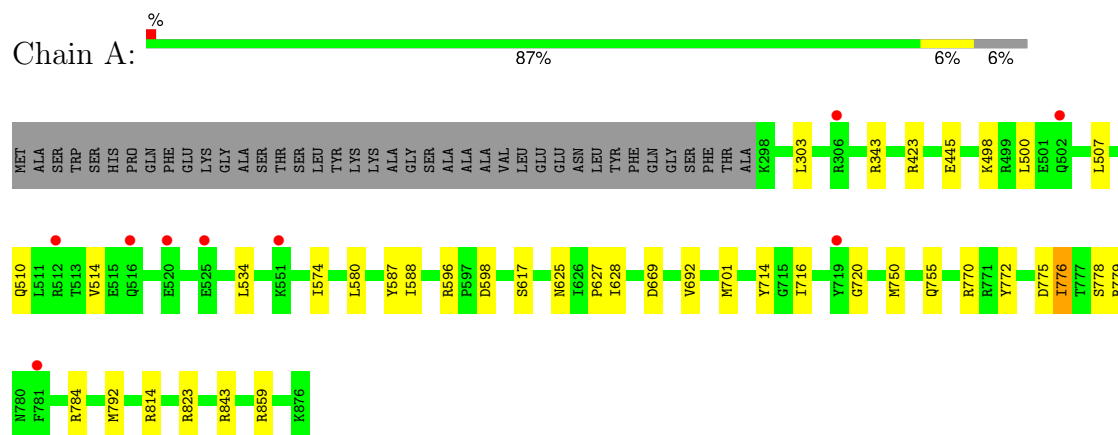
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	C	1	Total	O	0	0
			1	1		

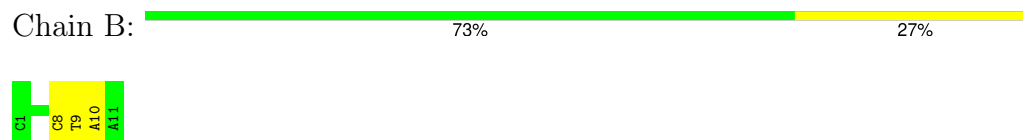
3 Residue-property plots [i](#)

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.

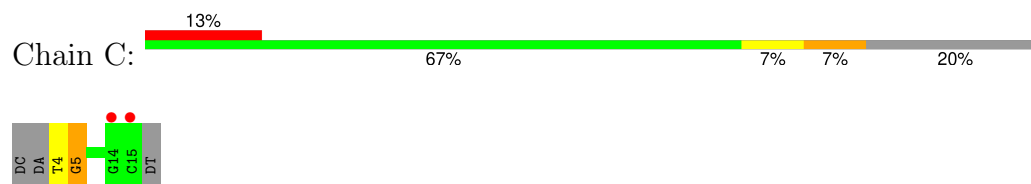
- Molecule 1: DNA POLYMERASE



- Molecule 2: 5'-D(*CP*CP*TP*GP*AP*CP*TP*CP*TP*AP*AP)-3'



- Molecule 3: 5'-D(*CP*AP*TP*FOXP*AP*GP*AP*GP*TP*CP*AP*GP*GP*CP*TP)-3'



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	87.90Å 94.03Å 105.34Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.02 – 2.00 47.02 – 1.97	Depositor EDS
% Data completeness (in resolution range)	99.7 (47.02-2.00) 99.2 (47.02-1.97)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.29 (at 1.97Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, R_{free}	0.187 , 0.215 0.191 , 0.221	Depositor DCC
R_{free} test set	2964 reflections (4.81%)	wwPDB-VP
Wilson B-factor (Å ²)	26.6	Xtriage
Anisotropy	0.542	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.38 , 43.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	5354	wwPDB-VP
Average B, all atoms (Å ²)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.62% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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: FOX, MG, SO4

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	0.74	0/4776	0.80	9/6454 (0.1%)
2	B	0.26	0/246	0.77	0/376
3	C	0.35	0/256	0.86	0/391
All	All	0.71	0/5278	0.81	9/7221 (0.1%)

There are no bond length outliers.

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	843	ARG	NE-CZ-NH1	7.14	123.87	120.30
1	A	859	ARG	NE-CZ-NH2	-7.14	116.73	120.30
1	A	859	ARG	NE-CZ-NH1	6.51	123.56	120.30
1	A	843	ARG	NE-CZ-NH2	-6.30	117.15	120.30
1	A	423	ARG	NE-CZ-NH2	-5.91	117.34	120.30

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.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	4675	0	4755	20	0
2	B	221	0	125	8	0
3	C	252	0	139	18	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	30	0	0	2	0
5	A	1	0	0	0	0
6	A	173	0	0	2	0
6	B	1	0	0	0	0
6	C	1	0	0	0	0
All	All	5354	0	5019	39	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 39 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:B:10:DA:H2	3:C:5:FOX:H22	1.23	0.82
3:C:4:DT:H2'	3:C:5:FOX:N9	1.95	0.81
3:C:5:FOX:O6	3:C:5:FOX:H8	1.85	0.76
1:A:750[B]:MET:HE3	1:A:792:MET:SD	2.35	0.67
3:C:4:DT:H2'	3:C:5:FOX:H9	1.62	0.63

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	583/619 (94%)	572 (98%)	9 (2%)	2 (0%)	41 37

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	778	SER
1	A	628	ILE

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	501/526 (95%)	495 (99%)	6 (1%)	71 76

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	617	SER
1	A	755	GLN
1	A	776	ILE
1	A	445	GLU
1	A	303	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	356	GLN
1	A	529	ASN
1	A	579	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](#)

1 non-standard protein/DNA/RNA residue is 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	FOX	C	5	2,3	19,24,25	0.83	1 (5%)	17,33,36	1.45	3 (17%)

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	FOX	C	5	2,3	-	4/7/24/25	0/2/2/2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	5	FOX	C6-N1	-2.28	1.34	1.38

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	5	FOX	O8-C8-N7	-3.02	120.00	124.59
3	C	5	FOX	C6'-C1'-N9	-2.69	107.80	112.67
3	C	5	FOX	C6'-C4'-C5'	-2.08	108.66	112.78

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	C	5	FOX	C6'-C4'-C5'-O5'
3	C	5	FOX	C3'-C4'-C5'-O5'
3	C	5	FOX	C2'-C1'-N9-C4
3	C	5	FOX	N3-C4-N9-C1'

There are no ring outliers.

1 monomer is involved in 15 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	C	5	FOX	15	0

5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 7 ligands modelled in this entry, 1 is monoatomic - leaving 6 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	$\# Z > 2$
4	SO4	A	1878	-	4,4,4	0.24	0	6,6,6	1.00	0
4	SO4	A	1880	-	4,4,4	0.35	0	6,6,6	0.35	0
4	SO4	A	1877	-	4,4,4	0.52	0	6,6,6	0.28	0
4	SO4	A	1881	-	4,4,4	0.31	0	6,6,6	0.31	0
4	SO4	A	1882	-	4,4,4	0.79	0	6,6,6	0.48	0
4	SO4	A	1879	-	4,4,4	0.77	0	6,6,6	0.70	0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1880	SO4	1	0
4	A	1881	SO4	1	0

5.7 Other polymers [i](#)

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 [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	579/619 (93%)	-0.08	9 (1%) 72 70	15, 27, 50, 72	0
2	B	11/11 (100%)	0.17	0 100 100	38, 47, 67, 85	0
3	C	11/15 (73%)	0.64	2 (18%) 1 1	35, 41, 79, 94	0
All	All	601/645 (93%)	-0.06	11 (1%) 68 66	15, 28, 52, 94	0

The worst 5 of 11 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	15	DC	3.7
1	A	520	GLU	2.9
1	A	719	TYR	2.5
1	A	502	GLN	2.3
1	A	516	GLN	2.3

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	FOX	C	5	23/24	0.84	0.20	55,66,71,74	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(\AA^2)	Q<0.9
4	SO4	A	1879	5/5	0.92	0.22	49,51,56,66	0
4	SO4	A	1882	5/5	0.92	0.33	25,29,31,34	5
4	SO4	A	1880	5/5	0.93	0.41	34,36,38,41	5
4	SO4	A	1881	5/5	0.95	0.13	36,37,38,41	5
5	MG	A	1883	1/1	0.95	0.12	43,43,43,43	0
4	SO4	A	1878	5/5	0.97	0.12	38,40,46,50	0
4	SO4	A	1877	5/5	0.98	0.07	48,49,49,52	0

6.5 Other polymers [i](#)

There are no such residues in this entry.