



wwPDB X-ray Structure Validation Summary Report ⓘ

May 14, 2020 – 08:52 am BST

PDB ID : 5D4B
Title : Structural Basis for a New Templated Activity by Terminal Deoxynucleotidyl Transferase: Implications for V(D)J Recombination
Authors : Loc'h, J.; Rosario, S.; Delarue, M.
Deposited on : 2015-08-07
Resolution : 2.66 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 following versions of software and data (see [references ⓘ](#)) 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 : 2.11
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) : 2.11

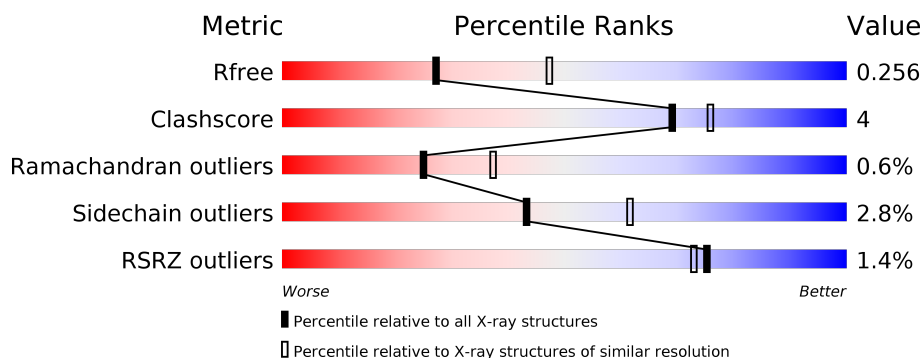
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.66 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	1332 (2.68-2.64)
Clashscore	141614	1374 (2.68-2.64)
Ramachandran outliers	138981	1349 (2.68-2.64)
Sidechain outliers	138945	1349 (2.68-2.64)
RSRZ outliers	127900	1318 (2.68-2.64)

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 on 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	400	<div> <div>%</div> <div> <div></div> <div>80%</div> <div>7%</div> <div>•</div> <div>13%</div> </div> </div>
1	B	400	<div> <div>2%</div> <div> <div></div> <div>82%</div> <div>5%</div> <div>•</div> <div>13%</div> </div> </div>
2	C	6	<div> <div></div> <div>100%</div> </div>
2	D	6	<div> <div></div> <div>100%</div> </div>
2	E	6	<div> <div></div> <div>83%</div> <div>17%</div> </div>
3	F	7	<div> <div></div> <div>43%</div> <div>57%</div> </div>

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 5871 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 Terminal deoxynucleotidyltransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	350	Total	C	N	O	S	0	0	0
			2624	1687	446	476	15			
1	B	349	Total	C	N	O	S	0	0	0
			2564	1650	431	466	17			

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	111	MET	-	initiating methionine	UNP Q3UZ80
A	112	GLY	-	expression tag	UNP Q3UZ80
A	113	SER	-	expression tag	UNP Q3UZ80
A	114	SER	-	expression tag	UNP Q3UZ80
A	115	HIS	-	expression tag	UNP Q3UZ80
A	116	HIS	-	expression tag	UNP Q3UZ80
A	117	HIS	-	expression tag	UNP Q3UZ80
A	118	HIS	-	expression tag	UNP Q3UZ80
A	119	HIS	-	expression tag	UNP Q3UZ80
A	120	HIS	-	expression tag	UNP Q3UZ80
A	121	SER	-	expression tag	UNP Q3UZ80
A	122	SER	-	expression tag	UNP Q3UZ80
A	123	GLY	-	expression tag	UNP Q3UZ80
A	124	LEU	-	expression tag	UNP Q3UZ80
A	125	VAL	-	expression tag	UNP Q3UZ80
A	126	PRO	-	expression tag	UNP Q3UZ80
A	127	ARG	-	expression tag	UNP Q3UZ80
A	128	GLY	-	expression tag	UNP Q3UZ80
A	129	SER	-	expression tag	UNP Q3UZ80
A	130	HIS	-	expression tag	UNP Q3UZ80
A	131	MET	-	expression tag	UNP Q3UZ80
B	111	MET	-	initiating methionine	UNP Q3UZ80
B	112	GLY	-	expression tag	UNP Q3UZ80
B	113	SER	-	expression tag	UNP Q3UZ80
B	114	SER	-	expression tag	UNP Q3UZ80

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	115	HIS	-	expression tag	UNP Q3UZ80
B	116	HIS	-	expression tag	UNP Q3UZ80
B	117	HIS	-	expression tag	UNP Q3UZ80
B	118	HIS	-	expression tag	UNP Q3UZ80
B	119	HIS	-	expression tag	UNP Q3UZ80
B	120	HIS	-	expression tag	UNP Q3UZ80
B	121	SER	-	expression tag	UNP Q3UZ80
B	122	SER	-	expression tag	UNP Q3UZ80
B	123	GLY	-	expression tag	UNP Q3UZ80
B	124	LEU	-	expression tag	UNP Q3UZ80
B	125	VAL	-	expression tag	UNP Q3UZ80
B	126	PRO	-	expression tag	UNP Q3UZ80
B	127	ARG	-	expression tag	UNP Q3UZ80
B	128	GLY	-	expression tag	UNP Q3UZ80
B	129	SER	-	expression tag	UNP Q3UZ80
B	130	HIS	-	expression tag	UNP Q3UZ80
B	131	MET	-	expression tag	UNP Q3UZ80

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	6	Total	C	N	O	P	0	0	0
			120	59	28	28	5			
2	D	6	Total	C	N	O	P	0	0	0
			120	59	28	28	5			
2	E	6	Total	C	N	O	P	0	0	0
			113	55	26	27	5			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	F	7	Total	C	N	O	P	0	0	0
			131	65	16	44	6			

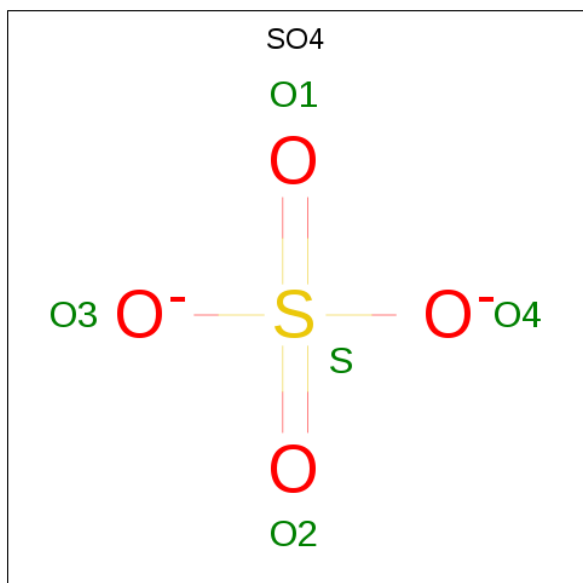
- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Mg	0	0
			1	1		
4	A	1	Total	Mg	0	0
			1	1		

- Molecule 5 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	1	Total	Na	0	0
			1	1		
5	A	1	Total	Na	0	0
			1	1		

- Molecule 6 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



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

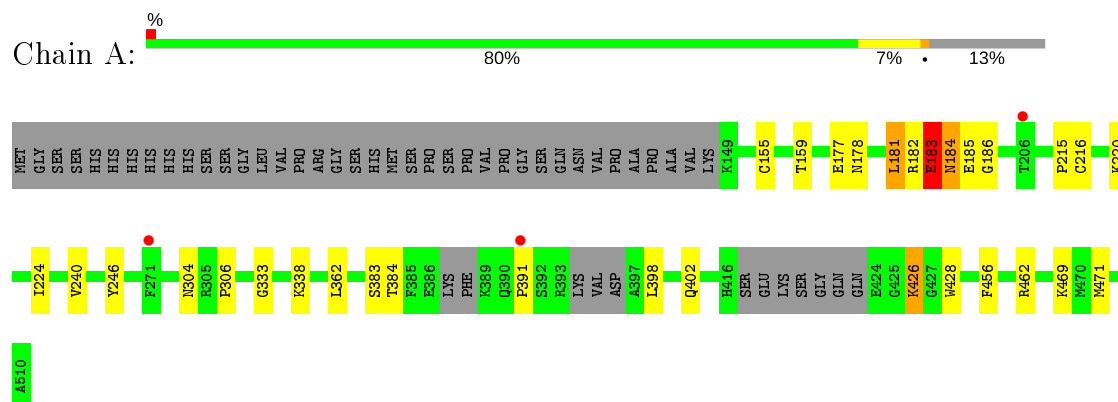
- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	64	Total 64	O 64	0	0
7	B	65	Total 65	O 65	0	0
7	C	6	Total 6	O 6	0	0
7	D	8	Total 8	O 8	0	0
7	E	5	Total 5	O 5	0	0
7	F	7	Total 7	O 7	0	0

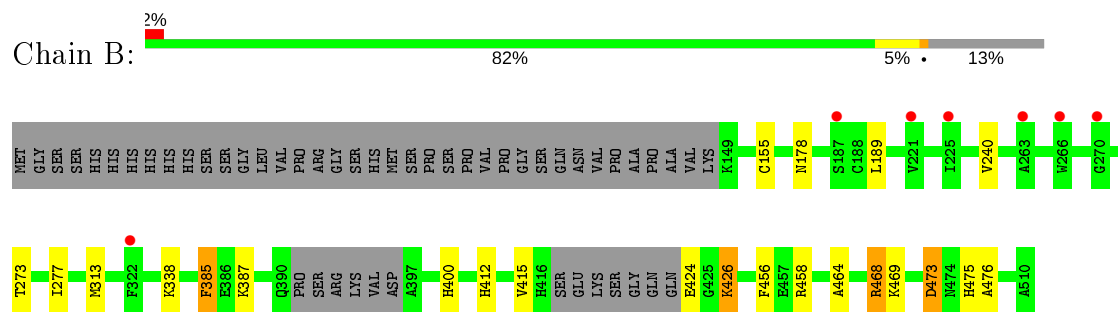
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA 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: Terminal deoxynucleotidyltransferase



- Molecule 1: Terminal deoxynucleotidyltransferase



- Molecule 2: DNA (5'-D(*AP*AP*AP*AP*AP*C)-3')




There are no outlier residues recorded for this chain.

- Molecule 2: DNA (5'-D(*AP*AP*AP*AP*AP*C)-3')



There are no outlier residues recorded for this chain.

- Molecule 2: DNA (5'-D(*AP*AP*AP*AP*AP*C)-3')

Chain E:  83% 17%



- Molecule 3: DNA (5'-D(*TP*TP*TP*TP*TP*GP*G)-3')

Chain F:  43% 57%



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	56.73 Å 71.45 Å 198.17 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.23 – 2.66 49.23 – 2.66	Depositor EDS
% Data completeness (in resolution range)	96.3 (49.23-2.66) 96.0 (49.23-2.66)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	0.11	Depositor
$\langle I/\sigma(I) \rangle$ ¹	0.94 (at 2.65 Å)	Xtriage
Refinement program	BUSTER 2.11.4	Depositor
R, R_{free}	0.207 , 0.247 0.215 , 0.256	Depositor DCC
R_{free} test set	1155 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	64.3	Xtriage
Anisotropy	0.238	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 75.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	5871	wwPDB-VP
Average B, all atoms (Å ²)	67.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 38.42 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 3.6945e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹ Intensities estimated from amplitudes.

² Theoretical 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 [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: NA, 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	0.48	0/2676	0.66	0/3619
1	B	0.48	0/2616	0.68	0/3552
2	C	1.04	0/136	0.85	0/207
2	D	1.15	0/136	0.88	0/207
2	E	0.85	0/128	0.81	0/195
3	F	0.95	0/144	1.03	0/221
All	All	0.55	0/5836	0.70	0/8001

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	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	183	GLU	Mainchain

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	2624	0	2447	21	0
1	B	2564	0	2315	18	0
2	C	120	0	66	0	0
2	D	120	0	66	0	0
2	E	113	0	62	1	0
3	F	131	0	80	4	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
5	A	1	0	0	0	0
5	B	1	0	0	0	0
6	A	10	0	0	0	0
6	B	20	0	0	1	0
6	C	10	0	0	0	0
7	A	64	0	0	0	0
7	B	65	0	0	0	0
7	C	6	0	0	0	0
7	D	8	0	0	0	0
7	E	5	0	0	0	0
7	F	7	0	0	1	0
All	All	5871	0	5036	40	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 40 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:464:ALA:O	1:B:468:ARG:HB2	1.72	0.88
1:A:391:PRO:HB2	1:A:471:MET:HE2	1.54	0.85
1:A:182:ARG:O	1:A:183:GLU:O	2.01	0.77
3:F:4:DT:H2"	3:F:5:DT:H5"	1.74	0.68
1:A:362:LEU:HD12	1:A:402:GLN:HG2	1.78	0.66

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	342/400 (86%)	330 (96%)	9 (3%)	3 (1%)	17	26
1	B	343/400 (86%)	329 (96%)	13 (4%)	1 (0%)	41	56
All	All	685/800 (86%)	659 (96%)	22 (3%)	4 (1%)	25	37

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	183	GLU
1	A	185	GLU
1	A	184	ASN
1	B	385	PHE

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	245/350 (70%)	239 (98%)	6 (2%)	49	67
1	B	227/350 (65%)	220 (97%)	7 (3%)	40	57
All	All	472/700 (67%)	459 (97%)	13 (3%)	43	61

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	426	LYS
1	B	313	MET

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	458	ARG
1	A	398	LEU
1	B	426	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	166	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry ⓘ

Of 12 ligands modelled in this entry, 4 are monoatomic - leaving 8 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	$\# Z > 2$
6	SO4	A	604	-	4,4,4	0.24	0	6,6,6	0.16	0
6	SO4	B	605	-	4,4,4	0.20	0	6,6,6	0.13	0
6	SO4	A	603	-	4,4,4	0.21	0	6,6,6	0.22	0
6	SO4	B	603	4	4,4,4	0.22	0	6,6,6	0.21	0
6	SO4	C	402	-	4,4,4	0.14	0	6,6,6	0.11	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
6	SO4	B	604	-	4,4,4	0.29	0	6,6,6	0.26	0
6	SO4	C	401	4	4,4,4	0.29	0	6,6,6	0.28	0
6	SO4	B	606	-	4,4,4	0.15	0	6,6,6	0.21	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.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	B	605	SO4	1	0

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	350/400 (87%)	0.01	3 (0%) 84 83	46, 63, 87, 117	0
1	B	349/400 (87%)	0.03	7 (2%) 65 60	45, 66, 101, 122	0
2	C	6/6 (100%)	-0.01	0 100 100	81, 84, 104, 122	0
2	D	6/6 (100%)	-0.39	0 100 100	51, 64, 89, 103	0
2	E	6/6 (100%)	0.04	0 100 100	75, 87, 108, 118	0
3	F	7/7 (100%)	-0.41	0 100 100	61, 66, 72, 75	0
All	All	724/825 (87%)	0.01	10 (1%) 75 73	45, 64, 99, 122	0

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	221	VAL	3.2
1	A	271	PHE	2.7
1	B	263	ALA	2.4
1	A	391	PRO	2.3
1	B	270	GLY	2.3

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no carbohydrates 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
6	SO4	B	606	5/5	0.78	0.30	132,132,133,134	0
5	NA	B	602	1/1	0.82	0.11	67,67,67,67	0
6	SO4	A	603	5/5	0.82	0.25	123,124,125,125	0
6	SO4	C	402	5/5	0.83	0.15	161,161,161,161	0
6	SO4	B	604	5/5	0.91	0.17	108,110,110,111	0
5	NA	A	602	1/1	0.93	0.12	66,66,66,66	0
4	MG	A	601	1/1	0.94	0.17	67,67,67,67	0
6	SO4	B	605	5/5	0.95	0.12	95,96,99,100	0
6	SO4	A	604	5/5	0.95	0.23	101,101,101,101	0
4	MG	B	601	1/1	0.96	0.15	50,50,50,50	0
6	SO4	B	603	5/5	0.97	0.15	69,72,74,76	0
6	SO4	C	401	5/5	0.97	0.12	66,71,73,75	0

6.5 Other polymers [i](#)

There are no such residues in this entry.