Full wwPDB X-ray Structure Validation Report (i)

Oct 11, 2023 – 07:26 AM EDT

PDB ID	:	5HT2
Title	:	Mouse Tdp2 reaction product (5'-phosphorylated DNA)-Mg2+ complex with 1-N6-etheno-adenine
Authors	:	Schellenberg, M.J.; Vilas, C.K.; Williams, R.S.
Deposited on	:	2016-01-26
Resolution	:	1.43 Å(reported)

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

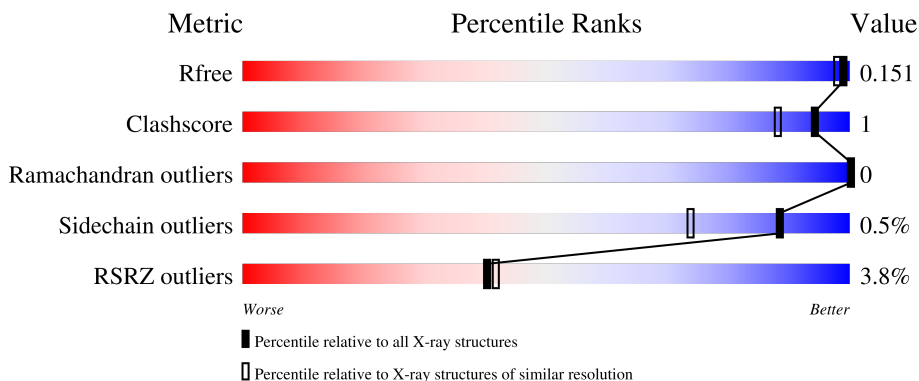|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (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:
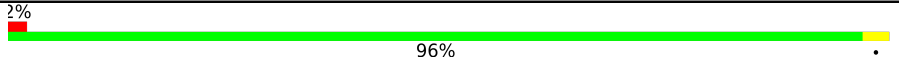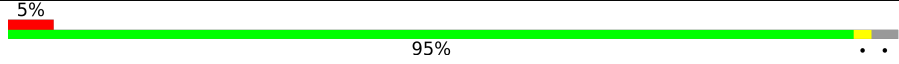*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.43 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		0.151
Clashscore		1
Ramachandran outliers		0
Sidechain outliers		0.5%
RSRZ outliers		3.8%

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$_{free}$	130704	2021 (1.46-1.42)
Clashscore	141614	2086 (1.46-1.42)
Ramachandran outliers	138981	2047 (1.46-1.42)
Sidechain outliers	138945	2047 (1.46-1.42)
RSRZ outliers	127900	1993 (1.46-1.42)

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	256	2% 96% ·
1	B	256	5% 95% · ·
2	C	9	89% 11%
2	D	9	67% 33%

2   Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 9749 atoms, of which 4446 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 Tyrosyl-DNA phosphodiesterase 2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	255	Total 4259	C 1348	H 2149	N 360	O 386	S 16	0	18	0
1	B	249	Total 4164	C 1320	H 2098	N 354	O 375	S 17	0	18	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	115	SER	-	expression tag	UNP Q9JJX7
A	116	LYS	-	expression tag	UNP Q9JJX7
A	117	GLY	-	expression tag	UNP Q9JJX7
B	115	SER	-	expression tag	UNP Q9JJX7
B	116	LYS	-	expression tag	UNP Q9JJX7
B	117	GLY	-	expression tag	UNP Q9JJX7

- Molecule 2 is a DNA chain called DNA.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	C	9	Total 279	C 90	H 91	N 35	O 54	P 9	0	0	0
2	D	9	Total 279	C 90	H 91	N 35	O 54	P 9	0	0	0

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total 1	Mg 1	0	0
3	B	2	Total 2	Mg 2	0	1

- Molecule 4 is 4-(2-HYDROXYETHYL)-1-PIPERAZINE ETHANESULFONIC ACID

(three-letter code: EPE) (formula: $C_8H_{18}N_2O_4S$).
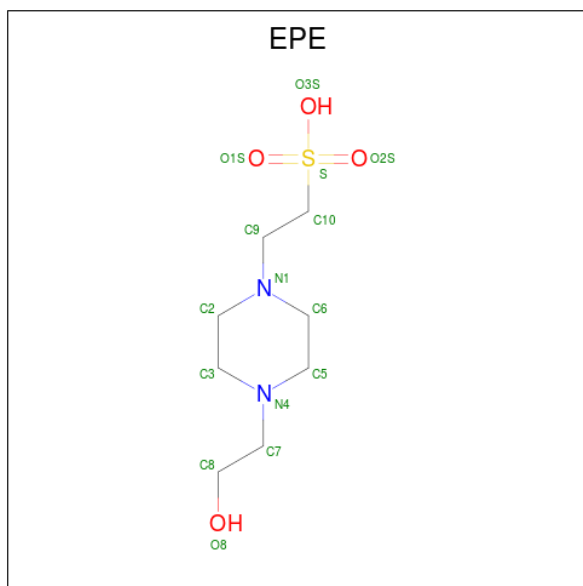
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
			Total	C	H	N	O	S		
4	B	1	32	8	17	2	4	1	0	0

- Molecule 5 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).
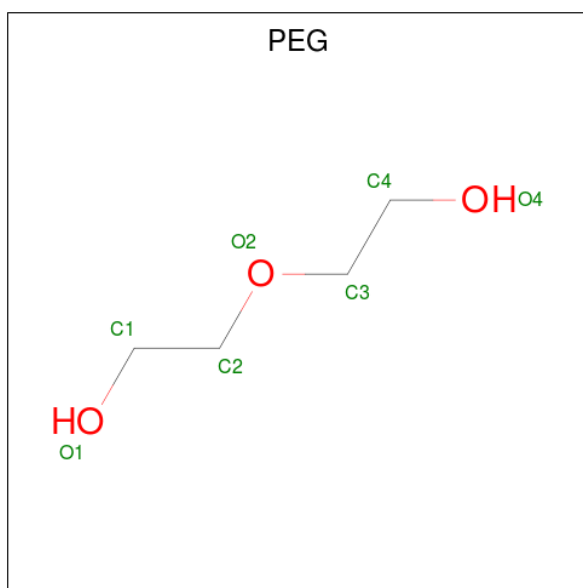
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
5	B	1	7	4	3	0	0

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	362	Total 373	O 373	0	19
6	B	274	Total 284	O 284	0	13
6	C	36	Total 36	O 36	0	1
6	D	30	Total 33	O 33	0	3

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: Tyrosyl-DNA phosphodiesterase 2
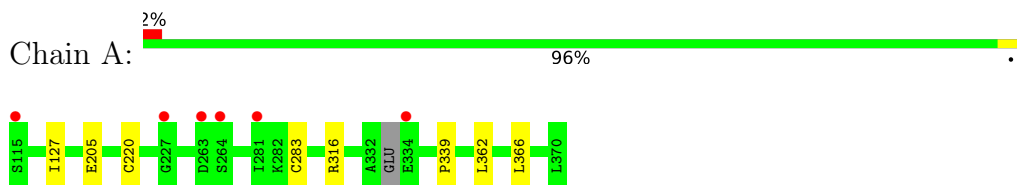
Chain A:

- Molecule 1: Tyrosyl-DNA phosphodiesterase 2

Chain B:

- Molecule 2: DNA

Chain C:

- Molecule 2: DNA

Chain D:
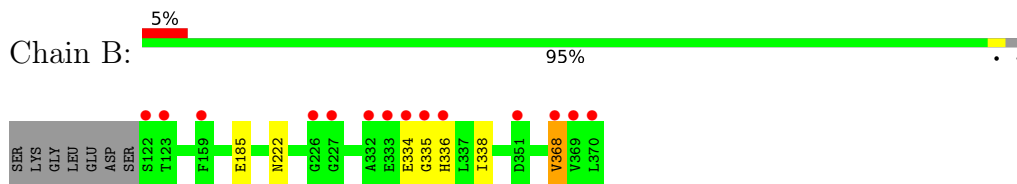
4 Data and refinement statistics ⓘ

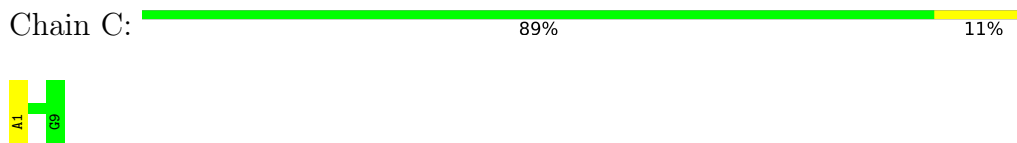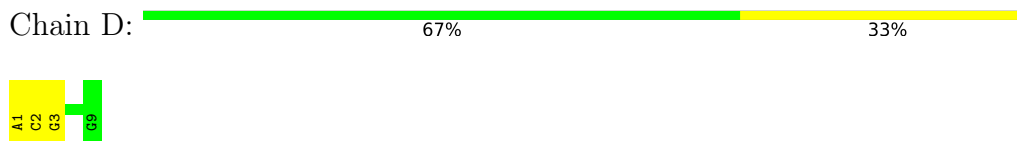| Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	54.88Å 68.60Å 167.25Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.93 – 1.43 31.74 – 1.43	Depositor EDS
% Data completeness (in resolution range)	94.2 (29.93-1.43) 91.0 (31.74-1.43)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.50 (at 1.43Å)	Xtriage
Refinement program	PHENIX 1.8.4_1496	Depositor
R, R_{free}	0.116 , 0.150 0.118 , 0.151	Depositor DCC
R_{free} test set	5612 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	16.4	Xtriage
Anisotropy	0.564	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.44 , 62.5	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.98	EDS
Total number of atoms	9749	wwPDB-VP
Average B, all atoms (Å^2)	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.91% 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: PEG, EPE, EDA, MG

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	0.37	0/2208	0.60	0/2984
1	B	0.34	0/2163	0.58	0/2928
2	C	0.63	0/183	0.89	0/280
2	D	0.60	0/183	0.91	0/280
All	All	0.38	0/4737	0.62	0/6472

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 ⓘ

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	2110	2149	2148	4	0
1	B	2066	2098	2077	7	0
2	C	188	91	102	0	0
2	D	188	91	102	1	0
3	A	1	0	0	0	0
3	B	2	0	0	0	0
4	B	15	17	17	0	0
5	B	7	0	10	0	0
6	A	373	0	0	1	1

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	B	284	0	0	3	1
6	C	36	0	0	0	0
6	D	33	0	0	0	0
All	All	5303	4446	4456	12	1

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 (12) 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:339:PRO:HA	1:A:366:LEU:HD23	1.83	0.61
1:B:334:GLU:N	6:B:504:HOH:O	2.31	0.57
1:A:205[B]:GLU:HG3	1:A:220:CYS:HB2	1.87	0.57
1:B:336:HIS:O	1:B:368[A]:VAL:HA	2.11	0.51
1:A:316:ARG:NH1	6:A:511:HOH:O	2.43	0.50
1:B:335:GLY:HA2	1:B:338[B]:ILE:HD11	1.94	0.49
1:B:222[A]:ASN:ND2	6:B:512:HOH:O	2.45	0.49
1:B:336:HIS:O	1:B:368[B]:VAL:HA	2.13	0.48
2:D:2:DC:H2'	2:D:3:DG:C8	2.52	0.45
1:B:185:GLU:OE2	6:B:501:HOH:O	2.21	0.43
1:A:127:ILE:HA	1:A:362:LEU:O	2.20	0.41

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
6:A:851:HOH:O	6:B:692:HOH:O[1_565]	2.11	0.09

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	269/256 (105%)	267 (99%)	2 (1%)	0	100	100
1	B	264/256 (103%)	261 (99%)	3 (1%)	0	100	100
All	All	533/512 (104%)	528 (99%)	5 (1%)	0	100	100

There are no Ramachandran outliers to report.

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	241/224 (108%)	240 (100%)	1 (0%)	91	80
1	B	236/224 (105%)	234 (99%)	2 (1%)	81	61
All	All	477/448 (106%)	474 (99%)	3 (1%)	88	68

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	283	CYS
1	B	368[A]	VAL
1	B	368[B]	VAL

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

Mol	Chain	Res	Type
1	B	302	HIS

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	EDA	C	1	2,3	22,27,27	1.34	1 (4%)	23,41,41	1.23	2 (8%)
2	EDA	D	1	2,3	22,27,27	1.22	1 (4%)	23,41,41	1.16	2 (8%)

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	EDA	C	1	2,3	-	0/6/22/22	0/4/4/4
2	EDA	D	1	2,3	-	0/6/22/22	0/4/4/4

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	1	EDA	C5-C6	5.21	1.44	1.38
2	D	1	EDA	C5-C6	4.72	1.43	1.38

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	1	EDA	C2-N3-C4	-3.46	112.26	116.58
2	D	1	EDA	C2-N3-C4	-3.20	112.58	116.58
2	D	1	EDA	C4-C5-N7	2.68	112.19	109.40
2	C	1	EDA	C4-C5-N7	2.62	112.13	109.40

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 5 ligands modelled in this entry, 3 are monoatomic - leaving 2 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	PEG	B	403	-	6,6,6	0.57	0	5,5,5	0.35	0
4	EPE	B	402	-	15,15,15	0.79	0	18,20,20	0.61	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	PEG	B	403	-	-	2/4/4/4	-
4	EPE	B	402	-	-	3/9/19/19	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	B	402	EPE	C10-C9-N1-C6
4	B	402	EPE	C8-C7-N4-C3
5	B	403	PEG	O2-C3-C4-O4
4	B	402	EPE	C10-C9-N1-C2
5	B	403	PEG	C1-C2-O2-C3

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(Å^2)	Q<0.9
1	A	255/256 (99%)	-0.30	6 (2%)	59	60	12, 20, 46, 72	0
1	B	249/256 (97%)	-0.05	14 (5%)	24	24	16, 25, 44, 69	0
2	C	8/9 (88%)	-0.14	0	100	100	26, 37, 48, 52	0
2	D	8/9 (88%)	0.14	0	100	100	25, 41, 44, 47	0
All	All	520/530 (98%)	-0.17	20 (3%)	40	42	12, 23, 46, 72	0

All (20) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	370	LEU	6.1
1	A	334	GLU	5.5
1	B	369[A]	VAL	5.4
1	B	123[A]	THR	4.6
1	B	368[A]	VAL	4.3
1	B	122[A]	SER	3.9
1	A	115	SER	3.7
1	B	226	GLY	3.5
1	B	227	GLY	3.4
1	B	336	HIS	3.0
1	A	264	SER	3.0
1	A	281	ILE	2.9
1	B	335	GLY	2.6
1	A	263	ASP	2.5
1	B	334	GLU	2.4
1	B	333	GLU	2.4
1	A	227	GLY	2.3
1	B	159	PHE	2.3
1	B	351	ASP	2.3
1	B	332	ALA	2.0

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	EDA	C	1	24/24	0.96	0.08	18,21,27,31	0
2	EDA	D	1	24/24	0.99	0.07	14,19,25,27	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	PEG	B	403	7/7	0.86	0.10	50,56,64,66	0
3	MG	B	404[B]	1/1	0.93	0.61	30,30,30,30	1
4	EPE	B	402	15/15	0.96	0.21	35,66,83,84	0
3	MG	B	401	1/1	1.00	0.07	18,18,18,18	0
3	MG	A	401	1/1	1.00	0.07	14,14,14,14	0

6.5 Other polymers ⓘ

There are no such residues in this entry.