WORLDWIDE wwPDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Dec 16, 2023 – 08:05 am GMT

PDB ID	:	4A09
Title	:	Structure of hsDDB1-drDDB2 bound to a 15 bp CPD-duplex (purine at D-1 position) at 3.1 A resolution (CPD 2)
Authors	:	Scrima, A.; Fischer, E.S.; Iwai, S.; Gut, H.; Thoma, N.H.
Deposited on	:	2011-09-08
Resolution	:	3.10 Å(reported)

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

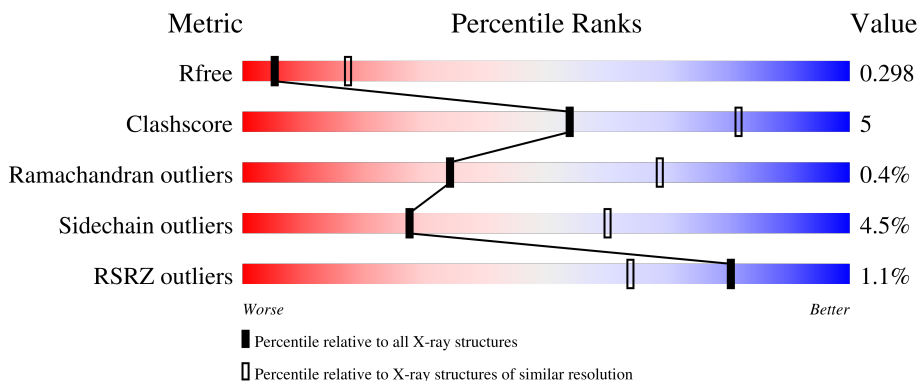|		
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 3.10 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		0.298
Clashscore		5
Ramachandran outliers		0.4%
Sidechain outliers		4.5%
RSRZ outliers		1.1%

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	1094 (3.10-3.10)
Clashscore	141614	1184 (3.10-3.10)
Ramachandran outliers	138981	1141 (3.10-3.10)
Sidechain outliers	138945	1141 (3.10-3.10)
RSRZ outliers	127900	1067 (3.10-3.10)

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	1159	% 81% 14% • 5%
2	B	382	75% 16% • 7%
3	G	14	50% 36% 14%
4	H	16	6% 62% 25% 6% 6%

2 Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 12113 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 DAMAGE-BINDING PROTEIN 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	1105	8646	5484	1453	1662	47	0	0	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-18	MET	-	expression tag	UNP Q16531
A	-17	HIS	-	expression tag	UNP Q16531
A	-16	HIS	-	expression tag	UNP Q16531
A	-15	HIS	-	expression tag	UNP Q16531
A	-14	HIS	-	expression tag	UNP Q16531
A	-13	HIS	-	expression tag	UNP Q16531
A	-12	HIS	-	expression tag	UNP Q16531
A	-11	VAL	-	expression tag	UNP Q16531
A	-10	ASP	-	expression tag	UNP Q16531
A	-9	GLU	-	expression tag	UNP Q16531
A	-8	ASN	-	expression tag	UNP Q16531
A	-7	LEU	-	expression tag	UNP Q16531
A	-6	TYR	-	expression tag	UNP Q16531
A	-5	PHE	-	expression tag	UNP Q16531
A	-4	GLN	-	expression tag	UNP Q16531
A	-3	GLY	-	expression tag	UNP Q16531
A	-2	GLY	-	expression tag	UNP Q16531
A	-1	GLY	-	expression tag	UNP Q16531
A	0	ARG	-	expression tag	UNP Q16531
A	194	ALA	GLU	engineered mutation	UNP Q16531

- Molecule 2 is a protein called DNA DAMAGE-BINDING PROTEIN 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	354	2839	1804	498	526	11	0	0	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	76	MET	-	expression tag	UNP Q2YDS1
B	77	HIS	-	expression tag	UNP Q2YDS1
B	78	HIS	-	expression tag	UNP Q2YDS1
B	79	HIS	-	expression tag	UNP Q2YDS1
B	80	HIS	-	expression tag	UNP Q2YDS1
B	81	HIS	-	expression tag	UNP Q2YDS1
B	82	HIS	-	expression tag	UNP Q2YDS1
B	83	ARG	-	expression tag	UNP Q2YDS1
B	84	ARG	-	expression tag	UNP Q2YDS1
B	85	LEU	-	expression tag	UNP Q2YDS1
B	86	VAL	-	expression tag	UNP Q2YDS1
B	87	PRO	-	expression tag	UNP Q2YDS1
B	88	ARG	-	expression tag	UNP Q2YDS1
B	89	GLY	-	expression tag	UNP Q2YDS1
B	90	SER	-	expression tag	UNP Q2YDS1
B	91	GLY	-	expression tag	UNP Q2YDS1
B	92	GLY	-	expression tag	UNP Q2YDS1
B	93	ARG	-	expression tag	UNP Q2YDS1
B	180	GLN	LEU	variant	UNP Q2YDS1
B	214	ARG	TRP	variant	UNP Q2YDS1

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	G	14	Total 313	C 149	N 64	O 86	P 14	0	0	0

- Molecule 4 is a DNA chain called 5'-D(*CP*CP*TP*GP*CP*TP*CP*CP*TP*TP*TP*CP *AP*CP*CP*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	H	15	Total 292	C 142	N 44	O 92	P 14	0	0	0

- Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	1	Total 1	Ca 1	0	0

- Molecule 6 is water.

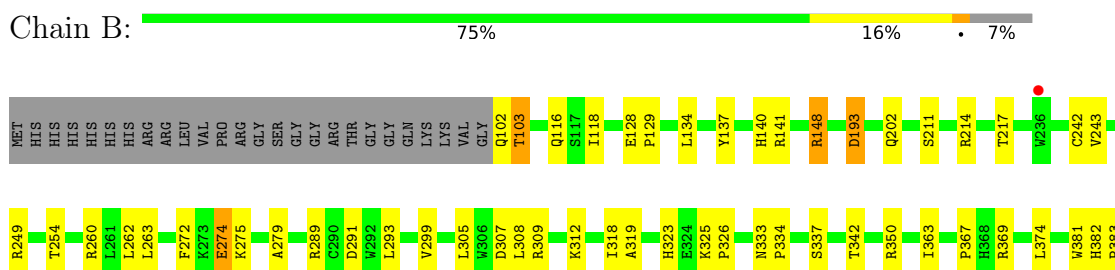| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	16	Total 16	O 16	0	0
6	B	4	Total 4	O 4	0	0
6	G	2	Total 2	O 2	0	0

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 DAMAGE-BINDING PROTEIN 1
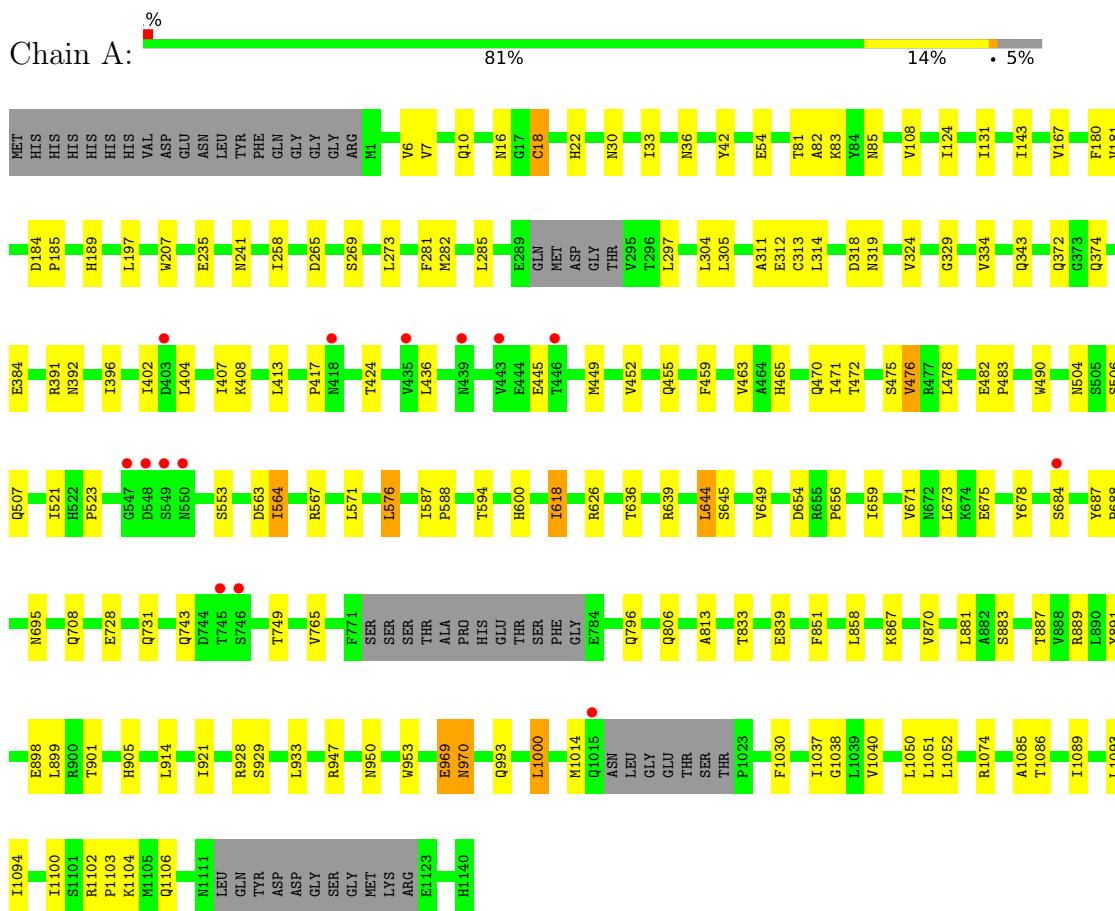
• Molecule 2: DNA DAMAGE-BINDING PROTEIN 2

I388 V389 R392 Y393 P394 D395 D402 T405 I408 R421 L441 A442 S443 F447 L450 I451 W452 N453 R454 E455 ASP THR

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

Chain G:

50%   36%   14%

G1 G2 T3 G4 A5 A6 A7 T8 G15

• Molecule 4: 5'-D(*CP*CP*TP*GP*CP*TP*CP*CP*TP*TP*TP*CP*AP*CP*CP*C)-3'

6%

Chain H:

62%   25%   6%   6%

C1 C2 T3 C8 T9 C15 DC
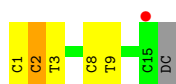
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	111.65Å 122.94Å 154.20Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.27 – 3.10	Depositor
	48.27 – 3.10	EDS
% Data completeness	100.0 (48.27-3.10)	Depositor
(in resolution range)	99.8 (48.27-3.10)	EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.55 (at 3.12Å)	Xtriage
Refinement program	REFMAC 5.5.0102	Depositor
R, R_{free}	0.245 , 0.308	Depositor
	0.239 , 0.298	DCC
R_{free} test set	1955 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	62.1	Xtriage
Anisotropy	0.087	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 45.3	EDS
L-test for twinning[2]	$< |L| > = 0.47, < L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	12113	wwPDB-VP
Average B, all atoms (Å^2)	56.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.45% 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: CA, TTD

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	0.32	0/8803	0.51	0/11924
2	B	0.32	0/2913	0.52	0/3957
3	G	0.61	0/308	1.03	1/473 (0.2%)
4	H	0.77	0/323	1.24	1/494 (0.2%)
All	All	0.35	0/12347	0.56	2/16848 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	G	5	DA	P-O3'-C3'	5.15	125.88	119.70
4	H	2	DC	P-O3'-C3'	5.01	125.71	119.70

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	8646	0	8626	80	0
2	B	2839	0	2785	38	0
3	G	313	0	170	5	0
4	H	292	0	172	4	0
5	B	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	A	16	0	0	0	0
6	B	4	0	0	0	0
6	G	2	0	0	0	0
All	All	12113	0	11753	126	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 5.

The worst 5 of 126 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:81:THR:HG22	1:A:83:LYS:H	1.41	0.86
1:A:167:VAL:HG12	1:A:180:PHE:HB3	1.65	0.78
1:A:482:GLU:HB3	1:A:483:PRO:HD3	1.67	0.75
1:A:184:ASP:HB2	1:A:185:PRO:HD2	1.70	0.71
1:A:329:GLY:HA3	1:A:384:GLU:HG2	1.73	0.70

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	1095/1159 (94%)	1011 (92%)	80 (7%)	4 (0%)	34	69
2	B	352/382 (92%)	327 (93%)	23 (6%)	2 (1%)	25	59
All	All	1447/1541 (94%)	1338 (92%)	103 (7%)	6 (0%)	34	69

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	407	ILE

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	417	PRO
1	A	688	PRO
2	B	193	ASP
2	B	118	ILE

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	967/1014 (95%)	923 (95%)	44 (5%)	27	59
2	B	313/335 (93%)	300 (96%)	13 (4%)	30	62
All	All	1280/1349 (95%)	1223 (96%)	57 (4%)	27	60

5 of 57 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	728	GLU
2	B	393	TYR
1	A	969	GLU
2	B	389	VAL
2	B	274	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 36 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	107	HIS
2	B	453	ASN
2	B	119	HIS
2	B	269	HIS
1	A	504	ASN

5.3.3   RNA ⓘ

There are no RNA molecules in this entry.

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

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	TTD	G	8	3	42,45,46	1.54	8 (19%)	62,74,77	2.70	17 (27%)

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	TTD	G	8	3	-	15/22/109/110	0/5/6/6

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	G	8	TTD	C1R-N1T	4.20	1.51	1.45
3	G	8	TTD	C1'-N1	3.51	1.50	1.45
3	G	8	TTD	PB-O5P	3.50	1.63	1.50
3	G	8	TTD	C2-N1	3.32	1.43	1.36
3	G	8	TTD	C6-N1	3.00	1.51	1.46

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	G	8	TTD	O4R-C1R-N1T	12.26	123.18	108.65
3	G	8	TTD	C5-C4-N3	6.71	121.91	116.06
3	G	8	TTD	C5T-C4T-N3T	6.71	121.90	116.06
3	G	8	TTD	N3-C2-N1	5.72	122.62	116.69
3	G	8	TTD	N3T-C2T-N1T	5.09	121.97	116.69

There are no chirality outliers.

5 of 15 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	G	8	TTD	C2'-C3R-O3R-PB
3	G	8	TTD	O4'-C1'-N1-C2
3	G	8	TTD	C5R-O5R-PB-O3R
3	G	8	TTD	O4R-C4'-C5R-O5R
3	G	8	TTD	C3'-C4'-C5R-O5R

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	G	8	TTD	2	0

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

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.

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	1105/1159 (95%)	0.03	14 (1%)	77	59	39, 56, 80, 96	0
2	B	354/382 (92%)	-0.13	1 (0%)	94	88	33, 49, 66, 75	0
3	G	13/14 (92%)	0.51	0	100	100	56, 73, 100, 102	0
4	H	15/16 (93%)	0.48	1 (6%)	17	7	47, 74, 107, 108	0
All	All	1487/1571 (94%)	0.00	16 (1%)	80	64	33, 54, 79, 108	0

The worst 5 of 16 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	547	GLY	3.5
1	A	446	THR	3.3
1	A	418	ASN	3.2
1	A	746	SER	3.2
1	A	439	ASN	2.9

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
3	TTD	G	8	40/41	0.91	0.24	63,66,66,66	0

6.3 Carbohydrates ⓘ

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
5	CA	B	1456	1/1	0.96	0.31	31,31,31,31	0

6.5   Other polymers (i)

There are no such residues in this entry.