



wwPDB X-ray Structure Validation Summary Report ⓘ

May 22, 2020 – 03:45 pm BST

PDB ID : 4RHD
Title : DNA Duplex with Novel ZP Base Pair
Authors : Zhang, W.; Zhang, L.; Benner, S.; Huang, Z.
Deposited on : 2014-10-01
Resolution : 1.70 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

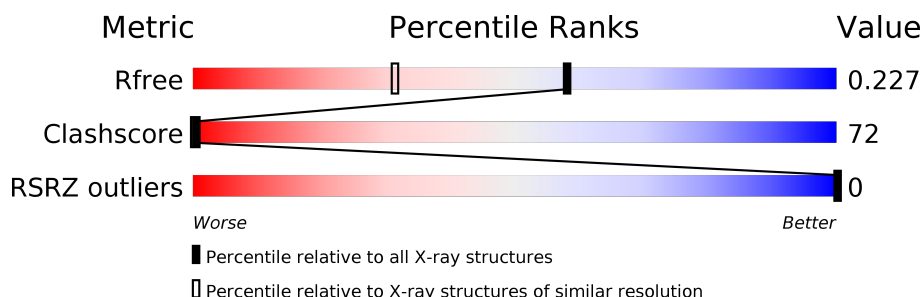
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION


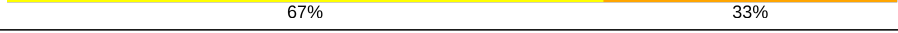

The reported resolution of this entry is 1.70 Å.

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	4298 (1.70-1.70)
Clashscore	141614	4695 (1.70-1.70)
RSRZ outliers	127900	4222 (1.70-1.70)

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	9	
1	C	9	
1	E	9	
1	G	9	
2	B	9	
2	D	9	
2	F	9	
2	H	9	

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
1	1W5	A	5	-	-	X	-
1	1W5	C	5	-	-	X	-
1	1W5	E	5	-	-	X	-
2	T5S	B	2	-	-	X	-
2	T5S	D	2	-	-	X	-
2	1WA	F	5	-	-	X	-

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 1538 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 DNA chain called DNA 9mer novel Z nucleobase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	9	Total	C	N	O	P	Se	0	0	0
			184	88	33	54	8	1			
1	C	9	Total	C	N	O	P	Se	0	0	0
			184	88	33	54	8	1			
1	E	9	Total	C	N	O	P	Se	0	0	0
			184	88	33	54	8	1			
1	G	9	Total	C	N	O	P	Se	0	0	0
			184	88	33	54	8	1			

- Molecule 2 is a DNA chain called DNA 9mer novel P nucleobase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	9	Total	C	N	O	P	Se	0	0	0
			184	88	35	52	8	1			
2	D	9	Total	C	N	O	P	Se	0	0	0
			184	88	35	52	8	1			
2	F	9	Total	C	N	O	P	Se	0	0	0
			184	88	35	52	8	1			
2	H	9	Total	C	N	O	P	Se	0	0	0
			184	88	35	52	8	1			

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	H	1	Total	Mg	0	0
			1	1		
3	C	1	Total	Mg	0	0
			1	1		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	5	Total 5	O 5	0	0
4	B	16	Total 16	O 16	0	0
4	C	17	Total 17	O 17	0	0
4	D	7	Total 7	O 7	0	0
4	E	4	Total 4	O 4	0	0
4	F	5	Total 5	O 5	0	0
4	G	9	Total 9	O 9	0	0
4	H	1	Total 1	O 1	0	0

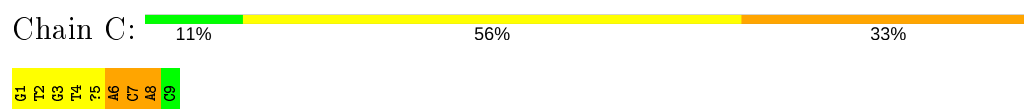
3 Residue-property plots [i](#)

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: DNA 9mer novel Z nucleobase



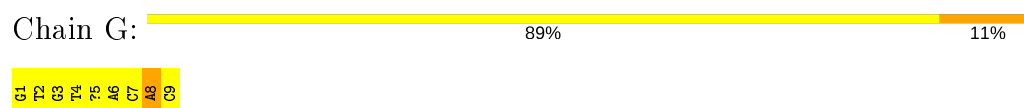
- Molecule 1: DNA 9mer novel Z nucleobase



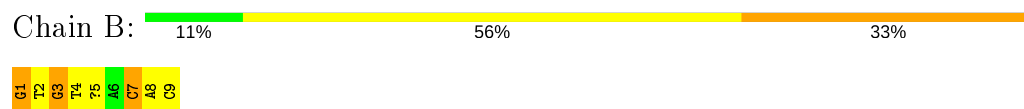
- Molecule 1: DNA 9mer novel Z nucleobase



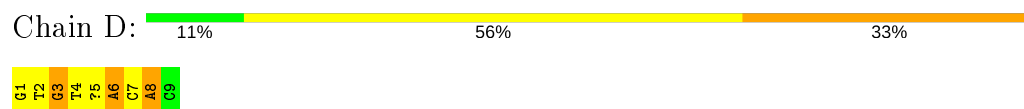
- Molecule 1: DNA 9mer novel Z nucleobase



- Molecule 2: DNA 9mer novel P nucleobase



- Molecule 2: DNA 9mer novel P nucleobase

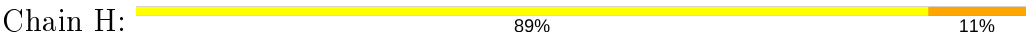


- Molecule 2: DNA 9mer novel P nucleobase



G1	T2	G3	T4	?5	A6	C7	A8	C9
----	----	----	----	----	----	----	----	----

- Molecule 2: DNA 9mer novel P nucleobase



G1	T2	G3	T4	?5	A6	C7	A8	C9
----	----	----	----	----	----	----	----	----

4 Data and refinement statistics

Property	Value	Source
Space group	P 3	Depositor
Cell constants a, b, c, α , β , γ	59.43 Å 59.43 Å 87.05 Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	87.05 – 1.70 29.02 – 1.70	Depositor EDS
% Data completeness (in resolution range)	99.6 (87.05-1.70) 99.7 (29.02-1.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	9.32 (at 1.70 Å)	Xtriage
Refinement program	REFMAC 5.8.0069	Depositor
R, R_{free}	0.223 , 0.226 0.224 , 0.227	Depositor DCC
R_{free} test set	1890 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å ²)	24.8	Xtriage
Anisotropy	0.138	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.42 , 43.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	0.025 for -h,-k,l 0.049 for h,-h-k,-l 0.038 for -k,-h,-l	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	1538	wwPDB-VP
Average B, all atoms (Å ²)	33.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.97% of the height of the origin peak. No significant pseudotranslation is detected.*

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: 1W5, T5S, MG, 1WA

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	1.41	1/156 (0.6%)	2.63	13/234 (5.6%)
1	C	1.28	1/156 (0.6%)	1.36	3/234 (1.3%)
1	E	1.35	0/156	2.06	9/234 (3.8%)
1	G	1.22	1/156 (0.6%)	1.45	3/234 (1.3%)
2	B	1.15	0/156	1.70	5/234 (2.1%)
2	D	1.10	1/156 (0.6%)	1.46	3/234 (1.3%)
2	F	1.26	0/156	1.68	3/234 (1.3%)
2	H	1.21	0/156	1.38	1/234 (0.4%)
All	All	1.25	4/1248 (0.3%)	1.76	40/1872 (2.1%)

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
2	F	0	1

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	G	8	DA	O3'-P	-7.11	1.52	1.61
1	C	7	DC	O3'-P	6.71	1.69	1.61
2	D	3	DG	O3'-P	-6.26	1.53	1.61
1	A	8	DA	P-OP1	5.82	1.58	1.49

The worst 5 of 40 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	9	DC	O5'-P-OP2	-14.47	92.67	105.70

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	7	DC	O5'-P-OP1	14.47	128.07	110.70
1	A	7	DC	O5'-P-OP2	-14.08	93.03	105.70
2	B	7	DC	O5'-P-OP1	10.45	123.24	110.70
1	A	9	DC	O5'-P-OP1	9.99	122.69	110.70

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	F	9	DC	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	184	0	100	20	0
1	C	184	0	99	23	0
1	E	184	0	100	24	0
1	G	184	0	100	21	0
2	B	184	0	101	19	0
2	D	184	0	101	22	0
2	F	184	0	101	28	0
2	H	184	0	101	15	0
3	C	1	0	0	0	0
3	H	1	0	0	0	0
4	A	5	0	0	1	0
4	B	16	0	0	0	0
4	C	17	0	0	0	0
4	D	7	0	0	0	0
4	E	4	0	0	0	0
4	F	5	0	0	0	0
4	G	9	0	0	0	0
4	H	1	0	0	0	0
All	All	1538	0	803	160	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 72.

The worst 5 of 160 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:F:8:DA:C2'	2:F:9:DC:H5'	1.88	1.01
2:F:4:DT:H2'	2:F:5:1WA:H8	1.37	1.01
2:D:1:DG:C5	2:D:2:T5S:SE	2.69	0.95
1:C:1:DG:H8	1:C:1:DG:HO5'	1.02	0.91
1:E:4:DT:C6	1:E:5:1W5:ON2	2.25	0.90

There are no symmetry-related clashes.

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

There are no protein molecules in this entry.

5.3.2 Protein sidechains [i](#)

There are no protein molecules in this entry.

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

16 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$
1	1W5	E	5	1	18,23,24	1.87	4 (22%)	20,33,36	2.65	9 (45%)
2	T5S	B	2	1,2	15,22,23	1.19	2 (13%)	16,31,34	4.09	5 (31%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
2	T5S	H	2	1,3,2	15,22,23	1.27	2 (13%)	16,31,34	6.17	6 (37%)
1	1W5	C	5	1	18,23,24	2.20	4 (22%)	20,33,36	2.03	8 (40%)
1	T5S	G	2	1,2	15,22,23	1.27	2 (13%)	16,31,34	3.13	8 (50%)
2	T5S	D	2	1,2	15,22,23	1.97	5 (33%)	16,31,34	6.64	5 (31%)
1	T5S	E	2	1,2	15,22,23	1.52	2 (13%)	16,31,34	4.83	5 (31%)
1	T5S	A	2	1,2	15,22,23	1.40	1 (6%)	16,31,34	3.72	5 (31%)
1	T5S	C	2	1,2	15,22,23	1.81	4 (26%)	16,31,34	2.72	5 (31%)
1	1W5	A	5	1	18,23,24	1.84	6 (33%)	20,33,36	2.68	8 (40%)
2	T5S	F	2	1,2	15,22,23	1.82	4 (26%)	16,31,34	5.11	5 (31%)
2	1WA	H	5	2	17,24,25	2.57	6 (35%)	19,35,38	2.78	10 (52%)
2	1WA	D	5	2	17,24,25	1.02	1 (5%)	19,35,38	2.01	4 (21%)
2	1WA	F	5	2	17,24,25	1.70	4 (23%)	19,35,38	3.20	9 (47%)
2	1WA	B	5	2	17,24,25	1.60	4 (23%)	19,35,38	2.94	8 (42%)
1	1W5	G	5	1	18,23,24	1.48	4 (22%)	20,33,36	2.99	7 (35%)

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
1	1W5	E	5	1	-	2/8/25/26	0/2/2/2
2	T5S	B	2	1,2	-	0/6/23/24	0/2/2/2
2	T5S	H	2	1,3,2	-	2/6/23/24	0/2/2/2
1	1W5	C	5	1	-	3/8/25/26	0/2/2/2
1	T5S	G	2	1,2	-	3/6/23/24	0/2/2/2
2	T5S	D	2	1,2	-	2/6/23/24	0/2/2/2
1	T5S	E	2	1,2	-	2/6/23/24	0/2/2/2
1	T5S	A	2	1,2	-	2/6/23/24	0/2/2/2
1	T5S	C	2	1,2	-	0/6/23/24	0/2/2/2
1	1W5	A	5	1	-	3/8/25/26	0/2/2/2
2	T5S	F	2	1,2	-	0/6/23/24	0/2/2/2
2	1WA	H	5	2	-	2/3/21/22	0/3/3/3
2	1WA	D	5	2	-	2/3/21/22	0/3/3/3
2	1WA	F	5	2	-	0/3/21/22	0/3/3/3
2	1WA	B	5	2	-	0/3/21/22	0/3/3/3
1	1W5	G	5	1	-	3/8/25/26	0/2/2/2

The worst 5 of 55 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	H	5	1WA	C2'-C3'	7.29	1.72	1.52
2	D	2	T5S	SE-C5	-4.93	1.86	1.91
1	C	5	1W5	C5-N	-4.70	1.37	1.45
1	C	5	1W5	C6-C5	-4.69	1.30	1.39
1	C	5	1W5	O4'-C1'	-4.67	1.39	1.44

The worst 5 of 107 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	2	T5S	CH3-SE-C5	-24.06	81.28	100.02
2	H	2	T5S	CH3-SE-C5	-20.91	83.73	100.02
1	E	2	T5S	CH3-SE-C5	-17.76	86.18	100.02
2	F	2	T5S	CH3-SE-C5	-16.04	87.52	100.02
2	B	2	T5S	CH3-SE-C5	-12.17	90.54	100.02

There are no chirality outliers.

5 of 26 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	H	2	T5S	C6-C5-SE-CH3
1	C	5	1W5	C6-C5-N-ON1
1	G	2	T5S	O4'-C1'-N1-C6
1	G	2	T5S	C6-C5-SE-CH3
1	G	2	T5S	C4-C5-SE-CH3

There are no ring outliers.

16 monomers are involved in 87 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	E	5	1W5	9	0
2	B	2	T5S	7	0
2	H	2	T5S	5	0
1	C	5	1W5	8	0
1	G	2	T5S	5	0
2	D	2	T5S	10	0
1	E	2	T5S	1	0
1	A	2	T5S	4	0
1	C	2	T5S	3	0
1	A	5	1W5	7	0
2	F	2	T5S	3	0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	H	5	1WA	2	0
2	D	5	1WA	4	0
2	F	5	1WA	13	0
2	B	5	1WA	3	0
1	G	5	1W5	3	0

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 2 ligands modelled in this entry, 2 are 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 [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	7/9 (77%)	0.29	0 100 100	22, 24, 31, 36	0
1	C	7/9 (77%)	0.53	0 100 100	29, 32, 40, 48	0
1	E	7/9 (77%)	0.32	0 100 100	22, 28, 29, 40	0
1	G	7/9 (77%)	0.17	0 100 100	26, 31, 33, 41	0
2	B	7/9 (77%)	0.23	0 100 100	24, 29, 37, 40	0
2	D	7/9 (77%)	0.13	0 100 100	27, 28, 36, 38	0
2	F	7/9 (77%)	0.10	0 100 100	26, 28, 34, 38	0
2	H	7/9 (77%)	0.34	0 100 100	30, 32, 33, 34	0
All	All	56/72 (77%)	0.27	0 100 100	22, 30, 40, 48	0

There are no RSRZ outliers to report.

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
1	1W5	E	5	22/23	0.77	0.22	25,38,55,66	0
1	1W5	A	5	22/23	0.82	0.15	19,28,46,57	0
2	1WA	H	5	22/23	0.86	0.22	18,41,77,81	0
1	1W5	C	5	22/23	0.90	0.16	21,29,49,63	0
1	T5S	G	2	21/22	0.90	0.19	25,47,59,68	2
2	T5S	F	2	21/22	0.91	0.14	25,32,43,45	2
1	1W5	G	5	22/23	0.91	0.15	18,28,55,70	0
1	T5S	E	2	21/22	0.92	0.14	24,37,43,48	2

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
2	T5S	B	2	21/22	0.93	0.11	22,30,39,41	2
2	1WA	D	5	22/23	0.93	0.12	19,26,39,43	0
2	1WA	B	5	22/23	0.93	0.12	17,28,40,54	0
1	T5S	A	2	21/22	0.93	0.16	22,34,40,42	2
2	T5S	H	2	21/22	0.94	0.12	22,26,44,49	2
2	T5S	D	2	21/22	0.94	0.12	25,31,38,45	2
1	T5S	C	2	21/22	0.94	0.13	16,29,39,51	2
2	1WA	F	5	22/23	0.96	0.12	17,24,34,44	0

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(\AA^2)	Q<0.9
3	MG	H	101	1/1	0.73	0.10	79,79,79,79	0
3	MG	C	101	1/1	0.81	0.09	77,77,77,77	0

6.5 Other polymers [i](#)

There are no such residues in this entry.