



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 27, 2019 – 08:13 PM EDT

PDB ID : 4A08
Title : Structure of hsDDB1-drDDB2 bound to a 13 bp CPD-duplex (purine at D-1 position) at 3.0 Å resolution (CPD 1)
Authors : Scrima, A.; Fischer, E.S.; Iwai, S.; Gut, H.; Thoma, N.H.
Deposited on : 2011-09-08
Resolution : 3.00 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.8.0 (224370), CSD as540be (2019)
Xtriage (Phenix) : 1.13
EDS : rb-20031633
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac : 5.8.0158
CCP4 : 7.0 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20031633

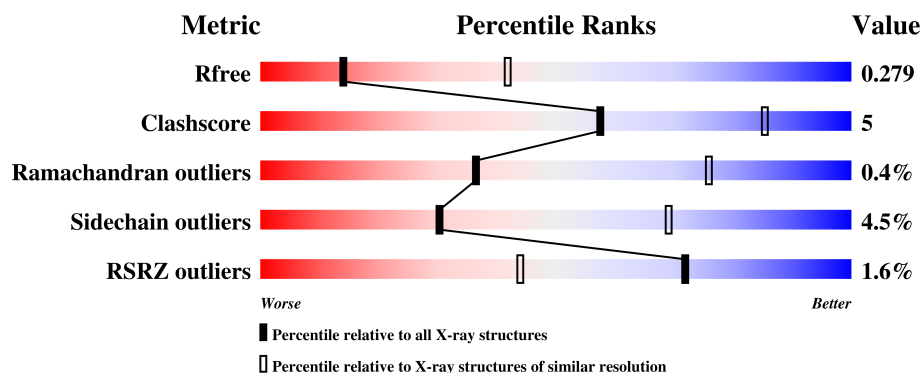
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.00 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}	111664	1851 (3.00-3.00)
Clashscore	122126	2167 (3.00-3.00)
Ramachandran outliers	120053	2101 (3.00-3.00)
Sidechain outliers	120020	2104 (3.00-3.00)
RSRZ outliers	108989	1751 (3.00-3.00)

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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. 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	1159	<div> <div>2%</div> <div>80%</div> <div>13%</div> <div>6%</div> </div>
2	B	382	<div> <div>%</div> <div>73%</div> <div>20%</div> <div>8%</div> </div>
3	G	13	<div> <div>8%</div> <div>54%</div> <div>15%</div> <div>8%</div> <div>23%</div> </div>
4	H	14	<div> <div>50%</div> <div>43%</div> <div>7%</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
5	MES	A	2141	-	-	-	X

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 11866 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
1	A	1088	Total	C	N	O	S	0	0	0
			8509	5412	1436	1615	46			

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
2	B	353	Total	C	N	O	S	0	0	0
			2824	1796	493	524	11			

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

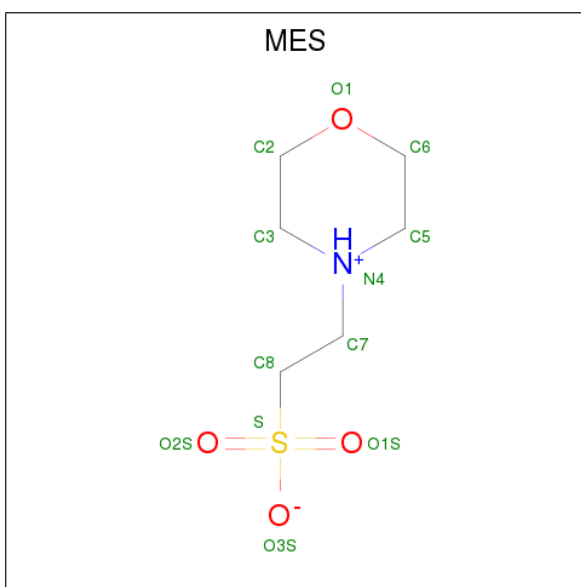
- Molecule 3 is a DNA chain called 5'-D(*AP*CP*GP*CP*GP*AP*(TTD)P*GP*CP*GP*C P*CP*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	G	10	Total	C	N	O	P	0	0	0
			225	106	41	67	11			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	H	13	Total	C	N	O	P	0	0	0
			266	124	50	79	13			

- Molecule 5 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (three-letter code: MES) (formula: C₆H₁₃NO₄S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	A	1	Total	C	N	O	S	0	0
			12	6	1	4	1		
5	A	1	Total	C	N	O	S	0	0
			12	6	1	4	1		

- Molecule 6 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	H	1	Total	Ca	0	0
			1	1		
6	B	1	Total	Ca	0	0
			1	1		

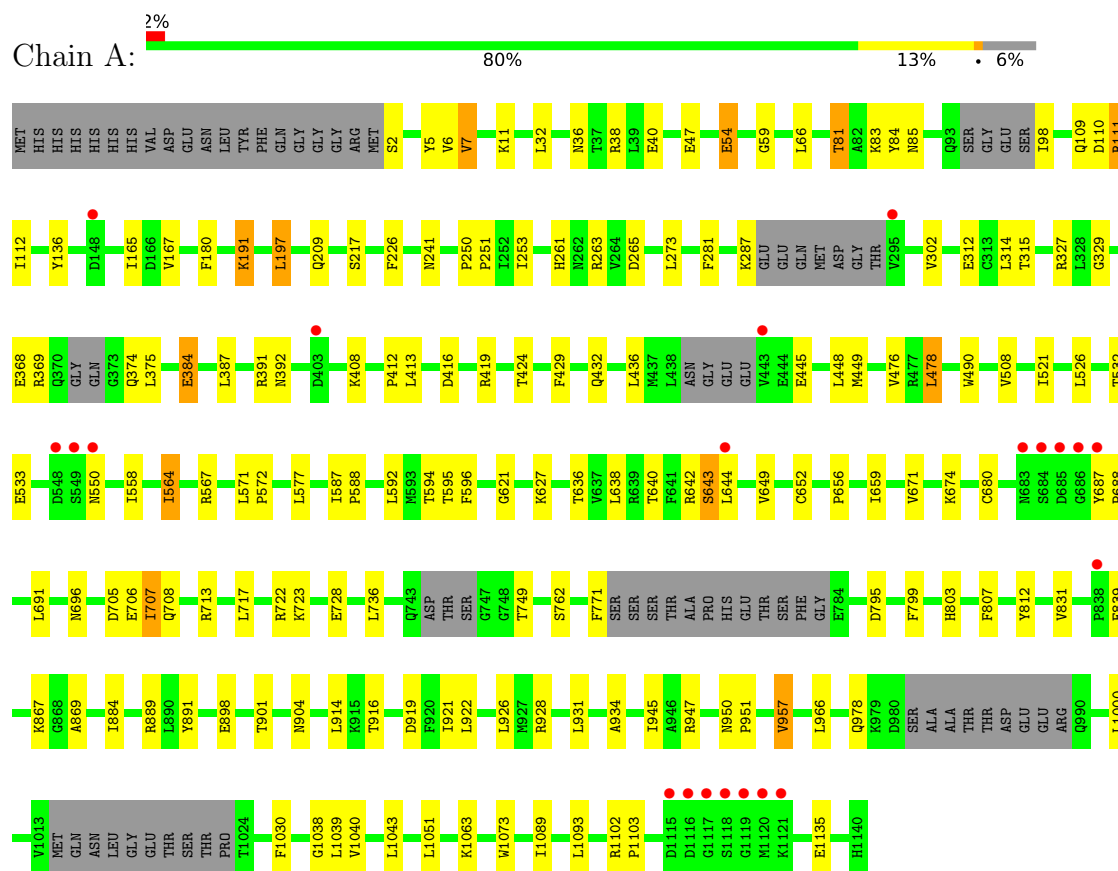
- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	11	Total	O	0	0
			11	11		
7	B	5	Total	O	0	0
			5	5		

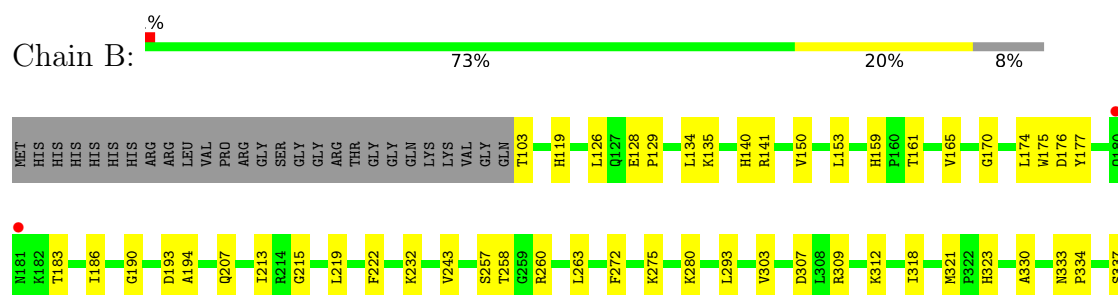
3 Residue-property plots [i](#)

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 DAMAGE-BINDING PROTEIN 1



• Molecule 2: DNA DAMAGE-BINDING PROTEIN 2





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



- Molecule 4: 5'-D(*TP*GP*GP*GP*CP*GP*CP*CP*CP*TP*CP*GP*CP*G)-3'



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	114.37Å 116.70Å 137.58Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.62 – 3.00 48.62 – 3.00	Depositor EDS
% Data completeness (in resolution range)	100.0 (48.62-3.00) 99.4 (48.62-3.00)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.04 (at 3.01Å)	Xtriage
Refinement program	REFMAC 5.5.0102	Depositor
R, R_{free}	0.234 , 0.294 0.229 , 0.279	Depositor DCC
R_{free} test set	1866 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	55.5	Xtriage
Anisotropy	0.004	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 40.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.021 for k,h,-l	Xtriage
F_o, F_c correlation	0.90	EDS
Total number of atoms	11866	wwPDB-VP
Average B, all atoms (Å ²)	48.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.97% 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 [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: CA, MES, 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	0/8662	0.50	0/11729
2	B	0.31	0/2898	0.49	0/3938
3	G	0.61	0/206	1.01	1/313 (0.3%)
4	H	0.69	0/297	1.05	0/456
All	All	0.34	0/12063	0.54	1/16436 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	G	2	DC	P-O3'-C3'	5.67	126.50	119.70

There are no chirality outliers.

There are no planarity outliers.

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	8509	0	8508	85	0
2	B	2824	0	2766	34	0
3	G	225	0	124	2	0
4	H	266	0	145	6	0
5	A	24	0	24	2	0
6	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	H	1	0	0	0	0
7	A	11	0	0	0	0
7	B	5	0	0	0	0
All	All	11866	0	11567	124	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 124 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:329:GLY:HA3	1:A:384:GLU:HG2	1.59	0.82
2:B:140:HIS:HB2	2:B:453:ASN:HB2	1.66	0.77
1:A:707:ILE:O	1:A:708:GLN:HB2	1.89	0.70
1:A:81:THR:HG22	1:A:85:ASN:H	1.55	0.70
1:A:263:ARG:HH11	1:A:263:ARG:HG2	1.56	0.69

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	1070/1159 (92%)	993 (93%)	74 (7%)	3 (0%)	43 80
2	B	351/382 (92%)	330 (94%)	19 (5%)	2 (1%)	27 67
All	All	1421/1541 (92%)	1323 (93%)	93 (6%)	5 (0%)	36 76

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	215	GLY

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	B	190	GLY
1	A	36	ASN
1	A	643	SER
1	A	564	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	948/1014 (94%)	905 (96%)	43 (4%)	30	69
2	B	311/335 (93%)	297 (96%)	14 (4%)	30	69
All	All	1259/1349 (93%)	1202 (96%)	57 (4%)	30	69

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

Mol	Chain	Res	Type
1	A	691	LEU
1	A	771	PHE
2	B	359	LYS
1	A	696	ASN
1	A	707	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	648	ASN
1	A	796	GLN
2	B	269	HIS
1	A	677	ASN
1	A	696	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	7	3	42,45,46	1.70	9 (21%)	62,74,77	2.55	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	7	3	-	0/22/109/110	0/5/6/6

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	G	7	TTD	O4'-C1'	2.13	1.47	1.42
3	G	7	TTD	C5T-C4T	2.14	1.55	1.51
3	G	7	TTD	C6T-N1T	2.61	1.50	1.46
3	G	7	TTD	C2T-N1T	2.78	1.42	1.36
3	G	7	TTD	C6-N1	3.02	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	7	TTD	O4-C4-C5	-3.79	119.85	122.88
3	G	7	TTD	C4-N3-C2	-3.37	121.48	126.67
3	G	7	TTD	C2'-C1'-N1	-3.32	111.10	115.59
3	G	7	TTD	C4T-N3T-C2T	-3.08	121.94	126.67
3	G	7	TTD	O4T-C4T-C5T	-3.03	120.46	122.88

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
3	G	7	TTD	1	0

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 4 ligands modelled in this entry, 2 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	MES	A	2141	-	12,12,12	2.12	1 (8%)	14,16,16	7.40	9 (64%)
5	MES	A	2142	-	12,12,12	2.11	2 (16%)	14,16,16	7.36	9 (64%)

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	MES	A	2141	-	-	0/6/14/14	0/1/1/1
5	MES	A	2142	-	-	0/6/14/14	0/1/1/1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	2141	MES	C8-S	-6.96	1.67	1.77
5	A	2142	MES	C8-S	-6.68	1.68	1.77
5	A	2142	MES	O1S-S	2.35	1.52	1.45

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	2142	MES	O1S-S-C8	-17.52	85.81	106.92
5	A	2141	MES	O2S-S-C8	-17.22	86.17	106.92
5	A	2141	MES	O1S-S-C8	-14.95	88.92	106.92
5	A	2142	MES	O2S-S-C8	-14.61	89.32	106.92
5	A	2141	MES	O3S-S-C8	-12.52	85.51	105.77

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	2142	MES	2	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	1088/1159 (93%)	-0.06	21 (1%) 66 37	30, 47, 68, 82	0
2	B	353/382 (92%)	-0.14	2 (0%) 89 71	32, 46, 60, 65	0
3	G	9/13 (69%)	0.65	1 (11%) 5 2	59, 69, 93, 98	0
4	H	13/14 (92%)	0.23	0 100 100	57, 72, 79, 79	0
All	All	1463/1568 (93%)	-0.08	24 (1%) 72 44	30, 47, 67, 98	0

The worst 5 of 24 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	181	ASN	4.5
1	A	1117	GLY	3.5
1	A	686	GLY	3.3
1	A	687	TYR	3.2
1	A	1116	ASP	3.2

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	TTD	G	7	40/41	0.90	0.20	64,66,67,67	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
5	MES	A	2141	12/12	0.77	0.41	116,116,117,117	0
6	CA	H	1015	1/1	0.79	0.31	78,78,78,78	0
5	MES	A	2142	12/12	0.86	0.29	83,84,84,84	0
6	CA	B	1456	1/1	0.92	0.09	56,56,56,56	0

6.5 Other polymers [i](#)

There are no such residues in this entry.