WORLDWIDE PDB
PROTEIN DATA BANK

Full wwPDB EM Validation Report ⓘ

Mar 9, 2025 – 12:08 am GMT

PDB ID	:	9HDP
EMDB ID	:	EMD-52071
Title	:	The Human LINE-1 ORF2p target-primed reverse transcription complex with the fingers domain in a closed conformation
Authors	:	Ghanim, G.E.; Hu, H.; Nguyen, T.H.D.
Deposited on	:	2024-11-12
Resolution	:	2.50 Å(reported)
Based on initial model	:	.

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

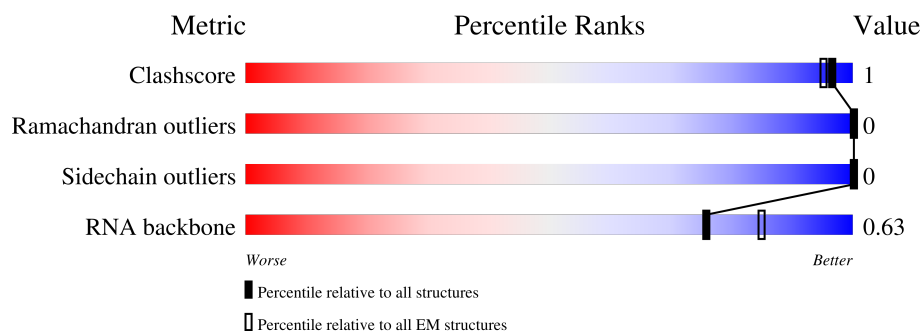|		
EMDB validation analysis	:	FAILED
Mogul	:	1.8.4, CSD as541be (2020)
MolProbity	:	4.02b-467
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
MapQ	:	FAILED
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.41

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
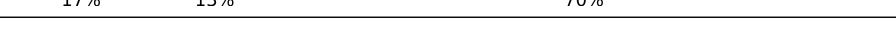*ELECTRON MICROSCOPY*

The reported resolution of this entry is 2.50 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
Clashscore		1
Ramachandran outliers		0
Sidechain outliers		0
RNA backbone		0.63

Worse   *Better*

▌ Percentile relative to all structures

▯ Percentile relative to all EM structures

Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	210492	15764
Ramachandran outliers	207382	16835
Sidechain outliers	206894	16415
RNA backbone	6643	2191

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the 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%

Mol	Chain	Length	Quality of chain
1	A	1797	56% · 43%
2	B	37	32% 14% 54%
3	C	30	33% 67%
4	D	30	17% 13% 70%
5	E	26	100%
6	F	19	100%
7	G	4	100%

WORLDWIDE
wwPDB
PROTEIN DATA BANK

2 Entry composition ⓘ

There are 10 unique types of molecules in this entry. The entry contains 10290 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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 LINE-1 retrotransposable element ORF2 protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	1025	Total 8505	C 5501	N 1460	O 1505	S 39	0	0

There are 79 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-521	MET	-	initiating methionine	UNP P0AEX9
A	-520	VAL	-	expression tag	UNP P0AEX9
A	-519	HIS	-	expression tag	UNP P0AEX9
A	-518	HIS	-	expression tag	UNP P0AEX9
A	-517	HIS	-	expression tag	UNP P0AEX9
A	-516	HIS	-	expression tag	UNP P0AEX9
A	-515	HIS	-	expression tag	UNP P0AEX9
A	-514	HIS	-	expression tag	UNP P0AEX9
A	-513	HIS	-	expression tag	UNP P0AEX9
A	-512	HIS	-	expression tag	UNP P0AEX9
A	-511	GLY	-	expression tag	UNP P0AEX9
A	-510	GLY	-	expression tag	UNP P0AEX9
A	-509	SER	-	expression tag	UNP P0AEX9
A	-508	ALA	-	expression tag	UNP P0AEX9
A	-507	TRP	-	expression tag	UNP P0AEX9
A	-506	SER	-	expression tag	UNP P0AEX9
A	-505	HIS	-	expression tag	UNP P0AEX9
A	-504	PRO	-	expression tag	UNP P0AEX9
A	-503	GLN	-	expression tag	UNP P0AEX9
A	-502	PHE	-	expression tag	UNP P0AEX9
A	-501	GLU	-	expression tag	UNP P0AEX9
A	-500	LYS	-	expression tag	UNP P0AEX9
A	-499	GLY	-	expression tag	UNP P0AEX9
A	-498	GLY	-	expression tag	UNP P0AEX9
A	-497	GLY	-	expression tag	UNP P0AEX9
A	-496	SER	-	expression tag	UNP P0AEX9
A	-495	GLY	-	expression tag	UNP P0AEX9
A	-494	GLY	-	expression tag	UNP P0AEX9

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	-493	GLY	-	expression tag	UNP P0AEX9
A	-492	SER	-	expression tag	UNP P0AEX9
A	-491	GLY	-	expression tag	UNP P0AEX9
A	-490	GLY	-	expression tag	UNP P0AEX9
A	-489	SER	-	expression tag	UNP P0AEX9
A	-488	ALA	-	expression tag	UNP P0AEX9
A	-487	TRP	-	expression tag	UNP P0AEX9
A	-486	SER	-	expression tag	UNP P0AEX9
A	-485	HIS	-	expression tag	UNP P0AEX9
A	-484	PRO	-	expression tag	UNP P0AEX9
A	-483	GLN	-	expression tag	UNP P0AEX9
A	-482	PHE	-	expression tag	UNP P0AEX9
A	-481	GLU	-	expression tag	UNP P0AEX9
A	-480	LYS	-	expression tag	UNP P0AEX9
A	-479	GLY	-	expression tag	UNP P0AEX9
A	-478	GLY	-	expression tag	UNP P0AEX9
A	-477	SER	-	expression tag	UNP P0AEX9
A	-167	VAL	ALA	conflict	UNP P0AEX9
A	-112	ASN	-	linker	UNP P0AEX9
A	-111	SER	-	linker	UNP P0AEX9
A	-110	SER	-	linker	UNP P0AEX9
A	-109	SER	-	linker	UNP P0AEX9
A	-108	ASN	-	linker	UNP P0AEX9
A	-107	ASN	-	linker	UNP P0AEX9
A	-106	ASN	-	linker	UNP P0AEX9
A	-105	ASN	-	linker	UNP P0AEX9
A	-104	ASN	-	linker	UNP P0AEX9
A	-103	ASN	-	linker	UNP P0AEX9
A	-102	ASN	-	linker	UNP P0AEX9
A	-101	ASN	-	linker	UNP P0AEX9
A	-100	ASN	-	linker	UNP P0AEX9
A	-99	ASN	-	linker	UNP P0AEX9
A	-98	LEU	-	linker	UNP P0AEX9
A	-97	GLY	-	linker	UNP P0AEX9
A	-34	THR	ARG	conflict	UNP Q12306
A	-27	GLU	ARG	conflict	UNP Q12306
A	5	THR	ASN	conflict	UNP O00370
A	15	ILE	VAL	conflict	UNP O00370
A	21	ALA	PRO	conflict	UNP O00370
A	152	THR	ILE	conflict	UNP O00370
A	175	ALA	THR	conflict	UNP O00370
A	595	MET	VAL	conflict	UNP O00370

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	625	THR	MET	conflict	UNP O00370
A	627	PHE	LEU	conflict	UNP O00370
A	744	THR	ASN	conflict	UNP O00370
A	795	GLU	ASP	conflict	UNP O00370
A	932	PHE	LEU	conflict	UNP O00370
A	1141	LEU	VAL	conflict	UNP O00370
A	1182	ASN	LYS	conflict	UNP O00370
A	1183	GLU	ASP	conflict	UNP O00370
A	1212	LYS	ASN	conflict	UNP O00370

- Molecule 2 is a DNA chain called Target DNA strand 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
2	B	17	345	170	43	115	17	0	0

- Molecule 3 is a DNA chain called Target DNA strand 2.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
3	C	10	205	98	43	54	10	0	0

- Molecule 4 is a RNA chain called Template P(A)30 RNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
4	D	9	198	90	45	54	9	0	0

- Molecule 5 is a DNA chain called Target DNA strand 3.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
5	E	26	524	260	64	174	26	0	0

- Molecule 6 is a DNA chain called Target DNA strand 4.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
6	F	19	399	190	95	95	19	0	0

- Molecule 7 is a DNA chain called Unassigned Nucleic Acid.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	G	4	Total	C	N	O	P	0	0
			84	40	20	20	4		

- Molecule 8 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		AltConf
8	A	1	Total	Mg	0
			1	1	

- Molecule 9 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		AltConf
9	A	1	Total	Zn	0
			1	1	

- Molecule 10 is 2',3'-DIDEOXY-THYMIDINE-5'-TRIPHOSPHATE (three-letter code: D3T) (formula: $C_{10}H_{17}N_2O_{13}P_3$).

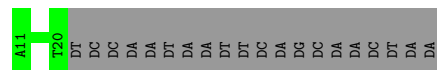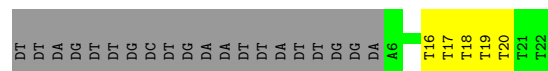| Mol	Chain	Residues	Atoms					AltConf
10	A	1	Total	C	N	O	P	0
			28	10	2	13	3	
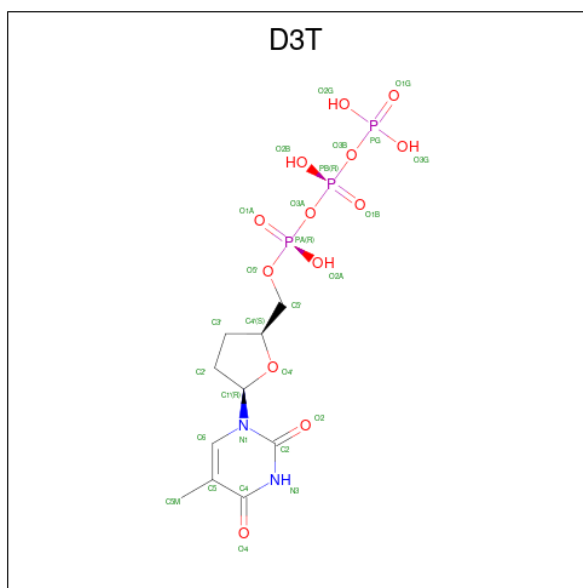
3 Residue-property plots ⓘ

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. 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. 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: LINE-1 retrotransposable element ORF2 protein
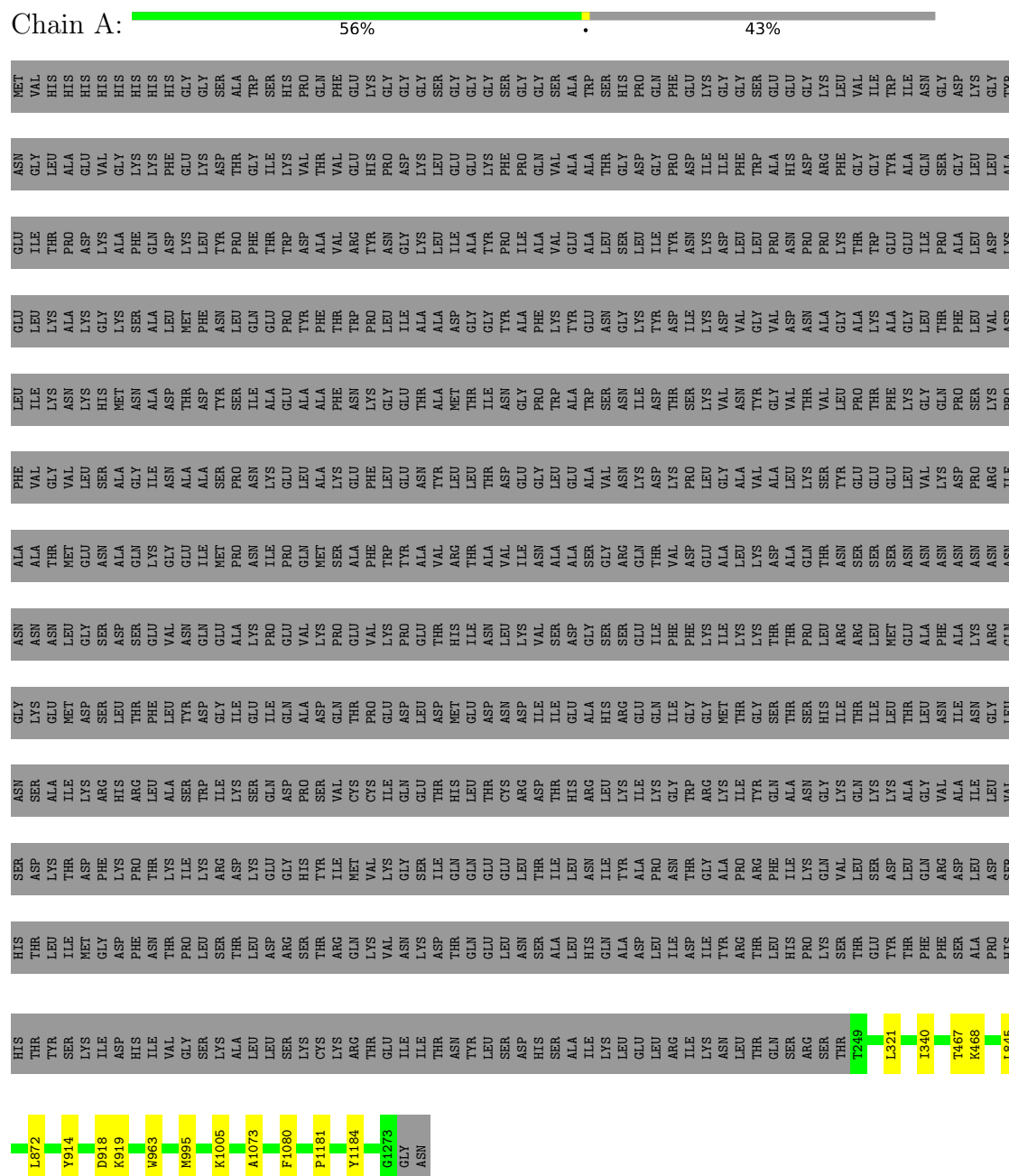
Chain A:

• Molecule 2: Target DNA strand 1

Chain B:

• Molecule 3: Target DNA strand 2

Chain C:

• Molecule 4: Template P(A)30 RNA

Chain D:

• Molecule 5: Target DNA strand 3

Chain E:

There are no outlier residues recorded for this chain.

• Molecule 6: Target DNA strand 4

Chain F:

There are no outlier residues recorded for this chain.

• Molecule 7: Unassigned Nucleic Acid

Chain G:

There are no outlier residues recorded for this chain.

4 Experimental information ⓘ

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	185228	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	TFS KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose $(e^-/\text{Å}^2)$	60	Depositor
Minimum defocus (nm)	8000	Depositor
Maximum defocus (nm)	22000	Depositor
Magnification	Not provided	
Image detector	TFS FALCON 4i (4k x 4k)	Depositor

5   Model quality ⓘ

5.1   Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: D3T, MG, ZN

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	0.30	0/8698	0.60	0/11724
2	B	0.62	0/381	1.08	0/587
3	C	0.66	0/231	0.96	0/353
4	D	0.48	0/224	1.08	0/347
5	E	0.65	0/579	1.13	0/892
6	F	0.59	0/455	1.08	0/699
7	G	0.73	0/95	0.83	0/144
All	All	0.38	0/10663	0.72	0/14746

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	8505	0	8793	8	0
2	B	345	0	202	4	0
3	C	205	0	112	0	0
4	D	198	0	100	1	0
5	E	524	0	309	0	0
6	F	399	0	210	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
7	G	84	0	45	0	0
8	A	1	0	0	0	0
9	A	1	0	0	0	0
10	A	28	0	13	0	0
All	All	10290	0	9784	13	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 1.

All (13) 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:467:THR:HG22	1:A:468:LYS:HG3	1.87	0.56
1:A:914:TYR:O	1:A:918:ASP:HB3	2.05	0.56
2:B:18:DT:H2'	2:B:19:DT:H71	1.86	0.56
1:A:1073:ALA:HB2	1:A:1080:PHE:CD2	2.45	0.52
2:B:19:DT:H2'	2:B:20:DT:H71	1.93	0.51
2:B:19:DT:C2'	2:B:20:DT:H71	2.43	0.48
1:A:919:LYS:HB2	1:A:963:TRP:CE2	2.51	0.45
2:B:16:DT:H2'	2:B:17:DT:H71	1.99	0.45
1:A:845:LEU:HD21	1:A:872:LEU:HD21	2.01	0.43
4:D:7:A:H2'	4:D:8:A:C8	2.55	0.41
1:A:321:LEU:HD11	1:A:340:ILE:HG23	2.02	0.41
1:A:1181:PRO:HG2	1:A:1184:TYR:CD2	2.56	0.41
1:A:995:MET:SD	1:A:1005:LYS:HG3	2.61	0.40

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 PDB entries followed by that with respect to all EM entries.

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	1023/1797 (57%)	1014 (99%)	9 (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 PDB entries followed by that with respect to all EM entries.

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	941/1593 (59%)	941 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (7) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	389	GLN
1	A	554	HIS
1	A	587	ASN
1	A	618	ASN
1	A	1055	ASN
1	A	1127	ASN
1	A	1266	HIS

5.3.3   RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
4	D	9/30 (30%)	1 (11%)	1 (11%)

All (1) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
4	D	1	A

All (1) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
4	D	0	A

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 oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 3 ligands modelled in this entry, 2 are monoatomic - leaving 1 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	$\#\|Z\| > 2$
10	D3T	A	1303	8	25,29,29	1.48	4 (16%)	35,45,45	1.11	3 (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
10	D3T	A	1303	8	-	8/22/31/31	0/2/2/2

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
10	A	1303	D3T	C6-C5	3.28	1.40	1.34
10	A	1303	D3T	C1'-N1	3.02	1.55	1.48
10	A	1303	D3T	C6-N1	2.56	1.42	1.38
10	A	1303	D3T	C3'-C2'	-2.22	1.47	1.54

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
10	A	1303	D3T	C3'-C2'-C1'	2.71	105.91	102.78
10	A	1303	D3T	N3-C2-N1	2.11	117.69	114.89
10	A	1303	D3T	C4-N3-C2	-2.03	124.73	127.35

There are no chirality outliers.

All (8) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
10	A	1303	D3T	C5'-O5'-PA-O2A
10	A	1303	D3T	PB-O3A-PA-O1A
10	A	1303	D3T	C5'-O5'-PA-O3A
10	A	1303	D3T	PA-O3A-PB-O2B
10	A	1303	D3T	PB-O3B-PG-O1G
10	A	1303	D3T	PB-O3B-PG-O2G
10	A	1303	D3T	PB-O3A-PA-O2A
10	A	1303	D3T	C5'-O5'-PA-O1A

There are no ring outliers.

No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand D3T A 1303

Bond lengths

Bond angles

Torsions

Rings
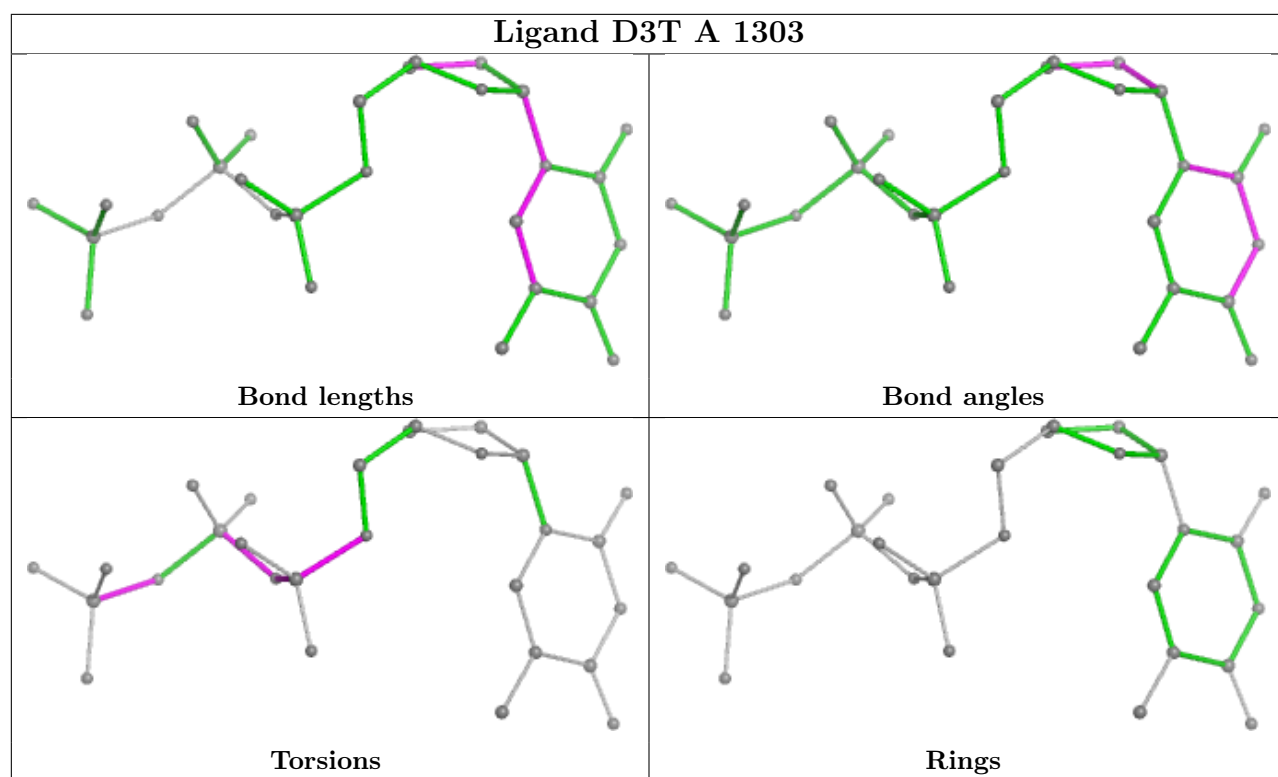
5.7　Other polymers ⓘ

There are no such residues in this entry.

5.8　Polymer linkage issues ⓘ

There are no chain breaks in this entry.