W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Apr 29, 2024 – 01:48 am BST

PDB ID	:	1H5R
Title	:	Thymidylyltransferase complexed with Thimidine and Glucose-1-phospate
Authors	:	Rosano, C.; Zuccotti, S.; Bolognesi, M.
Deposited on	:	2001-05-25
Resolution	:	1.90 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)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.90 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		17
Ramachandran outliers		0.3%
Sidechain outliers		5.3%

Worse   *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution
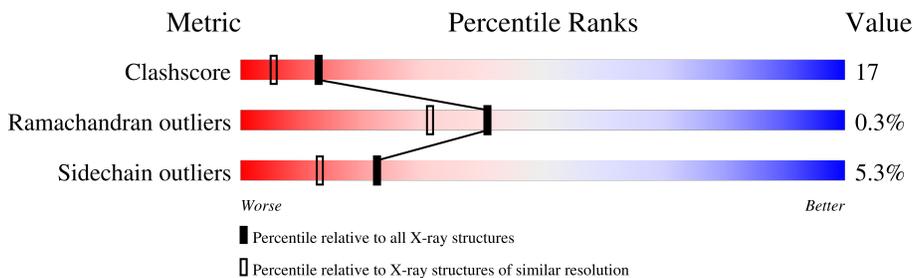
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)

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%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	293	78% 18% • ••
1	C	293	76% 20% • ••
1	D	293	73% 24% • •
2	B	293	70% 25% • ••

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
3	THM	B	1292	-	-	X	-
4	G1P	A	1294	-	-	X	-
4	G1P	D	1294	X	-	-	-
5	SO4	C	1295	-	-	X	-

2 Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 9791 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 GLUCOSE-1-PHOSPHATE THYMIDYLYLTRANSFERASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	290	Total	C	N	O	S	8	0	0
			2277	1459	378	429	11			
1	C	290	Total	C	N	O	S	9	0	0
			2273	1457	378	427	11			
1	D	290	Total	C	N	O	S	9	0	0
			2273	1457	378	427	11			

- Molecule 2 is a protein called GLUCOSE-1-PHOSPHATE THYMIDYLYLTRANSFERASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	290	Total	C	N	O	S	9	0	0
			2270	1455	378	428	9			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	217	LEU	MET	conflict	UNP P37744
B	288	GLN	MET	conflict	UNP P37744

- Molecule 3 is THYMIDINE (three-letter code: THM) (formula: $C_{10}H_{14}N_2O_5$).

| Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			17	10	2	5		
3	A	1	Total	C	N	O	0	0
			17	10	2	5		
3	B	1	Total	C	N	O	0	0
			17	10	2	5		
3	B	1	Total	C	N	O	0	0
			17	10	2	5		
3	C	1	Total	C	N	O	0	0
			17	10	2	5		
3	C	1	Total	C	N	O	0	0
			17	10	2	5		
3	D	1	Total	C	N	O	0	0
			17	10	2	5		
3	D	1	Total	C	N	O	0	0
			17	10	2	5		

- Molecule 4 is 1-O-phosphono-alpha-D-glucopyranose (three-letter code: G1P) (formula: $C_6H_{13}O_9P$).

G1P
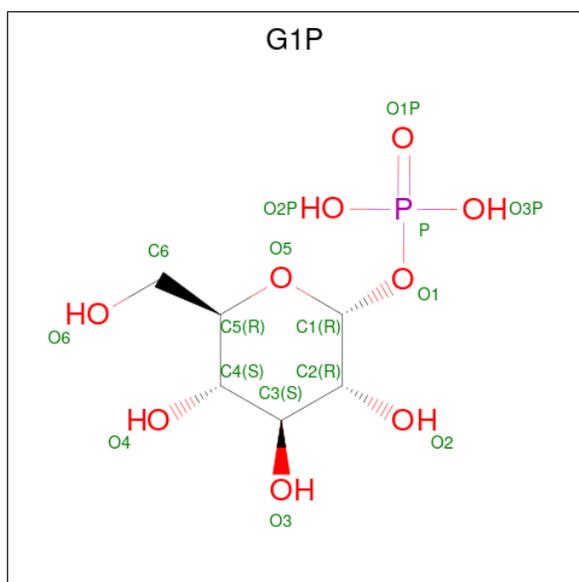
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total	C	O	P	0	0
			16	6	9	1		
4	B	1	Total	C	O	P	0	0
			16	6	9	1		
4	C	1	Total	C	O	P	0	0
			16	6	9	1		
4	D	1	Total	C	O	P	1	0
			16	6	9	1		

- Molecule 5 is SULFATE ION (three-letter code: SO4) (formula: O_4S).

SO4
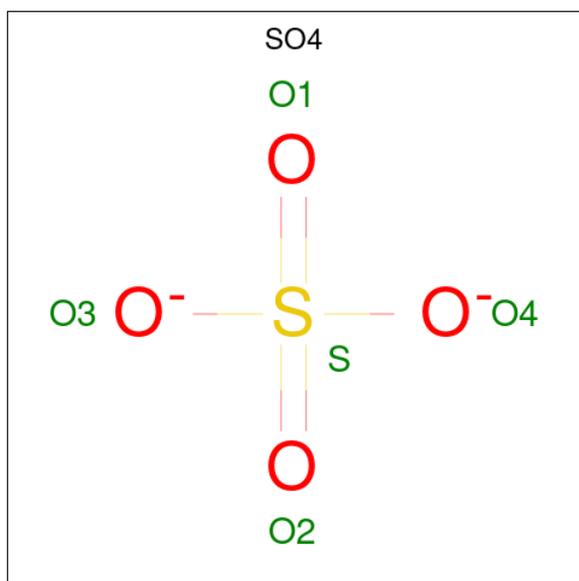
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total 5	O 4	S 1	0	0
5	A	1	Total 5	O 4	S 1	0	0
5	B	1	Total 5	O 4	S 1	0	0
5	C	1	Total 5	O 4	S 1	0	0
5	D	1	Total 5	O 4	S 1	0	0

• Molecule 6 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	134	Total 134	O 134	0	0
6	B	97	Total 97	O 97	0	0
6	C	108	Total 108	O 108	0	0
6	D	134	Total 134	O 134	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. 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.

Note EDS was not executed.

• Molecule 1: GLUCOSE-1-PHOSPHATE THYMIDYLYLTRANSFERASE
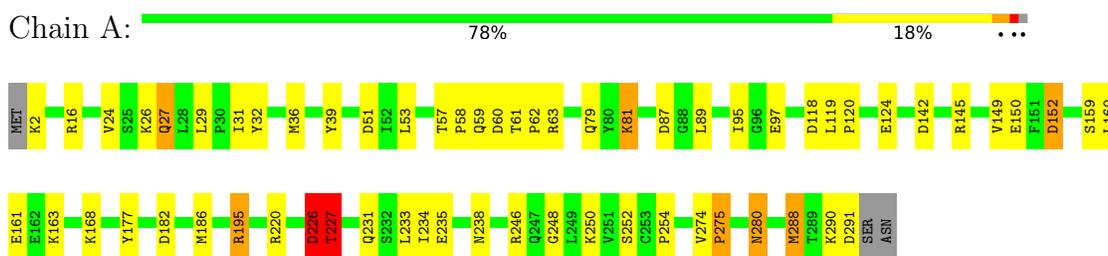
Chain A:

• Molecule 1: GLUCOSE-1-PHOSPHATE THYMIDYLYLTRANSFERASE

Chain C:

• Molecule 1: GLUCOSE-1-PHOSPHATE THYMIDYLYLTRANSFERASE

Chain D:
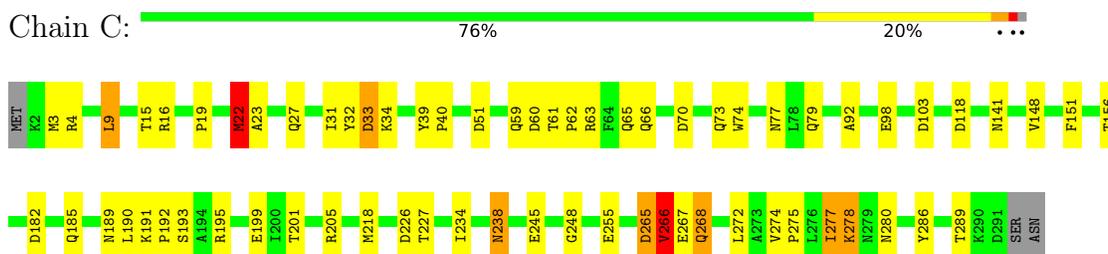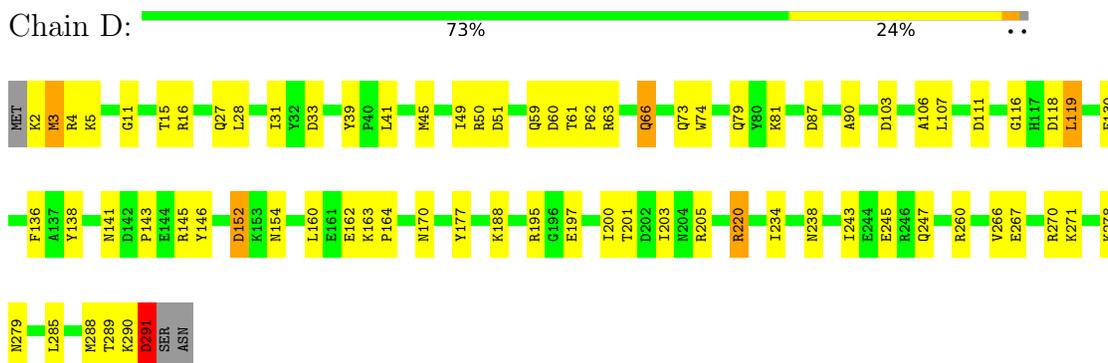
• Molecule 2: GLUCOSE-1-PHOSPHATE THYMIDYLYLTRANSFERASE

Chain B:
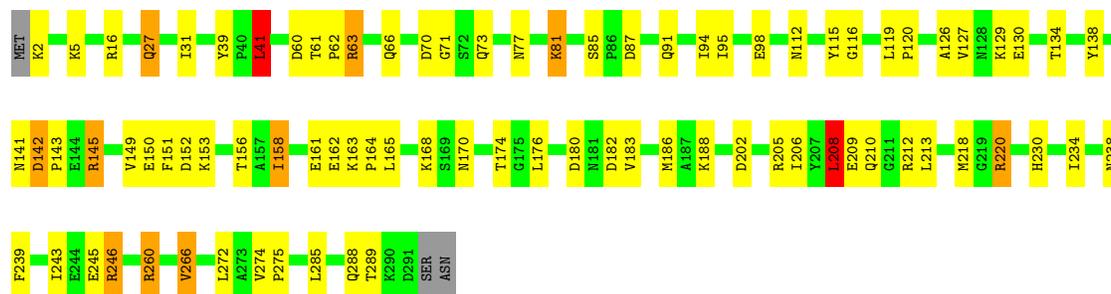
4 Data and refinement statistics ⓘ

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	72.81Å 118.96Å 80.59Å 90.00° 112.60° 90.00°	Depositor
Resolution (Å)	12.00 – 1.90	Depositor
% Data completeness (in resolution range)	85.4 (12.00-1.90)	Depositor
R_{merge}	0.03	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	REFMAC	Depositor
R, R_{free}	0.173 , 0.234	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	9791	wwPDB-VP
Average B, all atoms (Å^2)	28.0	wwPDB-VP

5 Model quality ⓘ

5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: THM, G1P, SO4

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	0.97	3/2324 (0.1%)	1.19	14/3145 (0.4%)
1	C	0.93	2/2320 (0.1%)	1.05	13/3139 (0.4%)
1	D	0.81	2/2320 (0.1%)	1.02	10/3140 (0.3%)
2	B	0.79	1/2317 (0.0%)	1.00	10/3138 (0.3%)
All	All	0.88	8/9281 (0.1%)	1.07	47/12562 (0.4%)

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
1	A	0	1
2	B	1	1
All	All	1	2

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	16	ARG	CG-CD	-21.04	0.99	1.51
1	A	152	ASP	CA-CB	20.32	1.98	1.53
1	A	16	ARG	CD-NE	-12.85	1.24	1.46
2	B	16	ARG	CG-CD	8.40	1.73	1.51
1	D	16	ARG	CG-CD	-8.18	1.31	1.51

The worst 5 of 47 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	16	ARG	CD-NE-CZ	-24.89	88.75	123.60
1	A	16	ARG	CG-CD-NE	16.33	146.09	111.80

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	152	ASP	CB-CA-C	-12.27	85.85	110.40
1	A	152	ASP	N-CA-CB	-11.60	89.72	110.60
2	B	152	ASP	N-CA-CB	10.67	129.81	110.60

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	B	174	THR	CB

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	195	ARG	Sidechain
2	B	208	LEU	Mainchain

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	2277	0	2284	68	0
1	C	2273	0	2281	68	0
1	D	2273	0	2280	92	0
2	B	2270	0	2270	87	0
3	A	34	0	28	0	0
3	B	34	0	28	9	0
3	C	34	0	28	5	0
3	D	34	0	28	5	0
4	A	16	0	11	13	0
4	B	16	0	11	1	0
4	C	16	0	11	0	0
4	D	16	0	10	2	0
5	A	10	0	0	0	0
5	B	5	0	0	0	0
5	C	5	0	0	4	0
5	D	5	0	0	0	0
6	A	134	0	0	16	0
6	B	97	0	0	12	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	C	108	0	0	5	0
6	D	134	0	0	25	0
All	All	9791	0	9270	306	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 17.

The worst 5 of 306 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:B:245:GLU:O	1:C:141:ASN:ND2	1.65	1.30
4:D:1294:G1P:P	4:D:1294:G1P:C1	2.18	1.30
2:B:142:ASP:OD2	2:B:145:ARG:NH1	1.61	1.30
2:B:161:GLU:CD	6:B:2055:HOH:O	1.73	1.25
2:B:161:GLU:OE1	6:B:2055:HOH:O	1.54	1.19

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	288/293 (98%)	286 (99%)	1 (0%)	1 (0%)	41	31
1	C	288/293 (98%)	282 (98%)	4 (1%)	2 (1%)	22	12
1	D	288/293 (98%)	284 (99%)	4 (1%)	0	100	100
2	B	288/293 (98%)	280 (97%)	8 (3%)	0	100	100
All	All	1152/1172 (98%)	1132 (98%)	17 (2%)	3 (0%)	41	31

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	32	TYR
1	C	32	TYR
1	C	266	VAL

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	244/247 (99%)	232 (95%)	12 (5%)	25	15
1	C	243/247 (98%)	231 (95%)	12 (5%)	25	15
1	D	243/247 (98%)	234 (96%)	9 (4%)	34	25
2	B	242/247 (98%)	223 (92%)	19 (8%)	12	5
All	All	972/988 (98%)	920 (95%)	52 (5%)	22	13

5 of 52 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	246	ARG
1	C	79	GLN
1	D	130	GLU
2	B	260	ARG
1	C	9	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 18 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	59	GLN
1	D	230	HIS
1	D	77	ASN
1	C	79	GLN
1	D	27	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

17 ligands 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	#$|Z| > 2$
5	SO4	A	1296	-	4,4,4	0.19	0	6,6,6	0.76	0
4	G1P	C	1294	-	15,16,16	0.96	1 (6%)	23,24,24	1.46	3 (13%)
3	THM	A	1292	-	18,18,18	2.18	6 (33%)	26,26,26	2.90	11 (42%)
5	SO4	B	1295	-	4,4,4	0.37	0	6,6,6	0.21	0
3	THM	B	1293	-	18,18,18	2.26	7 (38%)	26,26,26	2.65	10 (38%)
3	THM	C	1293	-	18,18,18	2.23	4 (22%)	26,26,26	2.70	9 (34%)
4	G1P	D	1294	-	15,16,16	8.00	2 (13%)	23,24,24	5.50	8 (34%)
5	SO4	C	1295	-	4,4,4	0.32	0	6,6,6	0.46	0
3	THM	D	1293	-	18,18,18	2.56	8 (44%)	26,26,26	2.67	12 (46%)
3	THM	A	1293	-	18,18,18	2.22	5 (27%)	26,26,26	3.28	9 (34%)
4	G1P	B	1294	-	15,16,16	1.10	1 (6%)	23,24,24	1.18	1 (4%)
3	THM	B	1292	-	18,18,18	2.33	7 (38%)	26,26,26	2.80	12 (46%)
3	THM	D	1292	-	18,18,18	2.35	6 (33%)	26,26,26	2.95	11 (42%)
5	SO4	D	1295	-	4,4,4	0.35	0	6,6,6	0.57	0
3	THM	C	1292	-	18,18,18	2.44	6 (33%)	26,26,26	2.15	8 (30%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	#|Z| > 2	Counts	RMSZ	#|Z| > 2
4	G1P	A	1294	-	15,16,16	1.34	3 (20%)	23,24,24	3.78	9 (39%)
5	SO4	A	1295	-	4,4,4	0.53	0	6,6,6	1.66	2 (33%)

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
4	G1P	C	1294	-	-	0/7/27/27	0/1/1/1
3	THM	A	1292	-	-	1/6/18/18	0/2/2/2
3	THM	B	1293	-	-	0/6/18/18	0/2/2/2
3	THM	C	1293	-	-	1/6/18/18	0/2/2/2
4	G1P	D	1294	-	1/1/6/6	3/7/27/27	0/1/1/1
3	THM	D	1293	-	-	0/6/18/18	0/2/2/2
3	THM	A	1293	-	-	0/6/18/18	0/2/2/2
4	G1P	B	1294	-	-	2/7/27/27	0/1/1/1
3	THM	B	1292	-	-	2/6/18/18	0/2/2/2
3	THM	D	1292	-	-	2/6/18/18	0/2/2/2
3	THM	C	1292	-	-	1/6/18/18	0/2/2/2
4	G1P	A	1294	-	-	3/7/27/27	0/1/1/1

The worst 5 of 56 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	D	1294	G1P	P-O1	-30.78	1.01	1.59
3	B	1292	THM	O4-C4	5.96	1.34	1.23
3	A	1293	THM	O2-C2	5.81	1.33	1.23
3	C	1292	THM	O2-C2	5.69	1.33	1.23
3	D	1292	THM	O4-C4	5.61	1.34	1.23

The worst 5 of 105 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	D	1294	G1P	O5-C1-O1	24.54	143.44	111.36
4	A	1294	G1P	O3-C3-C2	-9.24	88.99	110.35
3	A	1293	THM	C5-C4-N3	8.26	122.36	115.31
3	A	1293	THM	C4-N3-C2	-7.53	117.60	127.35
3	A	1292	THM	O4-C4-C5	-6.83	116.98	124.90

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
4	D	1294	G1P	C1

5 of 15 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	1294	G1P	C2-C1-O1-P
4	A	1294	G1P	O5-C1-O1-P
4	D	1294	G1P	O5-C1-O1-P
4	B	1294	G1P	C4-C5-C6-O6
4	B	1294	G1P	O5-C5-C6-O6

There are no ring outliers.

8 monomers are involved in 35 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	D	1294	G1P	2	0
5	C	1295	SO4	4	0
3	D	1293	THM	1	0
4	B	1294	G1P	1	0
3	B	1292	THM	9	0
3	D	1292	THM	4	0
3	C	1292	THM	5	0
4	A	1294	G1P	13	0

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 G1P C 1294

Bond lengths

Bond angles

Torsions

Rings

Ligand G1P D 1294

Bond lengths

Bond angles

Torsions

Rings

Ligand G1P B 1294

Bond lengths

Bond angles

Torsions

Rings

Ligand G1P A 1294

Bond lengths

Bond angles

Torsions

Rings
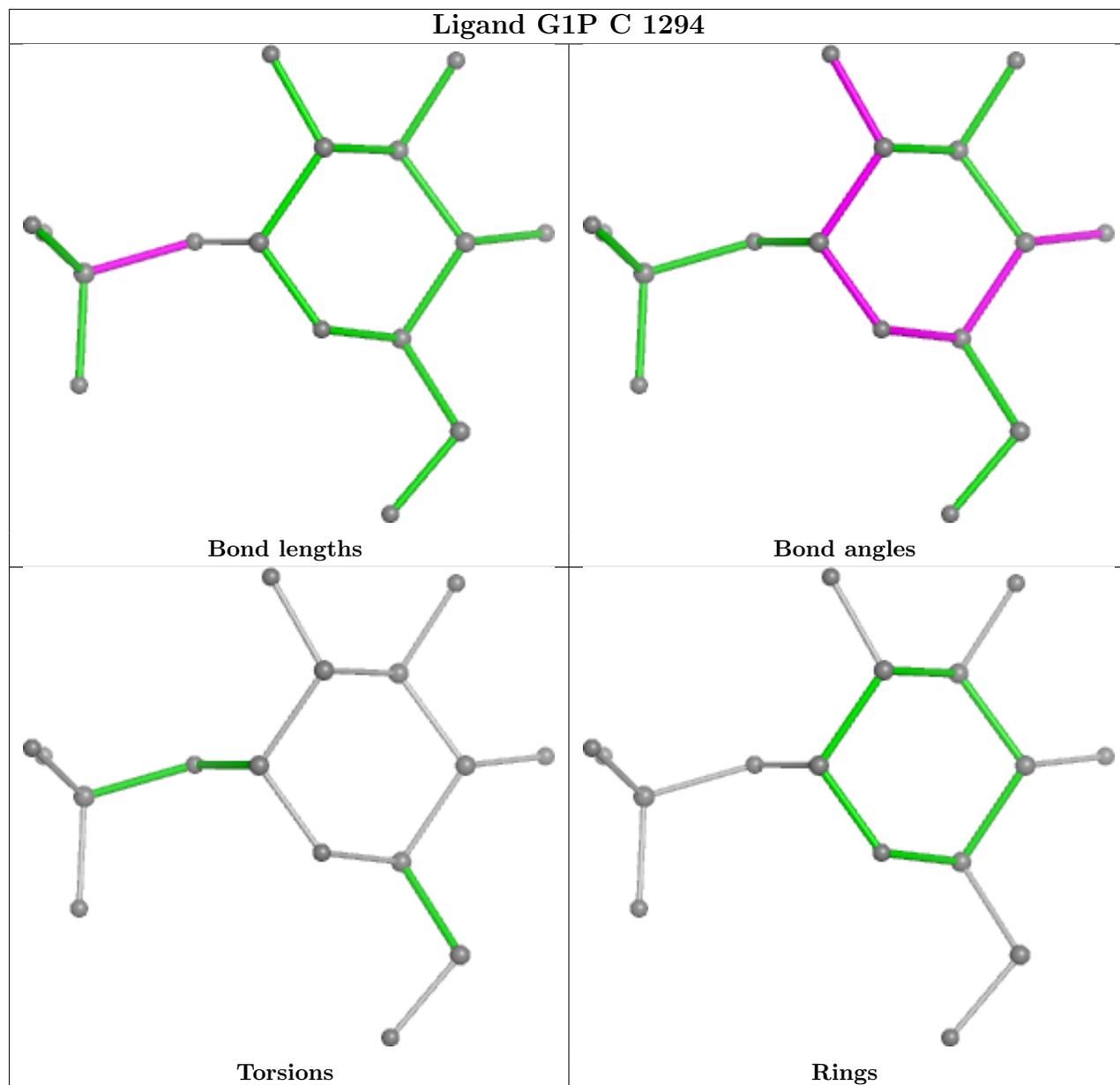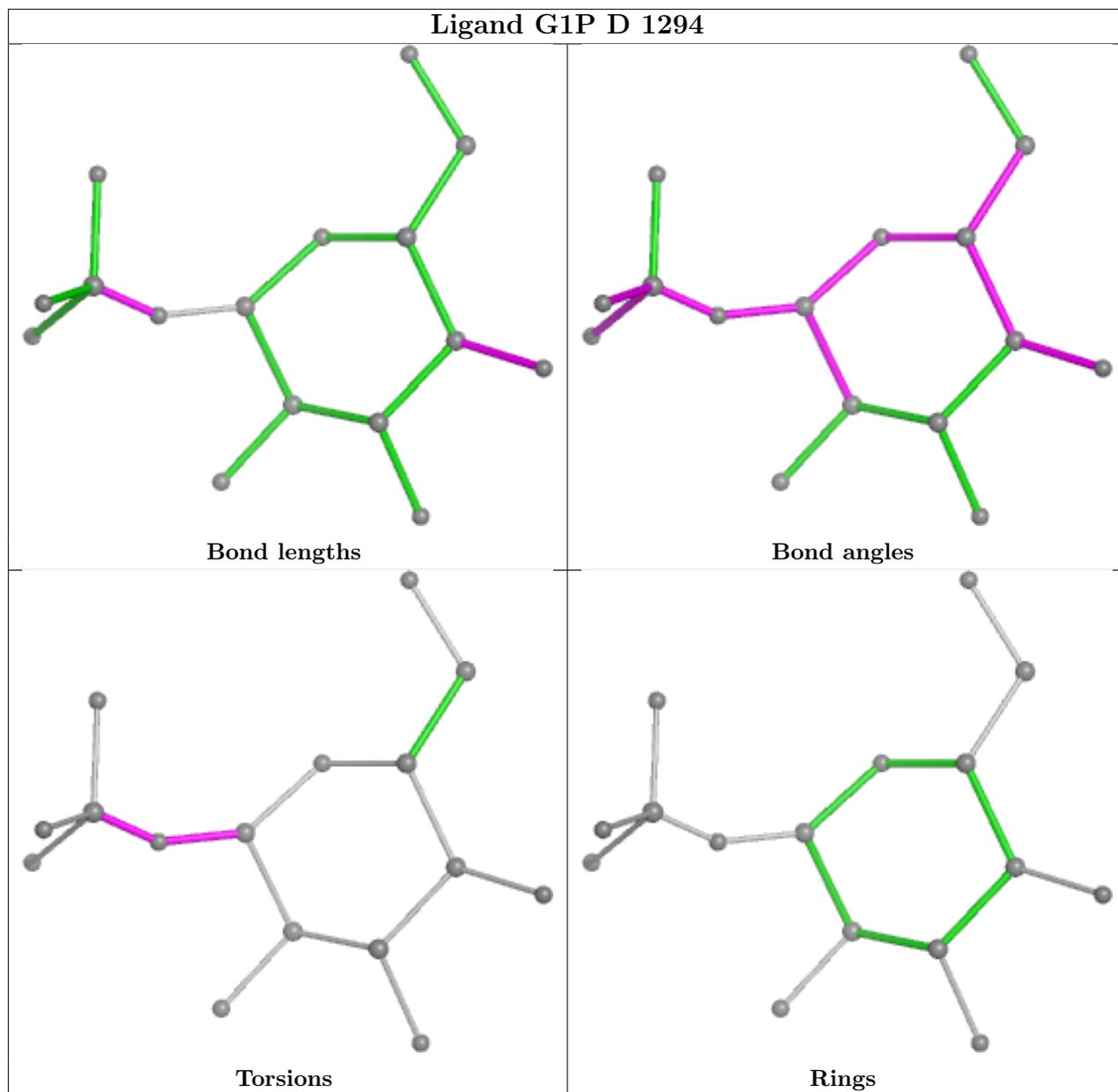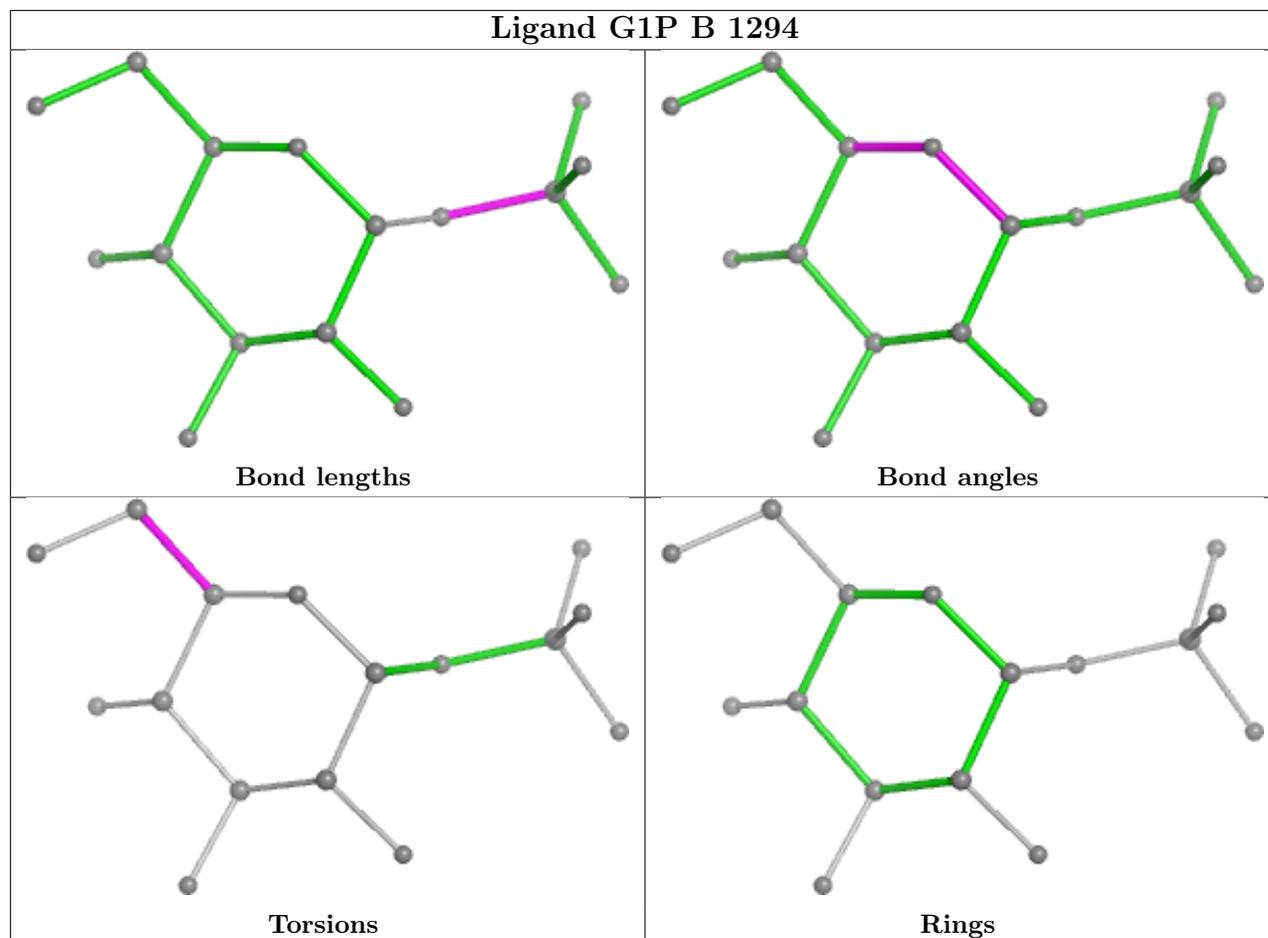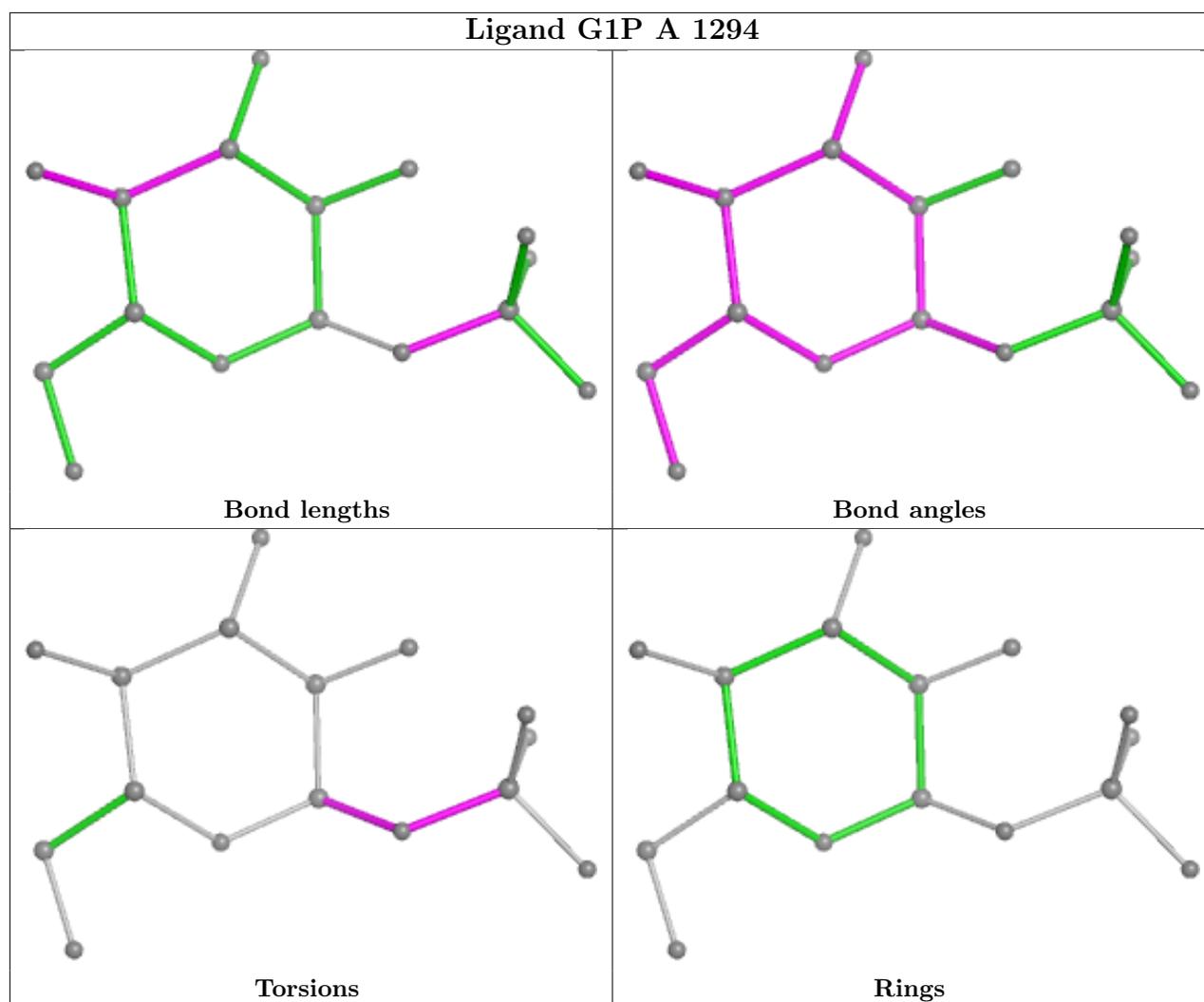
5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

6 Fit of model and data ⓘ

6.1 Protein, DNA and RNA chains ⓘ

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates ⓘ

EDS was not executed - this section is therefore empty.

6.4 Ligands ⓘ

EDS was not executed - this section is therefore empty.

6.5 Other polymers ⓘ

EDS was not executed - this section is therefore empty.