WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Dec 2, 2021 – 03:12 pm GMT

PDB ID : 7OTN
Title : HIV-1 REVERSE TRANSCRIPTASE COMPLEX WITH DNA AND IN-HIBITOR RMC-247
Authors : Martinez, S.E.; Singh, A.K.; Gu, W.; Das, K.
Deposited on : 2021-06-10
Resolution : 3.40 Å(reported)

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

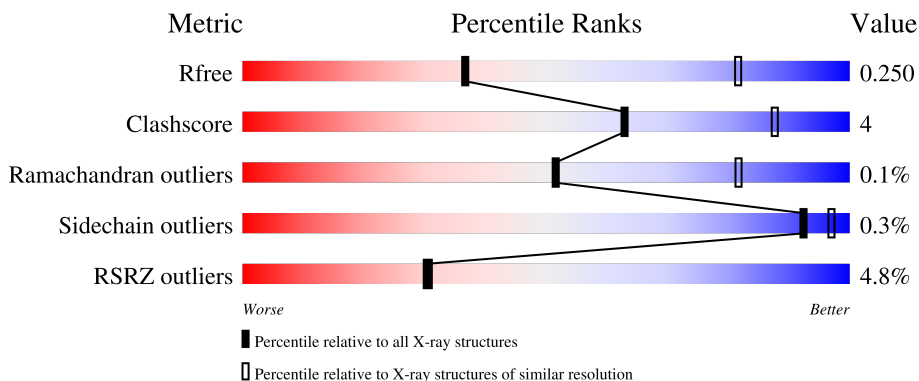|		
MolProbity	:	4.02b-467
Mogul	:	1.8.4 (270009), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 3.40 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		0.250
Clashscore		4
Ramachandran outliers		0.1%
Sidechain outliers		0.3%
RSRZ outliers		4.8%

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	1026 (3.48-3.32)
Clashscore	141614	1055 (3.48-3.32)
Ramachandran outliers	138981	1038 (3.48-3.32)
Sidechain outliers	138945	1038 (3.48-3.32)
RSRZ outliers	127900	2173 (3.50-3.30)

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	556	5% — 85% / 15%
1	C	556	7% — 88% / 12%
2	B	428	5% — 86% / 10% •
2	D	428	2% — 86% / 10% •
3	E	27	4% — 67% / 19% / 15%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	T	27	63% 26% 11%
4	F	21	76% 19% 5%
4	P	21	62% 33% 5%

2 Entry composition ⓘ

There are 7 unique types of molecules in this entry. The entry contains 17643 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 Reverse transcriptase/ribonuclease H.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	555	Total 4509	C 2919	N 750	O 832	S 8	0	0	0
1	A	556	Total 4521	C 2927	N 752	O 833	S 9	0	0	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-1	MET	-	initiating methionine	UNP P03366
C	0	VAL	-	expression tag	UNP P03366
C	258	CYS	GLN	conflict	UNP P03366
C	280	SER	CYS	conflict	UNP P03366
C	498	ASN	ASP	conflict	UNP P03366
A	-1	MET	-	initiating methionine	UNP P03366
A	0	VAL	-	expression tag	UNP P03366
A	258	CYS	GLN	conflict	UNP P03366
A	280	SER	CYS	conflict	UNP P03366
A	498	ASN	ASP	conflict	UNP P03366

- Molecule 2 is a protein called Reverse transcriptase/ribonuclease H.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	411	Total 3398	C 2213	N 563	O 616	S 6	0	0	0
2	B	412	Total 3391	C 2206	N 560	O 619	S 6	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	280	SER	CYS	conflict	UNP P03366
B	280	SER	CYS	conflict	UNP P03366

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

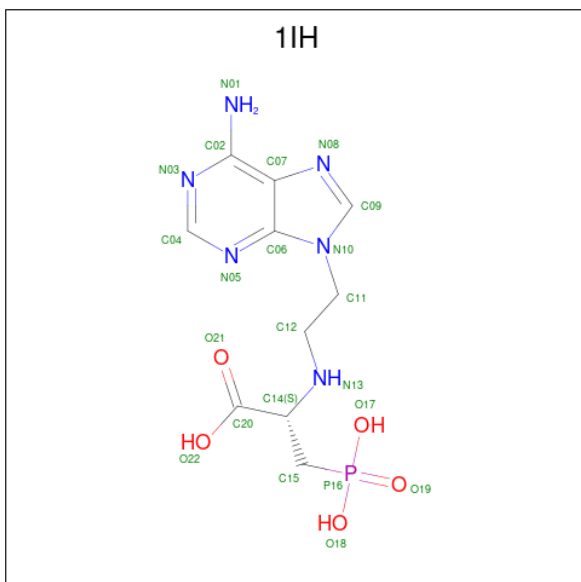| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	23	Total 477	C 223	N 95	O 136	P 23	0	0	0
3	T	24	Total 494	C 233	N 97	O 141	P 23	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	F	20	Total 403	C 192	N 72	O 120	P 19	0	0	0
4	P	20	Total 403	C 192	N 72	O 120	P 19	0	0	0

- Molecule 5 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	C	1	Total 1	Mn 1	0	0
5	A	1	Total 1	Mn 1	0	0

- Molecule 6 is (S)-2-((2-(6-amino-9H-purin-9-yl)ethyl)amino)-3-phosphonopropanoic acid (three-letter code: 1IH) (formula: $C_{10}H_{15}N_6O_5P$) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
6	C	1	Total 22	C 10	N 6	O 5	P 1	0	0
6	A	1	Total 22	C 10	N 6	O 5	P 1	0	0

- Molecule 7 is water.

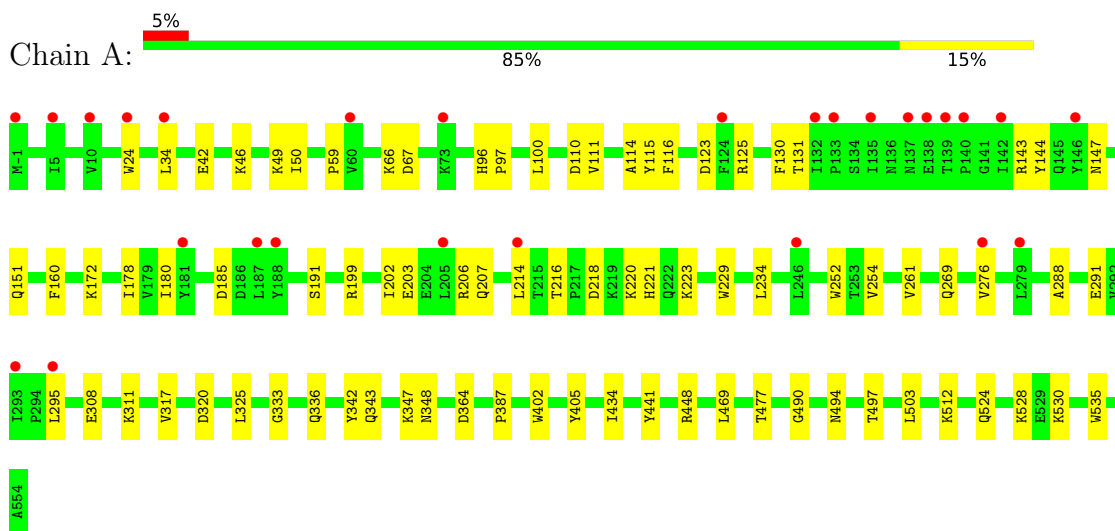| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	D	1	Total 1	O 1	0	0

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.
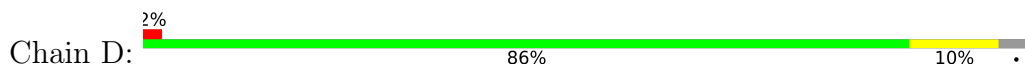
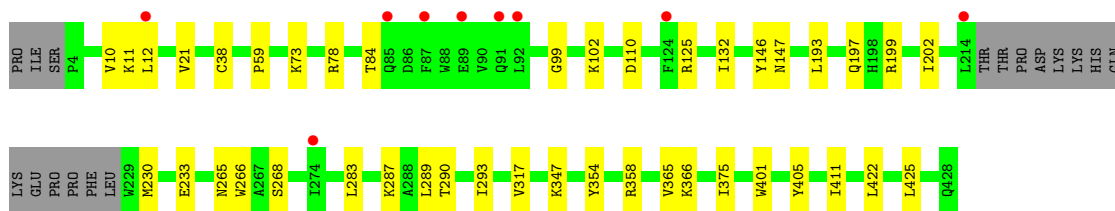- Molecule 1: Reverse transcriptase/ribonuclease H
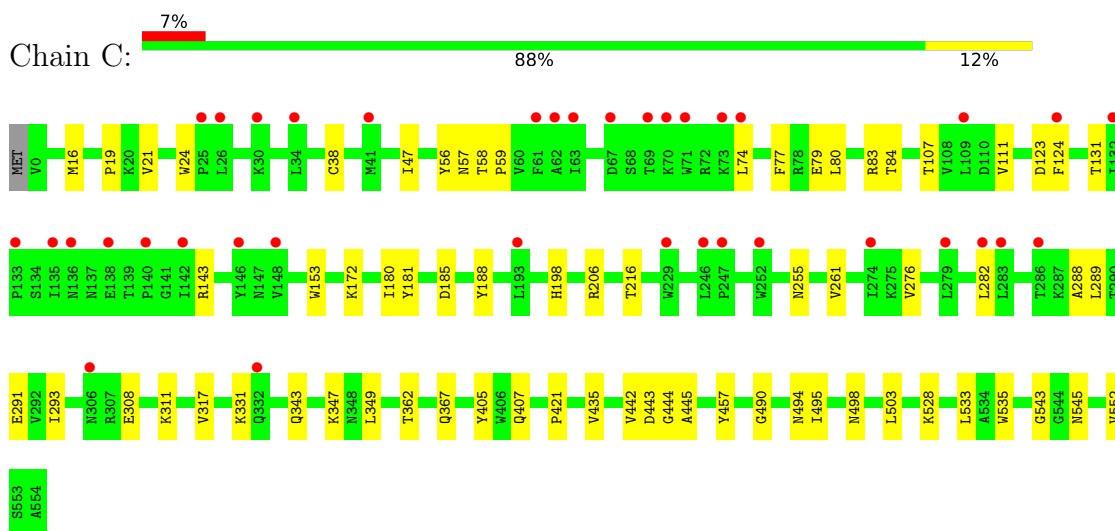
Chain C:
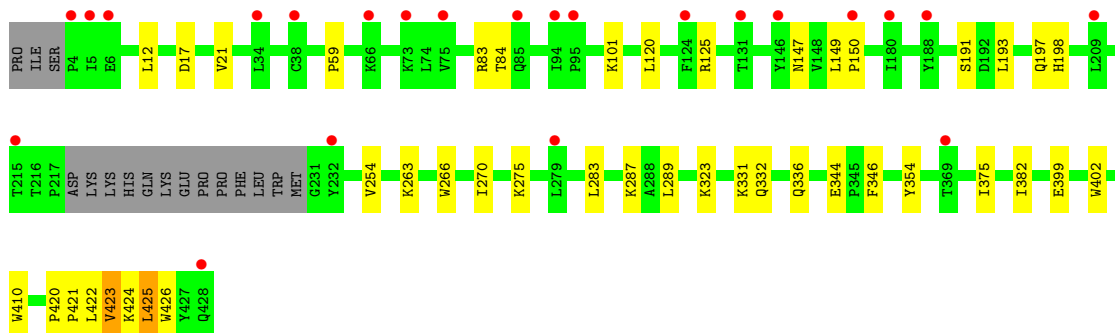
- Molecule 1: Reverse transcriptase/ribonuclease H
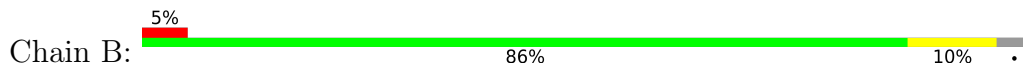

Chain A:

- Molecule 2: Reverse transcriptase/ribonuclease H

Chain D:

• Molecule 2: Reverse transcriptase/ribonuclease H

Chain B:

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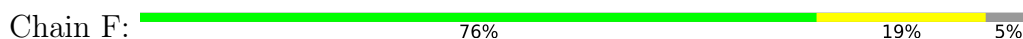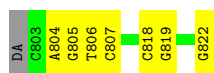

Chain E:

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

Chain T:

• Molecule 4: DNA (5'-D(*CP*AP*GP*TP*CP*CP*CP*TP*GP*TP*TP*CP*GP*GP*(MRG) P*CP*GP*CP*CP*(DDG))-3')

Chain F:

• Molecule 4: DNA (5'-D(*CP*AP*GP*TP*CP*CP*CP*TP*GP*TP*TP*CP*GP*GP*(MRG) P*CP*GP*CP*CP*(DDG))-3')

Chain P:

62% 33% 5%
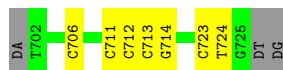
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	89.82Å 132.35Å 138.62Å 90.00° 98.45° 90.00°	Depositor
Resolution (Å)	95.23 – 3.40 95.23 – 3.40	Depositor EDS
% Data completeness (in resolution range)	99.7 (95.23-3.40) 99.8 (95.23-3.40)	Depositor EDS
R_{merge}	0.19	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.47 (at 3.41Å)	Xtriage
Refinement program	PHENIX 1.19.1_4122	Depositor
R, R_{free}	0.214 , 0.254 0.213 , 0.250	Depositor DCC
R_{free} test set	1272 reflections (2.88%)	wwPDB-VP
Wilson B-factor (Å^2)	110.5	Xtriage
Anisotropy	0.184	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< |L| > = 0.44, < L^2 > = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	17643	wwPDB-VP
Average B, all atoms (Å^2)	133.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.29% 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: DDG, MRG, 1IH, MN

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	0.24	0/4639	0.44	0/6303
1	C	0.23	0/4627	0.44	0/6289
2	B	0.23	0/3488	0.46	2/4740 (0.0%)
2	D	0.23	0/3496	0.45	0/4749
3	E	0.63	0/536	0.82	0/826
3	T	0.66	0/555	0.85	0/856
4	F	0.61	0/400	0.84	2/612 (0.3%)
4	P	0.80	1/400 (0.2%)	0.99	1/612 (0.2%)
All	All	0.31	1/18141 (0.0%)	0.51	5/24987 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	P	807	DC	C4'-O4'	8.11	1.53	1.45

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	F	820	DC	O4'-C4'-C3'	-6.30	101.98	104.50
4	P	807	DC	C1'-O4'-C4'	-6.10	104.00	110.10
2	B	425	LEU	CA-CB-CG	5.56	128.08	115.30
4	F	806	DT	P-O3'-C3'	5.55	126.36	119.70
2	B	425	LEU	CB-CG-CD2	-5.41	101.81	111.00

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	4521	0	4587	50	0
1	C	4509	0	4567	43	0
2	B	3391	0	3413	28	0
2	D	3398	0	3424	24	0
3	E	477	0	256	3	0
3	T	494	0	269	6	0
4	F	403	0	224	1	0
4	P	403	0	224	5	0
5	A	1	0	0	0	0
5	C	1	0	0	0	0
6	A	22	0	0	2	0
6	C	22	0	0	0	0
7	D	1	0	0	0	0
All	All	17643	0	16964	150	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 4.

The worst 5 of 150 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:263:LYS:HE3	2:B:425:LEU:HB2	1.67	0.77
1:A:50:ILE:HG12	1:A:143:ARG:HB2	1.71	0.71
1:C:111:VAL:HG22	1:C:185:ASP:HB2	1.75	0.69
1:C:84:THR:HG21	1:C:153:TRP:HE1	1.61	0.66
1:C:131:THR:HG22	1:C:143:ARG:HG2	1.78	0.63

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 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	554/556 (100%)	532 (96%)	22 (4%)	0	100	100
1	C	553/556 (100%)	538 (97%)	15 (3%)	0	100	100
2	B	408/428 (95%)	391 (96%)	16 (4%)	1 (0%)	47	78
2	D	407/428 (95%)	387 (95%)	20 (5%)	0	100	100
All	All	1922/1968 (98%)	1848 (96%)	73 (4%)	1 (0%)	51	82

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	423	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	497/497 (100%)	495 (100%)	2 (0%)	91	95
1	C	495/497 (100%)	494 (100%)	1 (0%)	93	98
2	B	372/390 (95%)	370 (100%)	2 (0%)	88	94
2	D	372/390 (95%)	371 (100%)	1 (0%)	92	97
All	All	1736/1774 (98%)	1730 (100%)	6 (0%)	92	97

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	405	TYR
2	B	410	TRP
2	B	422	LEU
2	D	422	LEU
1	C	123	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	474	ASN
1	C	545	ASN

5.3.3   RNA ⓘ

There are no RNA molecules in this entry.

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

4 non-standard protein/DNA/RNA residues are modelled in this entry.

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.

5.5   Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6   Ligand geometry ⓘ

Of 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

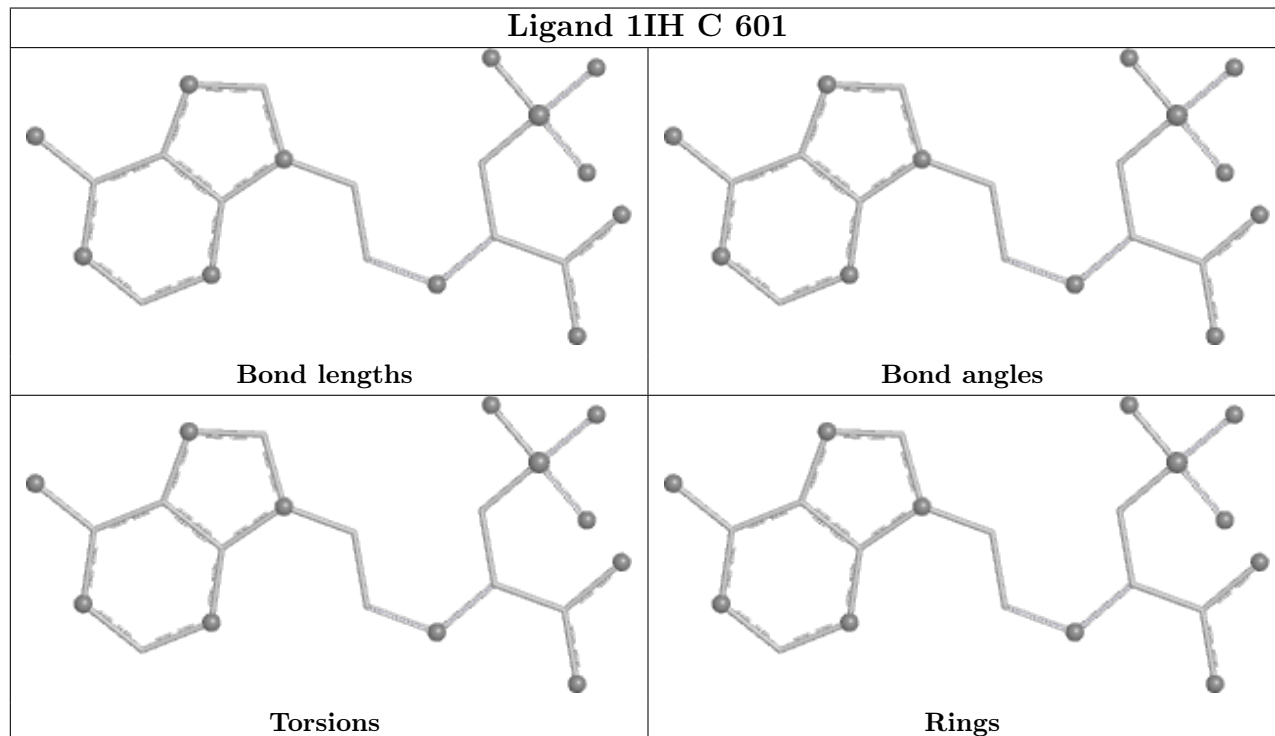There are no chirality outliers.
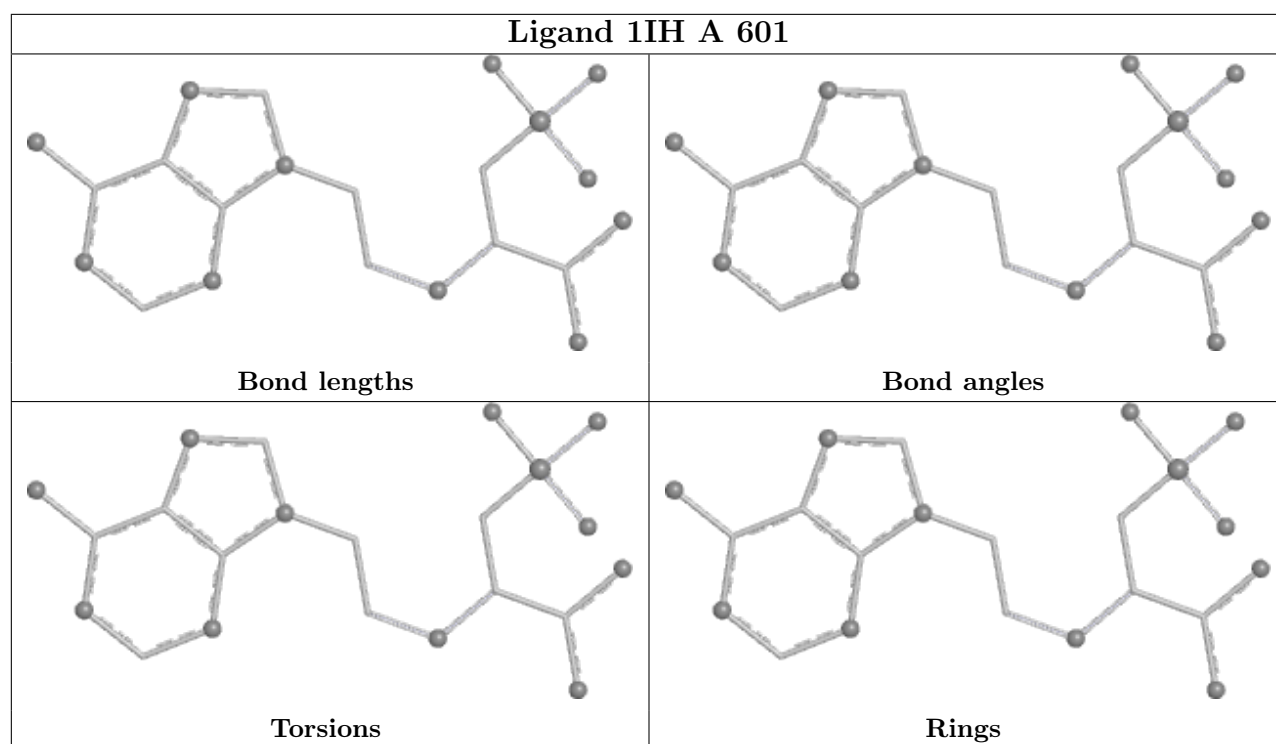
There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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 1IH C 601

Bond lengths

Bond angles

Torsions

Rings

Ligand 1IH A 601

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 ⓘ

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(Å²)	Q<0.9
1	A	556/556 (100%)	0.12	27 (4%)	29	29	56, 143, 212, 242	0
1	C	555/556 (99%)	0.34	37 (6%)	17	19	52, 134, 213, 246	0
2	B	412/428 (96%)	0.45	23 (5%)	24	25	60, 122, 174, 200	0
2	D	411/428 (96%)	0.21	9 (2%)	62	60	55, 109, 159, 184	0
3	E	23/27 (85%)	-0.22	1 (4%)	35	35	141, 167, 215, 272	0
3	T	24/27 (88%)	-0.51	0	100	100	136, 167, 245, 258	0
4	F	18/21 (85%)	-0.47	0	100	100	121, 166, 193, 199	0
4	P	18/21 (85%)	-0.78	0	100	100	116, 157, 201, 202	0
All	All	2017/2064 (97%)	0.24	97 (4%)	30	31	52, 127, 201, 272	0

The worst 5 of 97 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	4	PRO	10.8
1	C	133	PRO	7.2
1	C	26	LEU	6.2
1	C	229	TRP	5.8
1	A	132	ILE	5.8

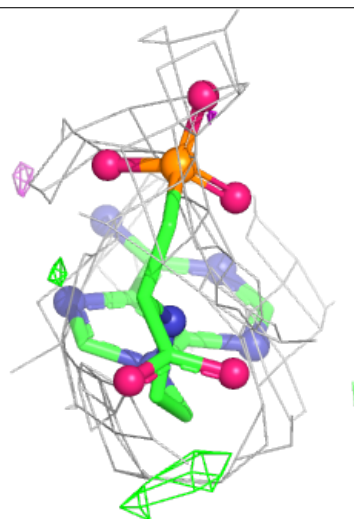## 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

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(Å²)	Q<0.9
4	DDG	P	822	21/22	0.82	0.13	147,166,179,184	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors($Å^2$)	Q<0.9
4	DDG	F	822	21/22	0.83	0.16	128,160,169,176	0
4	MRG	F	817	22/27	0.89	0.18	144,178,205,225	0
4	MRG	P	817	22/27	0.91	0.17	147,181,197,225	0

6.3 Carbohydrates ⓘ

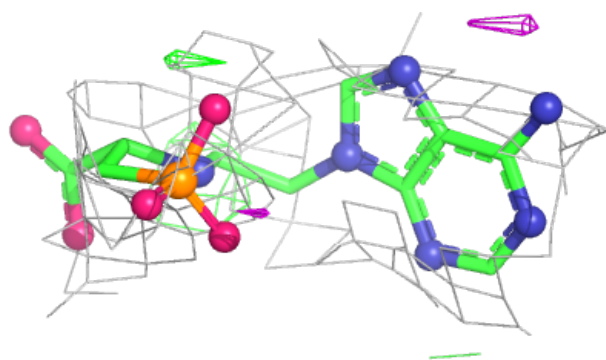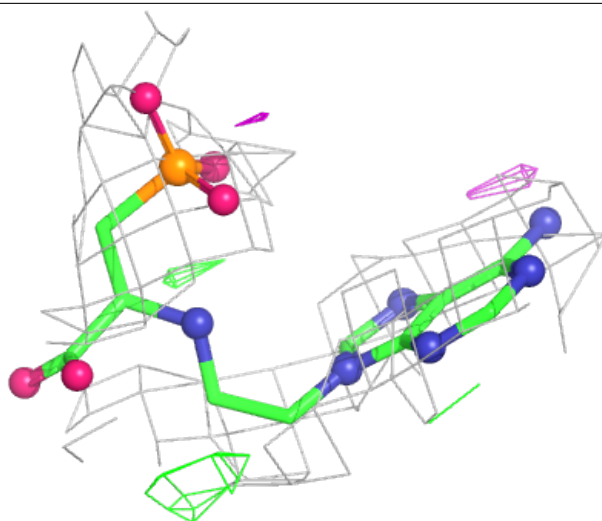There are no monosaccharides in this entry.

6.4 Ligands ⓘ

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.
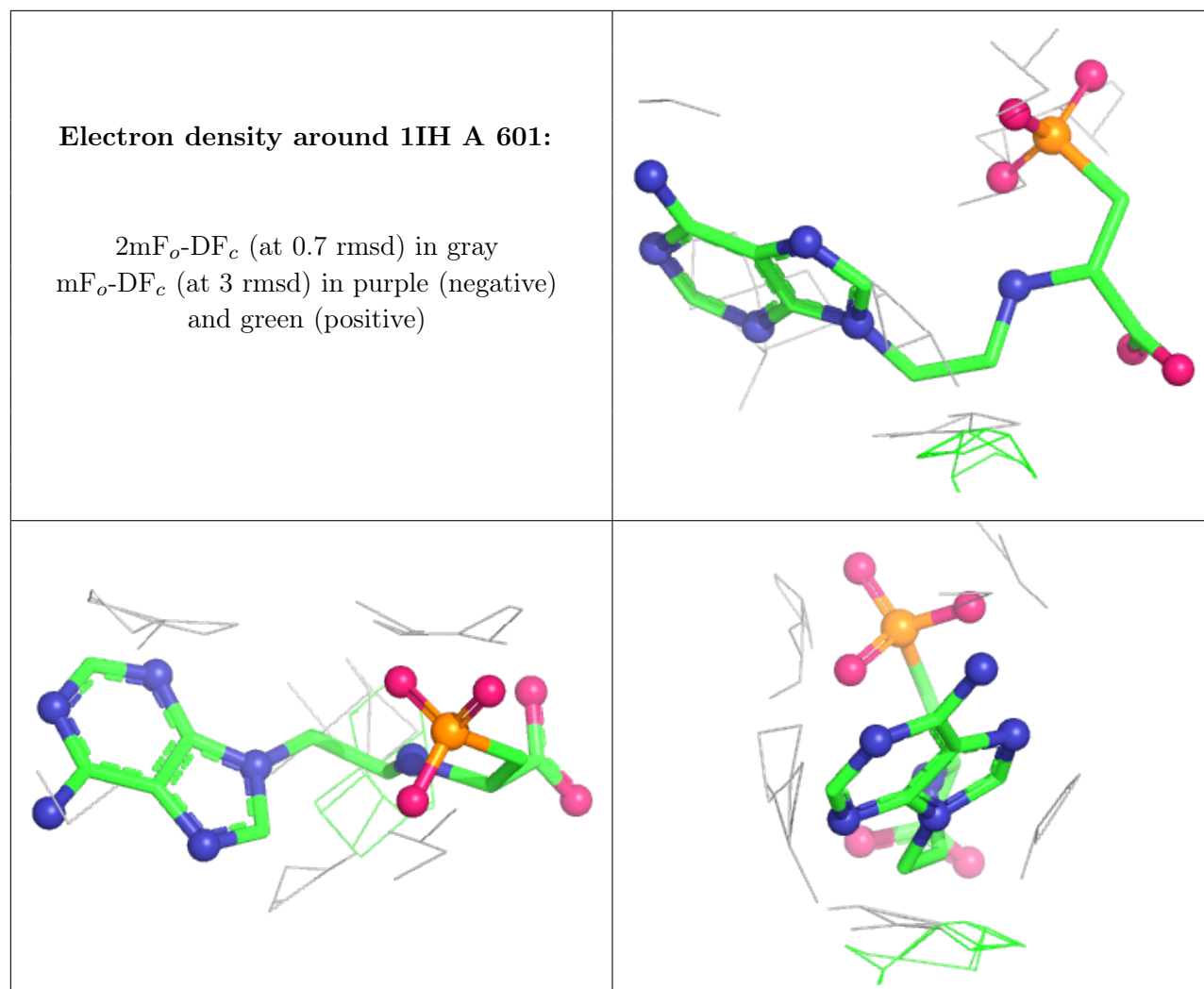
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors($Å^2$)	Q<0.9
6	1IH	C	601	22/22	0.64	0.22	163,204,228,252	0
5	MN	C	600	1/1	0.73	0.10	239,239,239,239	0
6	1IH	A	601	22/22	0.85	0.10	165,179,212,231	0
5	MN	A	600	1/1	0.98	0.13	198,198,198,198	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around 1IH C 601:

$2mF_o\text{-}DF_c$ (at 0.7 rmsd) in gray
$mF_o\text{-}DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)

Electron density around 1IH A 601:

$2mF_o\text{-}DF_c$ (at 0.7 rmsd) in gray
$mF_o\text{-}DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)

6.5   Other polymers ⓘ

There are no such residues in this entry.