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)

May 19, 2020 – 07:00 am BST

PDB ID	:	3VNV
Title	:	Complex structure of viral RNA polymerase II
Authors	:	Takeshita, D.; Tomita, K.
Deposited on	:	2012-01-18
Resolution	:	2.60 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 (i) symbol.

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

|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
buster-report	:	1.1.7 (2018)
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 (i)

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

The reported resolution of this entry is 2.60 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.
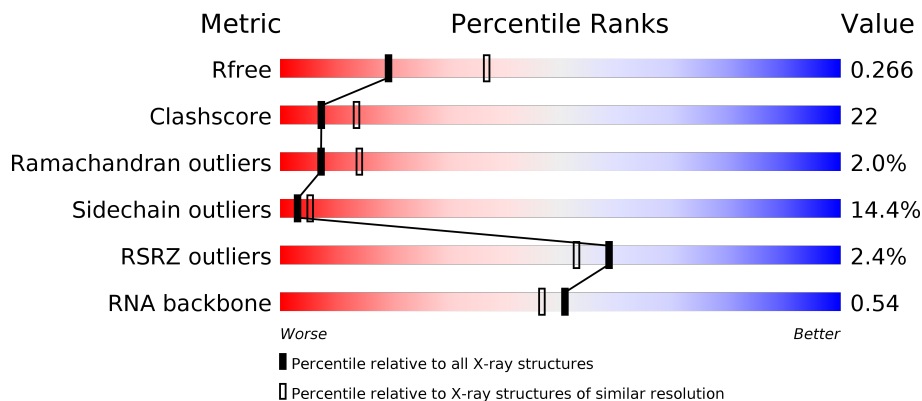
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)
RNA backbone	3102	1040 (2.90-2.30)

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 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 <=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	1289	2% / 55% / 31% / 7% / 7%
2	G	7	43% / 43% / 14%
3	T	8	13% / 38% / 63%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 9680 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 Elongation factor Ts, Elongation factor Tu, LINKER, Q beta replicase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1203	Total 9287	C 5865	N 1605	O 1772	S 45	0	0	0

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	284	HIS	-	LINKER	UNP P0A6P3
A	1284	HIS	-	EXPRESSION TAG	UNP Q8LTE0
A	1285	HIS	-	EXPRESSION TAG	UNP Q8LTE0
A	1286	HIS	-	EXPRESSION TAG	UNP Q8LTE0
A	1287	HIS	-	EXPRESSION TAG	UNP Q8LTE0
A	1288	HIS	-	EXPRESSION TAG	UNP Q8LTE0
A	1289	HIS	-	EXPRESSION TAG	UNP Q8LTE0

- Molecule 2 is a RNA chain called RNA (5'-R(*CP*CP*CP*UP*AP*CP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	G	7	Total 139	C 64	N 22	O 47	P 6	0	0	0

- Molecule 3 is a RNA chain called RNA (5'-R(*GP*GP*GP*UP*AP*GP*GP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	T	8	Total 177	C 79	N 37	O 54	P 7	8	0	0

- Molecule 4 is 3'-DEOXY-CYTIDINE-5'-TRIPHOSPHATE (three-letter code: CH1) (formula: $C_9H_{16}N_3O_{13}P_3$).

CH1
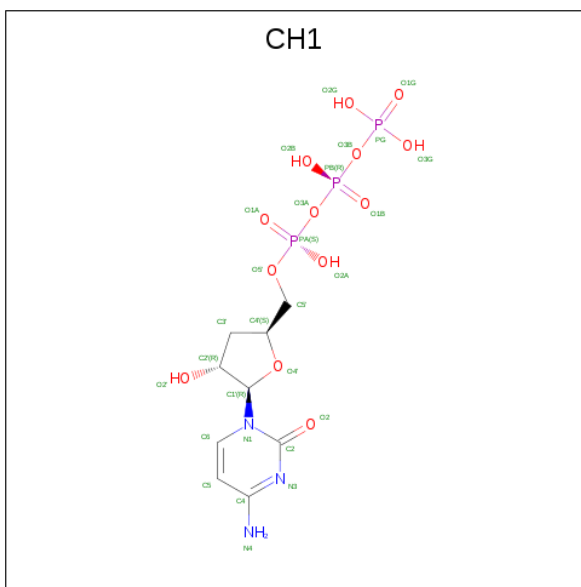
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
4	A	1	28	9	3	13	3	0	0

- Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Ca		
5	A	2	2	2	0	0

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
6	A	47	47	47	0	0

3 Residue-property plots ⓘ

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.
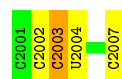
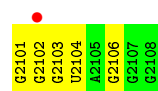• Molecule 1: Elongation factor Ts, Elongation factor Tu, LINKER, Q beta replicase

Chain A:

S887 T888 H889 F890 D891 I892 R893 I894 S895 D896 I897 S898 P899 F900 N901 T905 K908 N909 S910 K911 T912 D913 R914 C915 I916 G921 W922 N923 M924 Q927 L928 G929 I930 I933 R937 L938 D947 Q948 T949 I950 N951 Q952 R953 R954 A955 H956 E957 G958 S959 V960 T961 N962 N963

V967 D968 L969 S970 A971 A972 I976 C981 L984 L985 W989 F990 E991 V992 L993 M994 D995 L996 R997 G1001 R1002 L1003 P1004 E1012 K1013 I1014 S1015 S1016 M1017 N1019 G1020 Y1021 T1022 L1033 A1034 R1035 D1044 T1049 V1050 D1053 D1054 I1055 I1056 L1057 P1058 V1068 F1069

K1070 Y1071 V1072 G1073 F1074 T1075 T1076 N1077 T1081 F1087 R1088 E1089 K1093 H1094 Y1095 D1100 V1101 T1102 Y1105 V1111 L1116 N1121 Y1124 T1128 K1143 Y1144 P1149 K1150 Q1151 L1152 Q1153 R1154 I1157 A1165 L1166 V1167 N1173 P1174 F1175 A1176 K1177 I1182 R1183 Y1184

V1185 P1186 R1193 L1200 L1204 Y1205 D1206 L1207 R1210 N1216 ASP GLY LEU PRO LEU ARG GLY PRO SER GLY CYS ASP SER ALA ASP LEU PHE A1234 Q1237 L1238 I1239 C1240 R1241 S1242 N1243 P1244 T1245 K1246 R1249 S1250 T1251 F1254 R1264 VAL LEU ALA PRO TYR GLY VAL PHE

GLN GLY THR LYS ALA VAL SER LEU HIS GLU ALA HIS HIS HIS HIS HIS HIS

• Molecule 2: RNA (5'-R(*CP*CP*CP*UP*AP*CP*C)-3')

Chain G:

43%   43%   14%

C2001 C2002 C2003 U2004 C2007

• Molecule 3: RNA (5'-R(*GP*GP*GP*UP*AP*GP*GP*G)-3')

13%

Chain T:

38%   63%

G2101 G2102 G2103 U2104 A2105 G2106 G2107 G2108

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	139.73Å 256.04Å 101.72Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.94 – 2.60 48.08 – 2.60	Depositor EDS
% Data completeness (in resolution range)	99.3 (19.94-2.60) 99.3 (48.08-2.60)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.25 (at 2.61Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.6.4_486)	Depositor
R, R_{free}	0.220 , 0.279 0.209 , 0.266	Depositor DCC
R_{free} test set	2826 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å^2)	49.5	Xtriage
Anisotropy	0.360	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 40.3	EDS
L-test for twinning [2]	$< |L| > = 0.45, < L^2 > = 0.28$	Xtriage
Estimated twinning fraction	0.024 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.024 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	9680	wwPDB-VP
Average B, all atoms (Å^2)	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.82% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: CH1, CA

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	0.49	1/9456 (0.0%)	0.63	0/12787
2	G	0.46	0/153	0.84	0/235
3	T	0.64	0/199	1.20	0/311
All	All	0.50	1/9808 (0.0%)	0.65	0/13333

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

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	981	CYS	CB-SG	-6.00	1.72	1.81

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	849	SER	Peptide

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	9287	0	9273	421	0
2	G	139	0	78	5	0
3	T	177	0	89	7	0
4	A	28	0	12	2	0
5	A	2	0	0	0	0
6	A	47	0	0	2	0
All	All	9680	0	9452	428	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 22.

The worst 5 of 428 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:G:2003:C:H42	3:T:2106:G:H1	1.07	0.98
1:A:1157:ILE:HG12	1:A:1165:ALA:HB3	1.43	0.97
1:A:1100:ASP:OD1	1:A:1102:THR:HG23	1.62	0.97
1:A:961:THR:HG22	1:A:963:ASN:H	1.28	0.97
1:A:741:ASN:HD21	1:A:745:GLU:HB2	1.33	0.93

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	1193/1289 (93%)	1079 (90%)	90 (8%)	24 (2%)	7	14

5 of 24 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	266	GLU
1	A	285	MET
1	A	325	GLY
1	A	806	ARG
1	A	972	ALA

5.3.2 Protein sidechains (i)

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	995/1060 (94%)	852 (86%)	143 (14%)	3 5

5 of 143 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	705	LEU
1	A	816	SER
1	A	1152	LEU
1	A	717	ILE
1	A	785	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 28 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	729	ASN
1	A	857	ASN
1	A	1151	GLN
1	A	741	ASN
1	A	763	ASN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	G	6/7 (85%)	2 (33%)	0
3	T	7/8 (87%)	2 (28%)	0

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
All	All	13/15 (86%)	4 (30%)	0

All (4) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	G	2003	C
2	G	2004	U
3	T	2103	G
3	T	2104	U

There are no RNA pucker outliers to report.

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 carbohydrates 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	#$|Z| > 2$
4	CH1	A	2501	5	22,29,29	3.15	4 (18%)	27,45,45	1.65	7 (25%)

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	CH1	A	2501	5	-	7/20/34/34	0/2/2/2

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	2501	CH1	C2'-C1'	-12.33	1.42	1.54
4	A	2501	CH1	C3'-C4'	-4.99	1.41	1.52
4	A	2501	CH1	C4-N4	4.80	1.49	1.35
4	A	2501	CH1	C3'-C2'	-2.01	1.47	1.52

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	2501	CH1	C2-N3-C4	3.82	120.21	116.34
4	A	2501	CH1	O3G-PG-O3B	3.49	116.34	104.64
4	A	2501	CH1	N4-C4-N3	3.29	121.70	116.49
4	A	2501	CH1	C5-C4-N3	-2.70	118.61	121.72
4	A	2501	CH1	O4'-C4'-C3'	2.21	107.93	105.07

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	2501	CH1	C3'-C4'-C5'-O5'
4	A	2501	CH1	O4'-C4'-C5'-O5'
4	A	2501	CH1	PB-O3B-PG-O3G
4	A	2501	CH1	PB-O3A-PA-O5'
4	A	2501	CH1	PB-O3B-PG-O2G

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	2501	CH1	2	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 CH1 A 2501

Bond lengths

Bond angles

Torsions

Rings
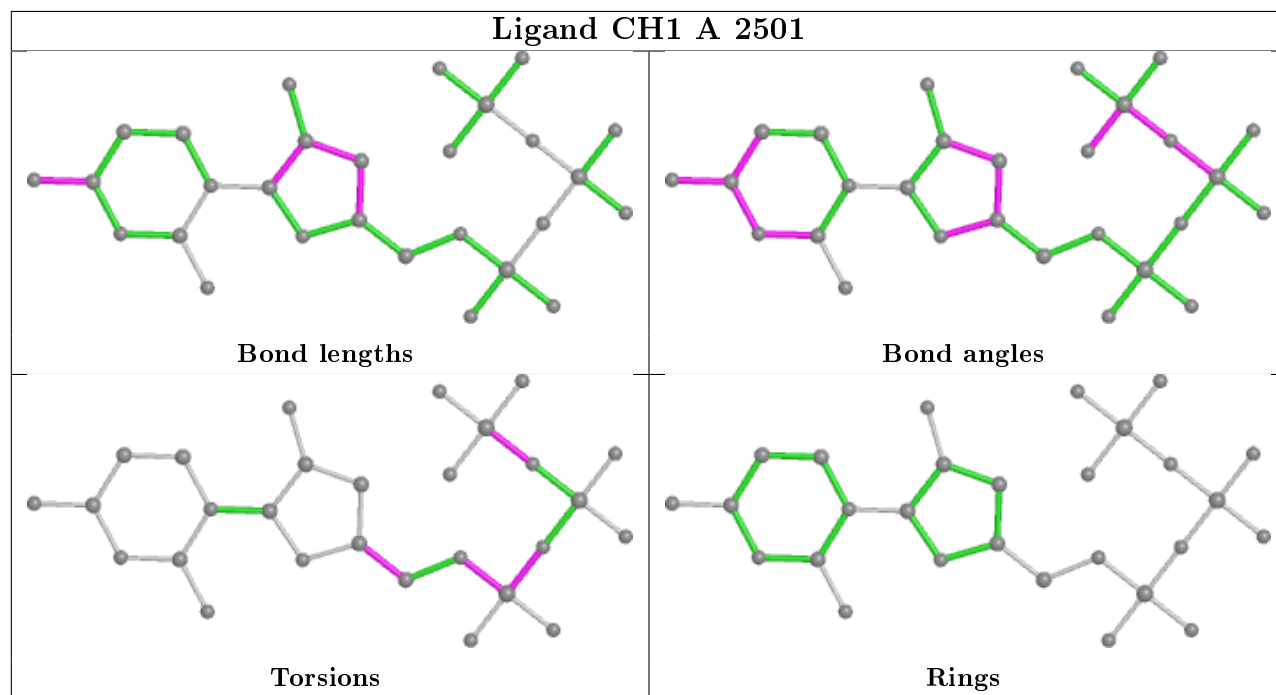
5.7 Other polymers (i)

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	1203/1289 (93%)	-0.11	28 (2%)	60	54	16, 49, 97, 122	0
2	G	7/7 (100%)	-0.25	0	100	100	59, 76, 99, 105	0
3	T	8/8 (100%)	0.62	1 (12%)	3	2	45, 69, 117, 122	1 (12%)
All	All	1218/1304 (93%)	-0.10	29 (2%)	59	53	16, 50, 99, 122	1 (0%)

The worst 5 of 29 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	700	SER	5.3
1	A	701	SER	5.3
1	A	8	LEU	4.7
1	A	189	GLU	4.7
1	A	193	LYS	4.4

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates (i)

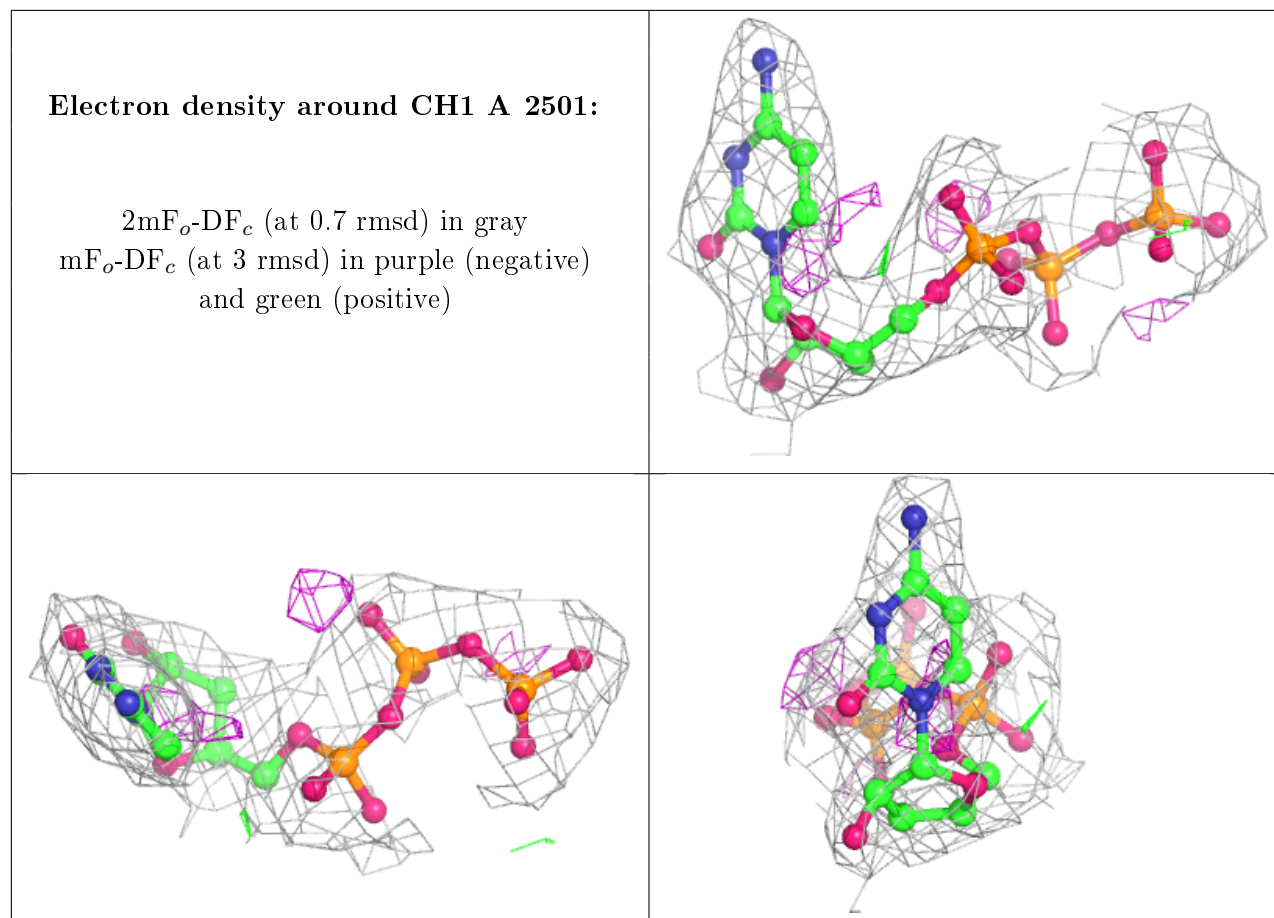There are no carbohydrates in this entry.

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(Å^2)	Q<0.9
5	CA	A	2503	1/1	0.92	0.20	53,53,53,53	0
4	CH1	A	2501	28/28	0.96	0.14	35,48,53,57	0
5	CA	A	2502	1/1	0.99	0.17	31,31,31,31	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 CH1 A 2501:

$2mF_o\text{-}DF_c$ (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.