wwPDB X-ray Structure Validation Summary Report (i)

Jan 4, 2021 – 01:05 pm GMT

PDB ID	:	7AIJ
Title	:	HIV-1 REVERSE TRANSCRIPTASE COMPLEX WITH DNA AND L-METHIONINE TENOFOVIR
Authors	:	Gu, W.; Martinez, S.E.; Nguyen, H.; Xu, H.; Herdewijn, P.; de Jonghe, S.; Das, K.
Deposited on	:	2020-09-27
Resolution	:	2.95 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:

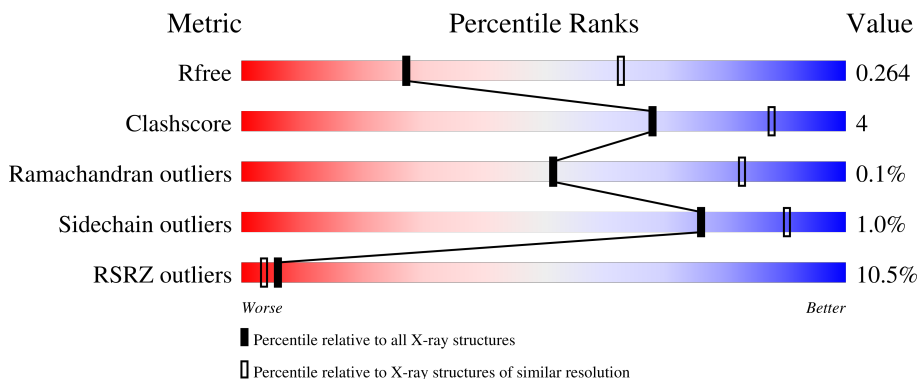|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.16
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.16

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 2.95 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.
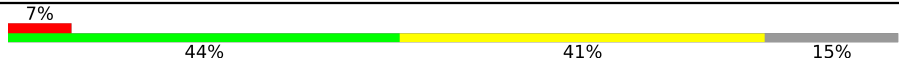
Metric	Percentile Ranks	Value
Rfree		0.264
Clashscore		4
Ramachandran outliers		0.1%
Sidechain outliers		1.0%
RSRZ outliers		10.5%

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	3104 (3.00-2.92)
Clashscore	141614	3462 (3.00-2.92)
Ramachandran outliers	138981	3340 (3.00-2.92)
Sidechain outliers	138945	3343 (3.00-2.92)
RSRZ outliers	127900	2986 (3.00-2.92)

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	16% / 90% / 9%
1	C	556	15% / 88% / 12%
2	B	428	2% / 87% / 8% · ·
2	D	428	5% / 85% / 11% ·
3	E	27	7% / 48% / 37% / 15%

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	T	27	7% green 44% yellow 41% grey 15%
4	F	21	green 62% yellow 33% grey 5%
4	P	21	green 76% yellow 19% grey 5%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 17687 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 Gag-Pol polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	555	Total 4513	C 2922	N 751	O 832	S 8	0	0	0
1	C	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
A	-1	MET	-	initiating methionine	UNP P03366
A	0	VAL	-	expression tag	UNP P03366
A	258	CYS	GLN	engineered mutation	UNP P03366
A	280	SER	CYS	engineered mutation	UNP P03366
A	498	ASN	ASP	engineered mutation	UNP P03366
C	-1	MET	-	initiating methionine	UNP P03366
C	0	VAL	-	expression tag	UNP P03366
C	258	CYS	GLN	engineered mutation	UNP P03366
C	280	SER	CYS	engineered mutation	UNP P03366
C	498	ASN	ASP	engineered mutation	UNP P03366

- Molecule 2 is a protein called Gag-Pol polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	411	Total 3401	C 2215	N 563	O 616	S 7	0	0	0
2	D	412	Total 3400	C 2212	N 563	O 619	S 6	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

| Chain	Residue	Modelled	Actual	Comment	Reference
B	280	SER	CYS	engineered mutation	UNP P03366
D	280	SER	CYS	engineered mutation	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	T	23	Total 477	C 223	N 95	O 136	P 23	0	0	0
3	E	23	Total 477	C 223	N 95	O 136	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	P	20	Total 403	C 192	N 72	O 120	P 19	0	0	0
4	F	20	Total 403	C 192	N 72	O 120	P 19	0	0	0

- Molecule 5 is L-Methionine Tenofovir (three-letter code: RFE) (formula: $C_{14}H_{23}N_6O_5PS$) (labeled as "Ligand of Interest" by depositor).
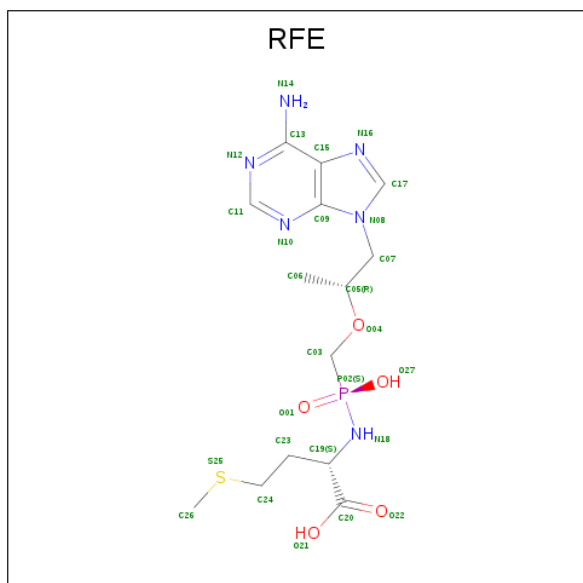
| Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
5	C	1	Total 27	C 14	N 6	O 5	P 1	S 1	0	0

- Molecule 6 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	17	Total 17	O 17	0	0
6	B	17	Total 17	O 17	0	0
6	C	14	Total 14	O 14	0	0
6	D	10	Total 10	O 10	0	0
6	T	4	Total 4	O 4	0	0
6	E	2	Total 2	O 2	0	0
6	F	1	Total 1	O 1	0	0

3 Residue-property plots (i)

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.
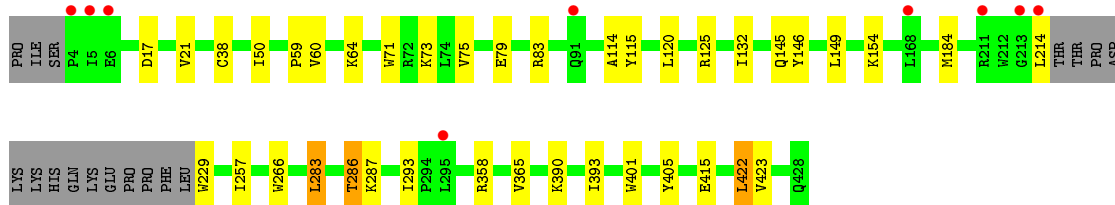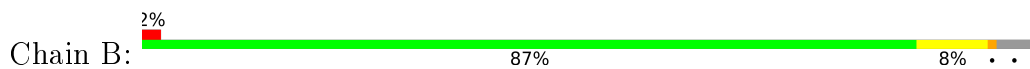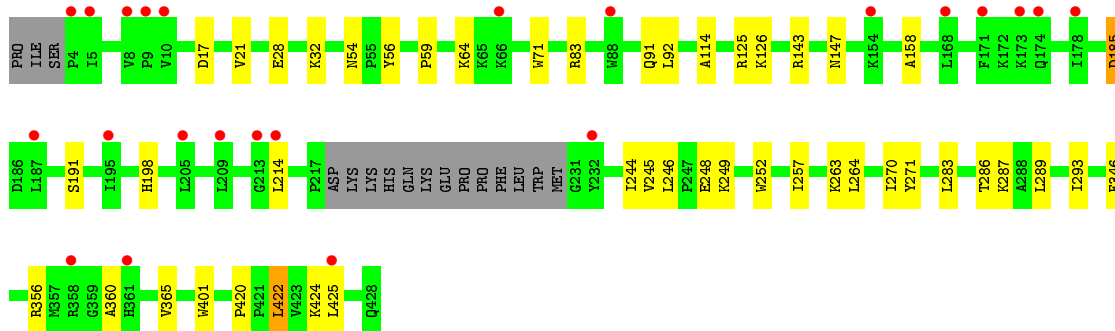
- Molecule 1: Gag-Pol polyprotein
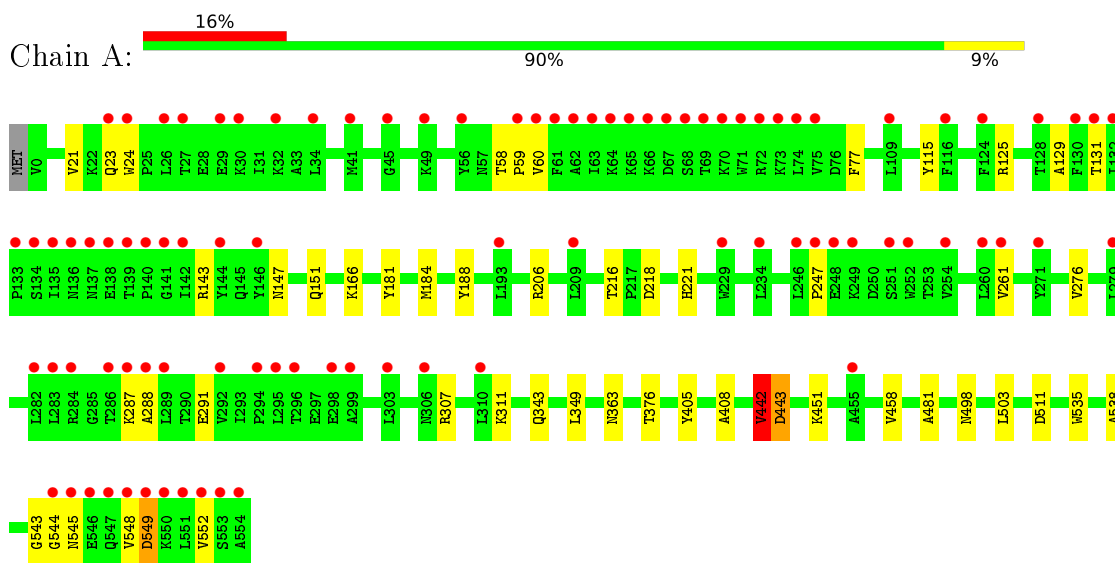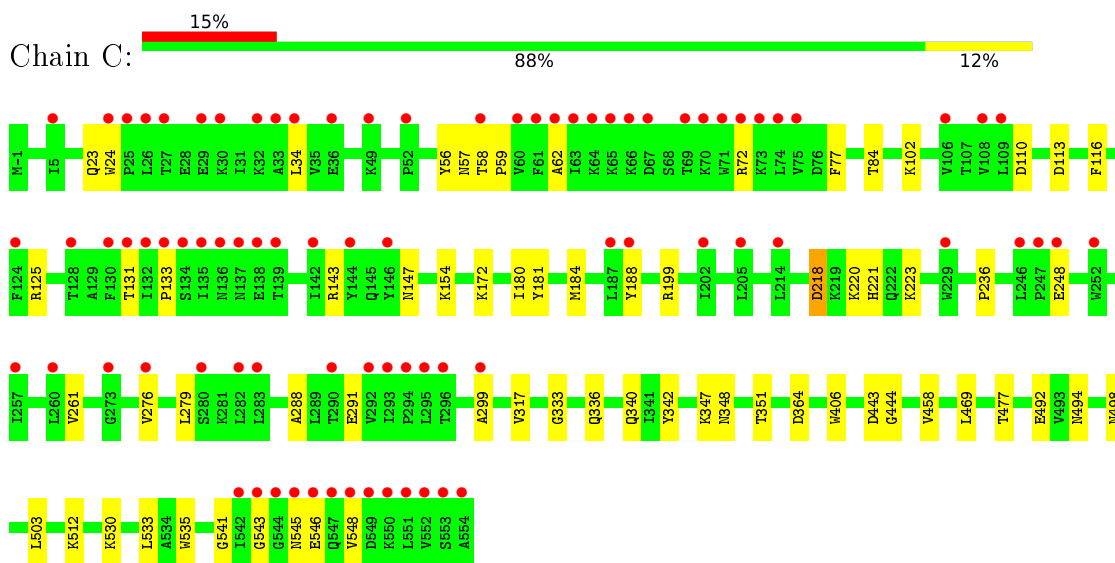
Chain A:

- Molecule 1: Gag-Pol polyprotein

Chain C:

- Molecule 2: Gag-Pol polyprotein

Chain B:

• Molecule 2: Gag-Pol polyprotein

Chain D:

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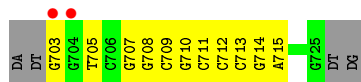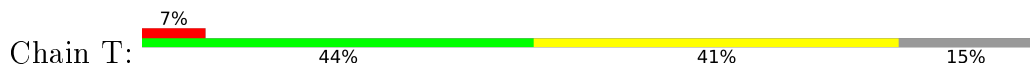
• Molecule 3: DNA (5'-D(P*GP*GP*TP*CP*GP*GP*CP*GP*CP*CP*CP*GP*AP*AP*CP*A
P*GP*GP*GP*AP*CP*TP*G)-3')
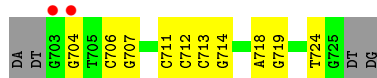
Chain E:

• Molecule 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:

● 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:

4 Data and refinement statistics (i)

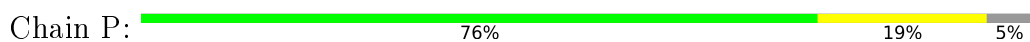| Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	89.62Å 132.93Å 139.98Å 90.00° 98.50° 90.00°	Depositor
Resolution (Å)	69.22 – 2.95 69.22 – 2.95	Depositor EDS
% Data completeness (in resolution range)	99.5 (69.22-2.95) 99.5 (69.22-2.95)	Depositor EDS
R_{merge}	0.16	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.39 (at 2.96Å)	Xtriage
Refinement program	PHENIX 1.18.2_3874	Depositor
R, R_{free}	0.226 , 0.263 0.226 , 0.264	Depositor DCC
R_{free} test set	1970 reflections (2.89%)	wwPDB-VP
Wilson B-factor (Å^2)	85.4	Xtriage
Anisotropy	0.272	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 60.8	EDS
L-test for twinning[2]	$< |L| > = 0.48, < L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	17687	wwPDB-VP
Average B, all atoms (Å^2)	114.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.76% 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: MRG, DDG, RFE

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	0.23	0/4631	0.40	0/6293
1	C	0.23	0/4639	0.40	0/6303
2	B	0.23	0/3499	0.40	0/4752
2	D	0.23	0/3497	0.40	0/4751
3	E	0.48	0/536	0.78	0/826
3	T	0.49	0/536	0.78	0/826
4	F	0.55	0/400	0.91	0/612
4	P	0.52	0/400	0.89	0/612
All	All	0.27	0/18138	0.47	0/24975

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 (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	4513	0	4578	33	0
1	C	4521	0	4587	37	0
2	B	3401	0	3431	25	0
2	D	3400	0	3433	26	0
3	E	477	0	256	6	0

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	T	477	0	256	7	0
4	F	403	0	224	6	0
4	P	403	0	224	3	0
5	C	27	0	0	1	0
6	A	17	0	0	1	0
6	B	17	0	0	0	0
6	C	14	0	0	0	0
6	D	10	0	0	0	0
6	E	2	0	0	0	0
6	F	1	0	0	0	0
6	T	4	0	0	0	0
All	All	17687	0	16989	129	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 129 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:C:543:GLY:N	2:D:283:LEU:O	2.26	0.67
1:C:56:TYR:O	1:C:143:ARG:NH2	2.29	0.65
1:A:131:THR:HG22	1:A:143:ARG:HG2	1.79	0.65
1:C:181:TYR:HB2	1:C:188:TYR:HB3	1.79	0.64
1:A:288:ALA:HB3	1:A:291:GLU:HB2	1.79	0.63

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	553/556 (100%)	534 (97%)	18 (3%)	1 (0%)	47	79

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	554/556 (100%)	536 (97%)	18 (3%)	0	100	100
2	B	407/428 (95%)	390 (96%)	17 (4%)	0	100	100
2	D	408/428 (95%)	388 (95%)	20 (5%)	0	100	100
All	All	1922/1968 (98%)	1848 (96%)	73 (4%)	1 (0%)	51	83

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	442	VAL

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	496/497 (100%)	491 (99%)	5 (1%)	76	90
1	C	497/497 (100%)	494 (99%)	3 (1%)	86	94
2	B	373/390 (96%)	368 (99%)	5 (1%)	69	87
2	D	374/390 (96%)	369 (99%)	5 (1%)	69	87
All	All	1740/1774 (98%)	1722 (99%)	18 (1%)	76	90

5 of 18 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	422	LEU
2	B	423	VAL
2	D	248	GLU
2	B	283	LEU
2	B	286	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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 (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

1 ligand is 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.

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.
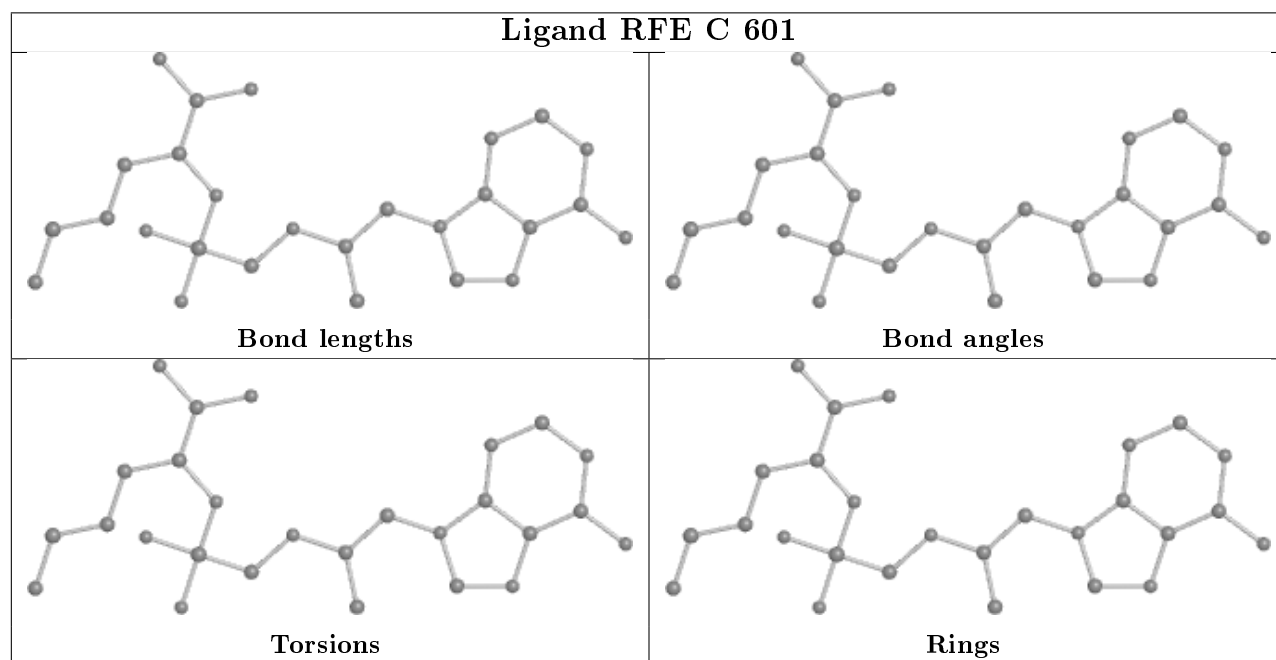
5.7 Other polymers (i)

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	555/556 (99%)	1.00	91 (16%)	1	1	41, 116, 214, 272	0
1	C	556/556 (100%)	0.95	84 (15%)	2	1	43, 119, 211, 243	0
2	B	411/428 (96%)	0.32	9 (2%)	62	45	40, 88, 140, 169	0
2	D	412/428 (96%)	0.50	23 (5%)	24	15	52, 100, 160, 185	0
3	E	23/27 (85%)	-0.11	2 (8%)	10	6	96, 146, 209, 249	0
3	T	23/27 (85%)	-0.04	2 (8%)	10	6	108, 156, 205, 267	0
4	F	18/21 (85%)	-0.53	0	100	100	83, 129, 207, 209	0
4	P	18/21 (85%)	-0.59	0	100	100	97, 141, 198, 200	0
All	All	2016/2064 (97%)	0.69	211 (10%)	6	4	40, 106, 191, 272	0

The worst 5 of 211 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	73	LYS	13.6
1	C	63	ILE	13.2
1	A	68	SER	11.8
1	C	544	GLY	11.8
1	C	26	LEU	11.6

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
4	MRG	P	817	22/27	0.90	0.09	138,158,172,174	0

Continued on next page...

Continued from previous page...

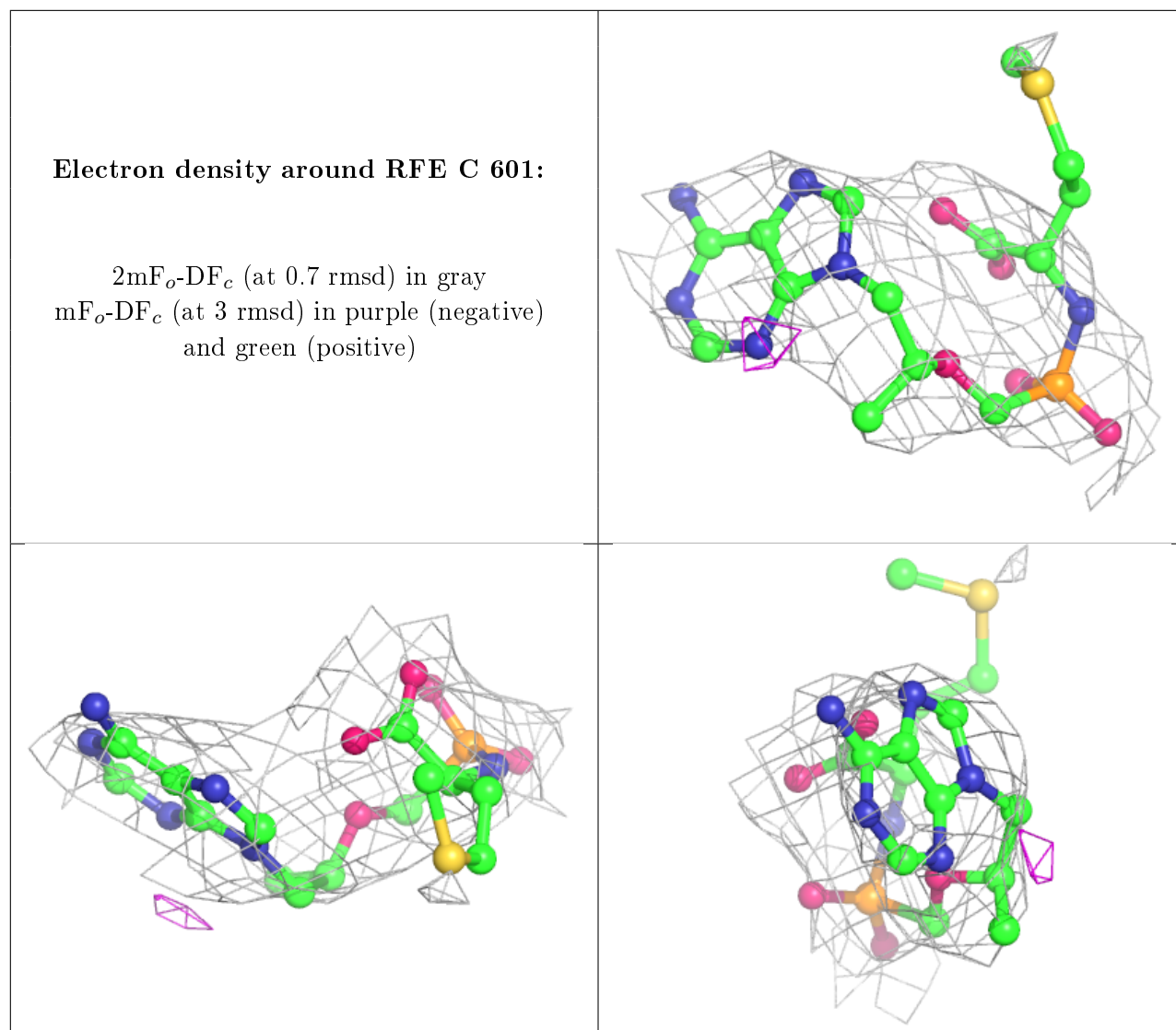| Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å²)	Q<0.9
4	MRG	F	817	22/27	0.91	0.10	124,140,155,170	0
4	DDG	F	822	21/22	0.92	0.21	103,119,126,130	0
4	DDG	P	822	21/22	0.94	0.14	104,118,127,129	0

6.3 Carbohydrates (i)

There are no monosaccharides 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(Å²)	Q<0.9
5	RFE	C	601	27/27	0.83	0.39	143,168,190,201	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 RFE C 601:

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

6.5 Other polymers (i)

There are no such residues in this entry.