



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 5, 2026 – 09:10 PM UTC

PDB ID : 6BSI / pdb_00006bsi
Title : Structure of HIV-1 RT complexed with an RNA/DNA hybrid containing the polypurine-tract sequence
Authors : Tian, L.; Kim, M.; Yang, W.
Deposited on : 2017-12-03
Resolution : 3.25 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 ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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

MolProbity : 4-5-2 with Phenix2.0
Mogul : 2022.3.0, CSD as543be (2022)
Xtriage (Phenix) : 2.0
EDS : 3.0
Buster-report : wwPDB partial adaption of 1.1.7 (2018)
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4 : 9.0.010 (Gargrove)
Density-Fitness : 1.0.12
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.49

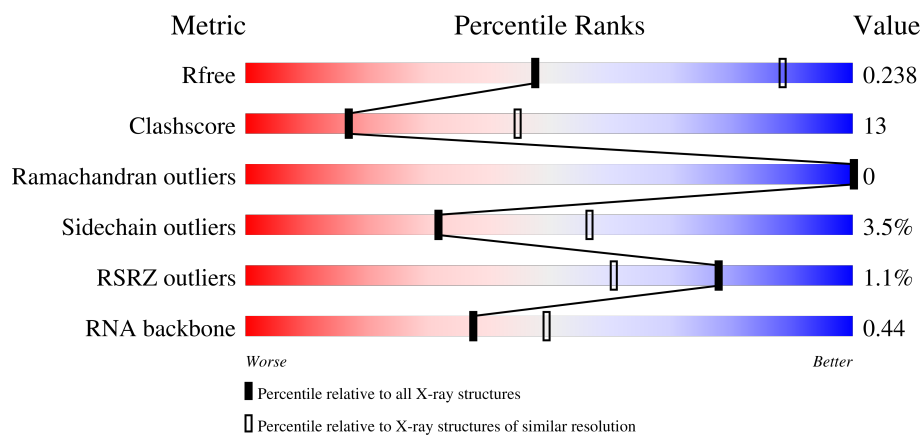
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.25 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}	180053	1605 (3.30-3.22)
Clashscore	190562	1660 (3.30-3.22)
Ramachandran outliers	187476	1630 (3.30-3.22)
Sidechain outliers	187428	1629 (3.30-3.22)
RSRZ outliers	180081	1605 (3.30-3.22)
RNA backbone	3983	1006 (3.54-2.98)

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 $\leq 5\%$. 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	558	67% 28% • •
2	B	441	67% 23% • 8%
3	D	23	57% 43%
4	R	25	44% 36% 20%

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
5	EFZ	A	601	-	X	-	-

2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 8566 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 P66 SUBUNIT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	543	Total	C	N	O	S	0	1	0
			4307	2784	703	812	8			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	GLY	-	expression tag	UNP Q74085
A	68	GLY	SER	conflict	UNP Q74085
A	83	LYS	ARG	conflict	UNP Q74085
A	357	MET	THR	conflict	UNP Q74085
A	411	VAL	ILE	conflict	UNP Q74085
A	461	LYS	ARG	conflict	UNP Q74085
A	483	HIS	TYR	conflict	UNP Q74085
A	512	GLN	LYS	conflict	UNP Q74085

- Molecule 2 is a protein called REVERSE TRANSCRIPTASE P51 SUBUNIT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	406	Total	C	N	O	S	0	0	0
			3227	2111	514	596	6			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	GLY	-	expression tag	UNP A0A076Q3N8
B	68	GLY	SER	conflict	UNP A0A076Q3N8
B	83	LYS	ARG	conflict	UNP A0A076Q3N8
B	411	VAL	ILE	conflict	UNP A0A076Q3N8

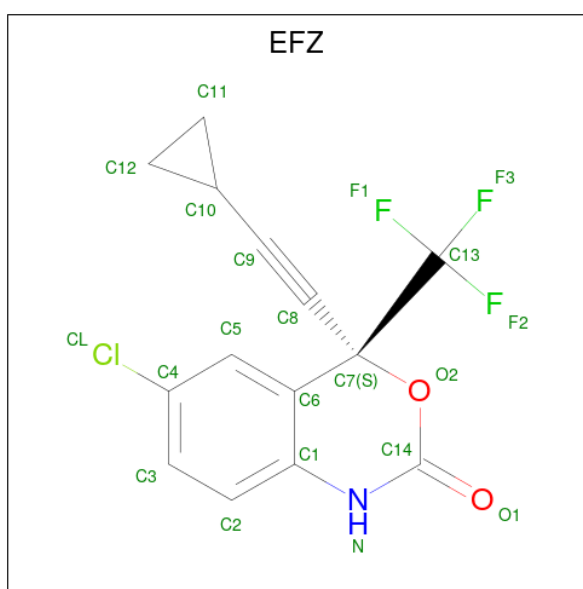
- Molecule 3 is a DNA chain called DNA (5'-D(*GP*TP*TP*TP*TP*TP*CP*TP*TP*TP*TP*GP*TP*TP*AP*TP*TP*GP*TP*GP*GP*CP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	23	Total	C	N	O	P	0	0	0
			465	227	67	149	22			

- Molecule 4 is a RNA chain called RNA (25-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	R	25	Total	C	N	O	P	0	0	0
			527	240	109	155	23			

- Molecule 5 is (-)-6-CHLORO-4-CYCLOPROPYLETHYNYL-4-TRIFLUOROMETHYL-1,4-DIHYDRO-2H-3,1-BENZOXAZIN-2-ONE (CCD ID: EFZ) (formula: C₁₄H₉ClF₃NO₂).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	A	1	Total	C	Cl	F	N	O	
			21	14	1	3	1	2	

- Molecule 6 is CALCIUM ION (CCD ID: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	1	Total Ca	0	0
			1 1		
6	D	1	Total Ca	0	0
			1 1		

- Molecule 7 is GLYCEROL (CCD ID: GOL) (formula: C₃H₈O₃).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
7	D	1	Total	C	O	0	0
			6	3	3		
7	R	1	Total	C	O	0	0
			6	3	3		

- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	4	Total	O	0	0
			4	4		
8	B	1	Total	O	0	0
			1	1		



- Molecule 3: DNA (5'-D(*GP*TP*TP*TP*TP*TP*CP*TP*TP*TP*TP*GP*TP*TP*AP*TP*TP*GP*TP*GP*GP*CP*C)-3')



- Molecule 4: RNA (25-MER)



4 Data and refinement statistics

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α , β , γ	163.85Å 163.85Å 129.57Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	47.30 – 3.25 47.30 – 3.25	Depositor EDS
% Data completeness (in resolution range)	99.7 (47.30-3.25) 92.8 (47.30-3.25)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.09	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.27 (at 3.26Å)	Xtriage
Refinement program	PHENIX (1.10_2155: ???)	Depositor
R, R_{free}	0.193 , 0.241 0.197 , 0.238	Depositor DCC
R_{free} test set	1592 reflections (4.97%)	wwPDB-VP
Wilson B-factor (Å ²)	115.9	Xtriage
Anisotropy	0.274	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 101.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.032 for -h,-k,l	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	8566	wwPDB-VP
Average B, all atoms (Å ²)	118.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.19% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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: CA, GOL, EFZ, 3DR

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	0.50	1/4423 (0.0%)	0.69	0/6039
2	B	0.42	0/3319	0.61	0/4538
3	D	0.50	0/516	0.71	0/796
4	R	0.41	0/580	0.66	0/899
All	All	0.46	1/8838 (0.0%)	0.66	0/12272

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	223	LYS	C-N	-6.14	1.20	1.33

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	4307	0	4211	121	0
2	B	3227	0	3135	71	0
3	D	465	0	269	11	0
4	R	527	0	275	17	0
5	A	21	0	9	3	0
6	A	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	D	1	0	0	0	0
7	D	6	0	8	1	0
7	R	6	0	8	0	0
8	A	4	0	0	0	0
8	B	1	0	0	0	0
All	All	8566	0	7915	208	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 13.

The worst 5 of 208 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:519:ASN:HA	1:A:522:ILE:HD12	1.52	0.89
3:D:11:DT:H3	4:R:20:A:H62	1.22	0.86
1:A:426:TRP:HB3	1:A:526:ILE:HD11	1.56	0.86
1:A:308:GLU:HA	1:A:311:LYS:HE2	1.68	0.74
2:B:111:VAL:HG21	2:B:187:LEU:HD13	1.70	0.73

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	540/558 (97%)	519 (96%)	21 (4%)	0	100	100
2	B	398/441 (90%)	369 (93%)	29 (7%)	0	100	100
All	All	938/999 (94%)	888 (95%)	50 (5%)	0	100	100

There are no Ramachandran outliers to report.

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	460/496 (93%)	442 (96%)	18 (4%)	28	54
2	B	341/399 (86%)	331 (97%)	10 (3%)	37	60
All	All	801/895 (90%)	773 (96%)	28 (4%)	32	56

5 of 28 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	447	SER
2	B	362	THR
1	A	517	LEU
2	B	261	VAL
1	A	515	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	494	ASN
1	A	512	GLN
2	B	265	ASN
2	B	137	ASN
1	A	265	ASN

5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
4	R	21/25 (84%)	5 (23%)	0

All (5) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
4	R	7	G
4	R	19	A

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
4	R	20	A
4	R	25	A
4	R	27	A

There are no RNA pucker outliers to report.

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

1 non-standard protein/DNA/RNA residue is 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	$\# Z > 2$
4	3DR	R	5	4	8,11,12	0.42	0	7,14,17	1.71	3 (42%)

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	3DR	R	5	4	-	0/3/15/16	0/1/1/1

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	R	5	3DR	O4'-C4'-C3'	2.80	107.85	103.73
4	R	5	3DR	C1'-C2'-C3'	2.56	106.00	103.26
4	R	5	3DR	C1'-O4'-C4'	2.14	111.73	108.37

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 oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 5 ligands modelled in this entry, 2 are monoatomic - leaving 3 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	$\# Z > 2$
5	EFZ	A	601	-	23,23,23	6.14	12 (52%)	36,36,36	5.37	24 (66%)
7	GOL	D	102	-	5,5,5	0.30	0	5,5,5	0.61	0
7	GOL	R	101	-	5,5,5	0.45	0	5,5,5	0.62	0

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
5	EFZ	A	601	-	-	9/10/32/32	0/3/3/3
7	GOL	D	102	-	-	2/4/4/4	-
7	GOL	R	101	-	-	2/4/4/4	-

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	601	EFZ	C1-C6	20.04	1.62	1.40
5	A	601	EFZ	C12-C11	12.08	1.92	1.48
5	A	601	EFZ	C5-C4	-7.94	1.25	1.38
5	A	601	EFZ	C2-C1	7.58	1.52	1.39

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	601	EFZ	C3-C4	-7.01	1.25	1.38

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	601	EFZ	C2-C1-C6	-16.26	103.83	119.87
5	A	601	EFZ	C12-C10-C11	12.09	90.46	59.04
5	A	601	EFZ	C2-C1-N	11.47	139.21	119.91
5	A	601	EFZ	C7-O2-C14	10.30	139.71	121.48
5	A	601	EFZ	O2-C7-C6	-9.80	104.22	111.61

There are no chirality outliers.

5 of 13 torsion outliers are listed below:

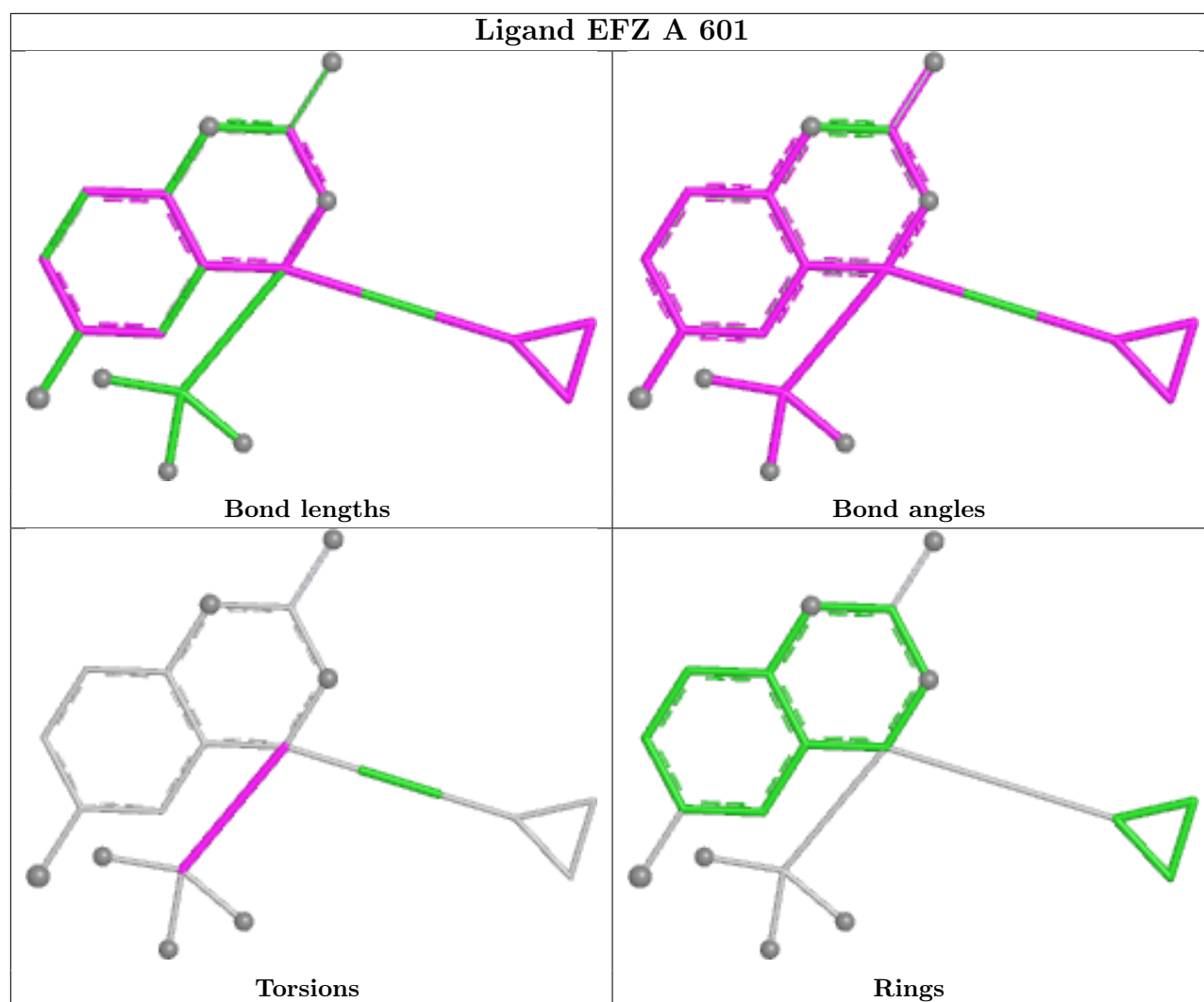
Mol	Chain	Res	Type	Atoms
5	A	601	EFZ	F1-C13-C7-O2
5	A	601	EFZ	F2-C13-C7-O2
5	A	601	EFZ	F3-C13-C7-O2
5	A	601	EFZ	F1-C13-C7-C6
5	A	601	EFZ	F2-C13-C7-C6

There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	601	EFZ	3	0
7	D	102	GOL	1	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 than 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	543/558 (97%)	-0.37	6 (1%) 78 60	50, 112, 156, 184	1 (0%)
2	B	406/441 (92%)	-0.32	5 (1%) 76 58	76, 114, 182, 206	0
3	D	23/23 (100%)	-0.50	0 100 100	86, 110, 181, 186	0
4	R	24/25 (96%)	-0.99	0 100 100	97, 117, 162, 178	0
All	All	996/1047 (95%)	-0.37	11 (1%) 78 60	50, 113, 173, 206	1 (0%)

The worst 5 of 11 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	92	LEU	5.0
1	A	500	GLN	3.0
2	B	202	ILE	2.9
2	B	152	GLY	2.6
2	B	189	VAL	2.6

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	3DR	R	5	11/12	0.91	0.10	117,153,170,173	0

6.3 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

6.4 Ligands

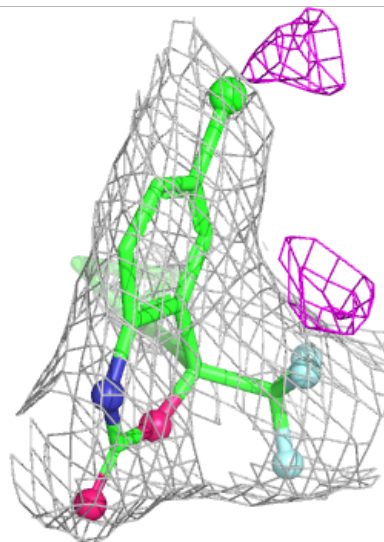
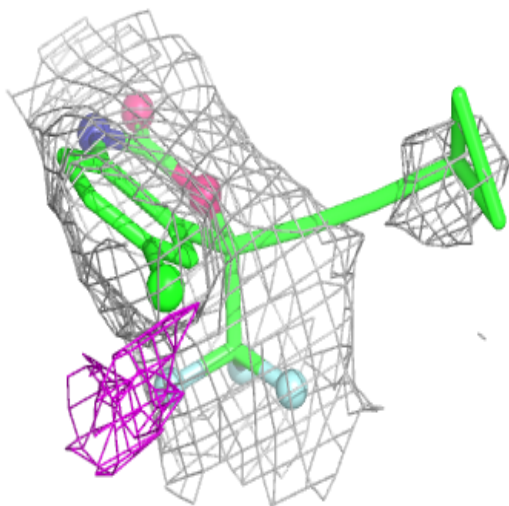
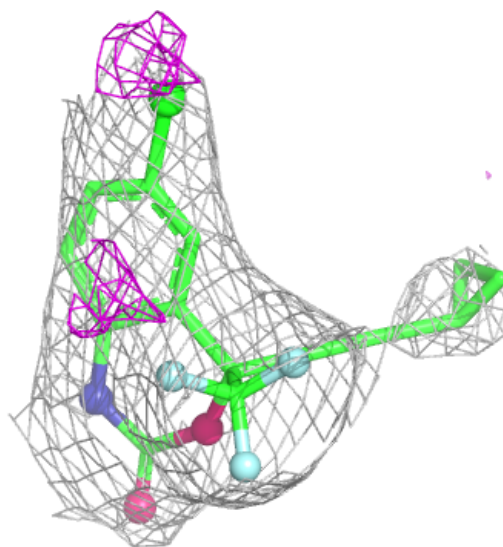
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å ²)	Q<0.9
7	GOL	R	101	6/6	0.58	0.16	119,125,126,131	0
7	GOL	D	102	6/6	0.71	0.36	106,122,127,131	0
6	CA	D	101	1/1	0.82	0.17	133,133,133,133	0
6	CA	A	602	1/1	0.93	0.11	140,140,140,140	0
5	EFZ	A	601	21/21	0.97	0.06	76,91,97,100	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 EFZ A 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.