wwPDB X-ray Structure Validation Summary Report (i)

Dec 10, 2022 – 05:14 PM EST

PDB ID	:	1JKH
Title	:	CRYSTAL STRUCTURE OF Y181C MUTANT HIV-1 REVERSE TRANSCRIPTASE IN COMPLEX WITH DMP-266(EFAVIRENZ)
Authors	:	Ren, J.; Nichols, C.; Bird, L.; Chamberlain, P.; Weaver, K.; Short, S.; Stuart, D.I.; Stammers, D.K.
Deposited on	:	2001-07-12
Resolution	:	2.50 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

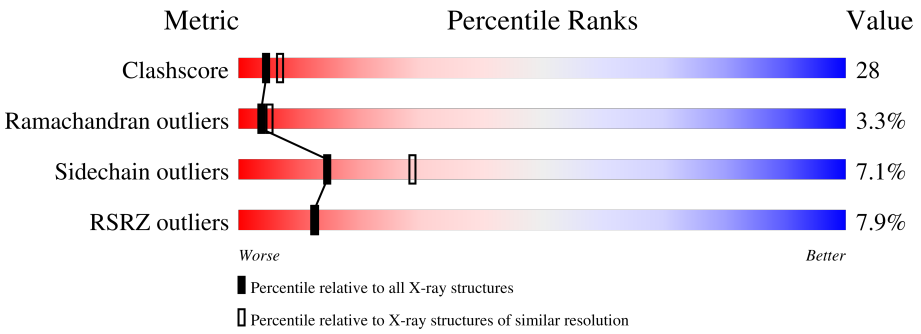
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.31.2
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.2

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 2.50 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))
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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	560	8%; 47%; 44%; 8%; •
2	B	440	7%; 51%; 34%; 5%; 10%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 7926 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 HIV-1 RT, A-CHAIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	551	4491	2899	750	833	9	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	181	CYS	TYR	modified residue	UNP P04585
A	280	CSD	CYS	modified residue	UNP P04585

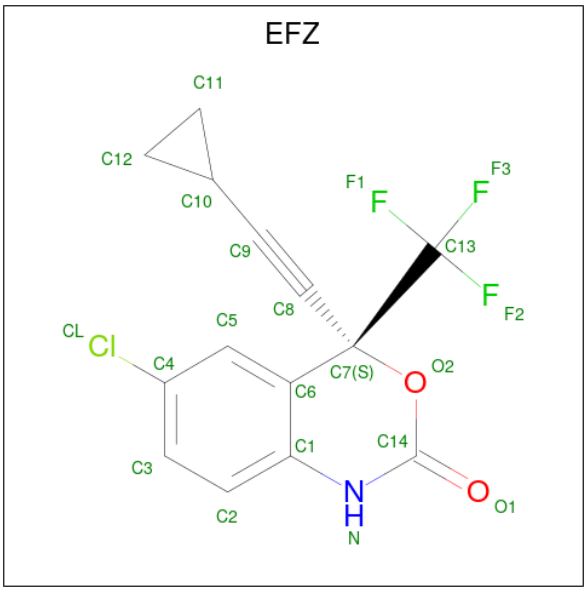
- Molecule 2 is a protein called HIV-1 RT, B-CHAIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	398	3291	2137	550	596	8	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	181	CYS	TYR	engineered mutation	UNP P04585

- Molecule 3 is (-)-6-CHLORO-4-CYCLOPROPYLETHYNYL-4-TRIFLUOROMETHYL-1,4-DIHYDRO-2H-3,1-BENZOXAZIN-2-ONE (three-letter code: EFZ) (formula: $C_{14}H_9ClF_3NO_2$).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
			Total	C	Cl	F	N	O		
3	A	1	21	14	1	3	1	2	0	0

- Molecule 4 is water.

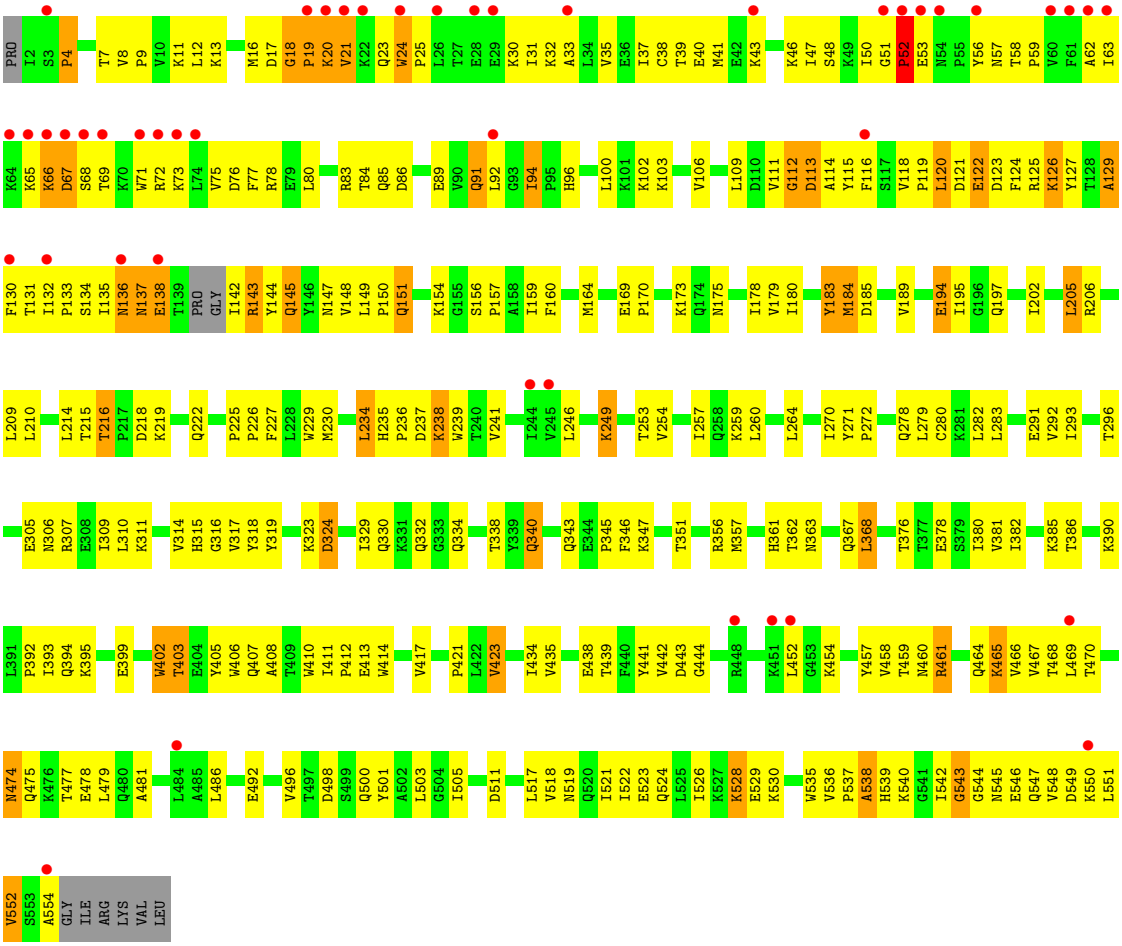
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	71	71	71	0	0
			Total	O		
4	B	52	52	52	0	0

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.

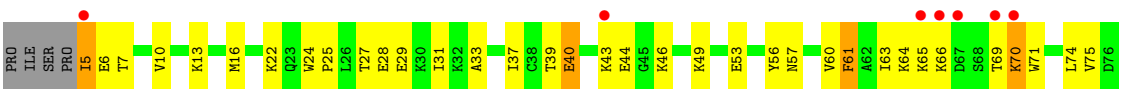
• Molecule 1: HIV-1 RT, A-CHAIN

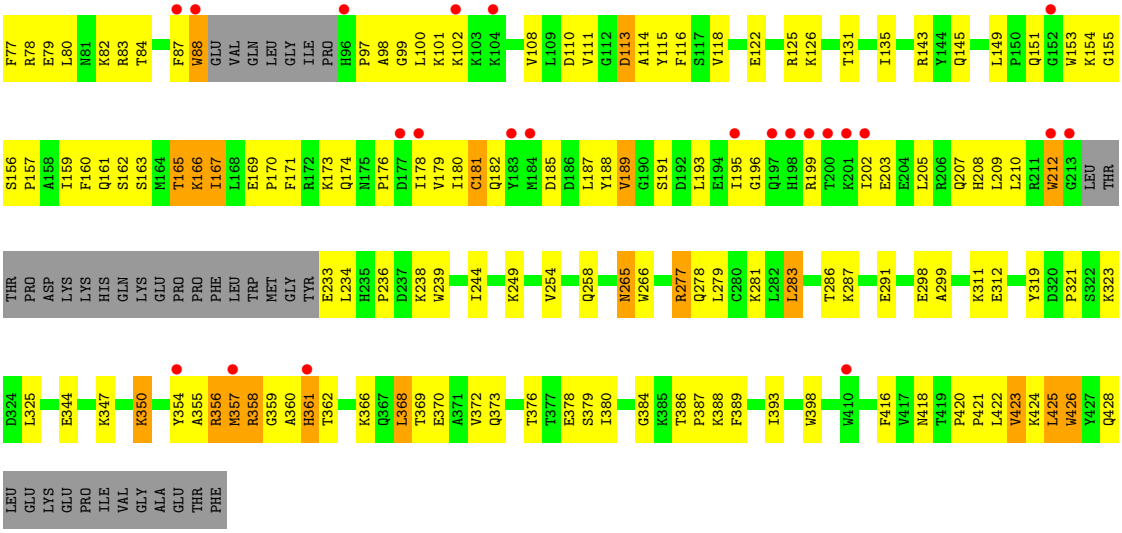
Chain A: 8% 47% 44% 8%



• Molecule 2: HIV-1 RT, B-CHAIN

Chain B: 7% 51% 34% 5% 10%




F77 R78 E79 L80 N81 K82 R83 T84 F87 W88 GLU VAL GLN LEU GLY ILE PRO H96 P97 A98 G99 L100 K101 K102 K103 K104 V108 L109 D110 V111 G112 D113 A114 Y115 F116 S117 V118 E122 R125 K126 T131 I135 R143 Y144 Q145 L149 P150 Q151 G152 W153 K154 G155
S156 P157 A158 I159 F160 Q161 S162 S163 M164 T165 K166 I167 L168 E169 P170 F171 R172 K173 Q174 N175 P176 D177 I178 V179 I180 C181 Q182 Y183 M184 D185 D186 L187 Y188 V189 G190 S191 D192 L193 E194 I195 G196 Q197 H198 R199 T200 K201 I202 E203 E204 L205 R206 Q207 H208 L209 L210 R211 W212 G213 LEU THR
THR PRO ASP LYS LYS HIS GLN LYS GLU PRO PRO PHE LEU TRP MET GLY TYR E233 L234 H235 P236 D237 K238 T369 W239 I244 K249 V254 Q258 N265 W266 R277 Q278 L279 C280 K281 L282 L283 T286 K287 E291 E298 A299 K311 E312 Y319 D320 P321 S322 K323
D324 L325 E344 K347 K350 Y354 A355 R356 M357 R358 G359 A360 H361 T362 K366 Q367 L368 T369 E370 A371 V372 Q373 T376 T377 E378 S379 I380 G384 K385 T386 P387 K388 F389 I393 W398 W410 F416 V417 N418 T419 P420 P421 L422 V423 K424 L425 W426 Y427 Q428
LEU GLU LYS GLU PRO ILE VAL GLY ALA GLU THR PHE

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	138.50Å 109.00Å 73.20Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.68 – 2.50 29.68 – 2.50	Depositor EDS
% Data completeness (in resolution range)	98.7 (29.68-2.50) 98.7 (29.68-2.50)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.63 (at 2.51Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.236 , 0.303 0.224 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	62.4	Xtriage
Anisotropy	0.321	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 65.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	7926	wwPDB-VP
Average B, all atoms (Å^2)	72.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.61% 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: EFZ, CSD

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	0.46	0/4596	0.70	1/6242 (0.0%)
2	B	0.42	0/3382	0.67	0/4591
All	All	0.45	0/7978	0.68	1/10833 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	18	GLY	N-CA-C	6.01	128.12	113.10

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	4491	0	4537	289	0
2	B	3291	0	3321	173	0
3	A	21	0	9	1	0
4	A	71	0	0	8	0
4	B	52	0	0	2	0
All	All	7926	0	7867	446	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 28.

The worst 5 of 446 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:50:ILE:HG13	1:A:51:GLY:H	1.12	1.11
1:A:57:ASN:HB2	1:A:143:ARG:HH22	1.23	1.02
2:B:180:ILE:HG12	2:B:189:VAL:HG13	1.44	0.99
1:A:175:ASN:HB3	1:A:178:ILE:HD13	1.48	0.95
2:B:350:LYS:HE2	2:B:378:GLU:OE2	1.67	0.94

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	546/560 (98%)	475 (87%)	47 (9%)	24 (4%)	2 3
2	B	392/440 (89%)	348 (89%)	37 (9%)	7 (2%)	8 14
All	All	938/1000 (94%)	823 (88%)	84 (9%)	31 (3%)	4 5

5 of 31 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	20	LYS
1	A	21	VAL
1	A	113	ASP
1	A	461	ARG
1	A	538	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	492/499 (99%)	455 (92%)	37 (8%)	13	26
2	B	362/400 (90%)	338 (93%)	24 (7%)	16	32
All	All	854/899 (95%)	793 (93%)	61 (7%)	14	28

5 of 61 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	368	LEU
2	B	356	ARG
1	A	517	LEU
2	B	350	LYS
2	B	425	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 25 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	474	ASN
2	B	175	ASN
2	B	428	GLN
2	B	57	ASN
2	B	197	GLN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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	$\#\|Z\| > 2$
1	CSD	A	280	1	3,7,8	0.85	0	1,8,10	6.14	1 (100%)

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
1	CSD	A	280	1	-	2/2/6/8	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	280	CSD	OD1-SG-CB	6.14	117.23	105.54

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	280	CSD	N-CA-CB-SG
1	A	280	CSD	CA-CB-SG-OD1

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.

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	$\#\|Z\| > 2$
3	EFZ	A	999	-	23,23,23	2.62	7 (30%)	36,36,36	1.38	2 (5%)

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
3	EFZ	A	999	-	-	0/10/32/32	0/3/3/3

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	999	EFZ	C10-C9	7.24	1.70	1.46
3	A	999	EFZ	C12-C11	-5.35	1.28	1.48
3	A	999	EFZ	C7-C6	4.32	1.57	1.51
3	A	999	EFZ	C11-C10	-3.69	1.24	1.48
3	A	999	EFZ	C12-C10	-3.69	1.24	1.48

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	999	EFZ	C12-C10-C9	-4.96	105.59	119.06
3	A	999	EFZ	C11-C10-C9	-4.94	105.65	119.06

There are no chirality outliers.

There are no torsion outliers.

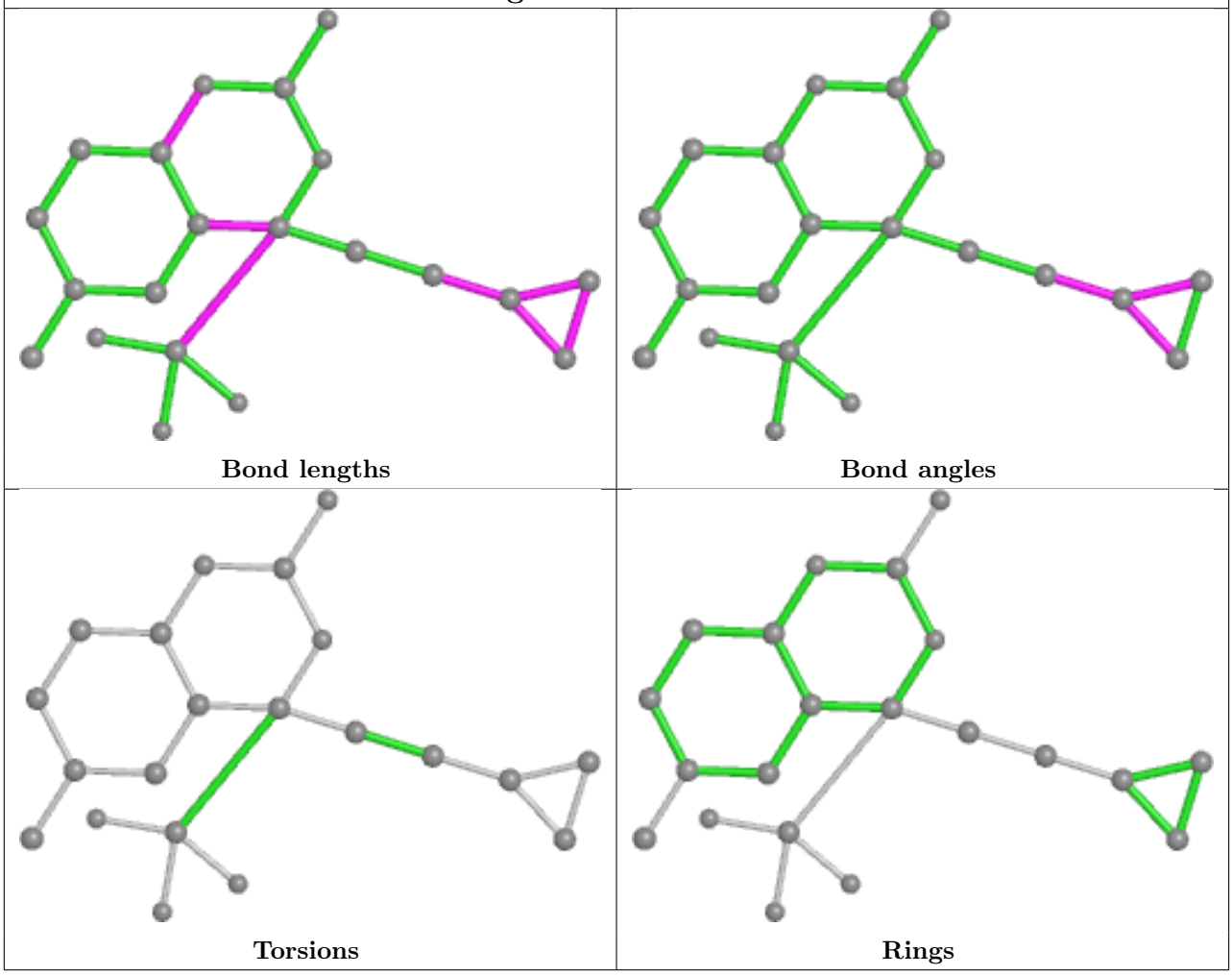
There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	999	EFZ	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 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 EFZ A 999



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	550/560 (98%)	0.29	45 (8%)	11	11	32, 68, 127, 150	0
2	B	398/440 (90%)	0.19	30 (7%)	14	14	34, 64, 119, 143	0
All	All	948/1000 (94%)	0.25	75 (7%)	12	12	32, 66, 124, 150	0

The worst 5 of 75 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	19	PRO	7.4
1	A	20	LYS	7.0
2	B	5	ILE	6.2
1	A	67	ASP	6.2
1	A	71	TRP	5.2

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
1	CSD	A	280	8/9	0.95	0.14	59,61,78,82	0

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

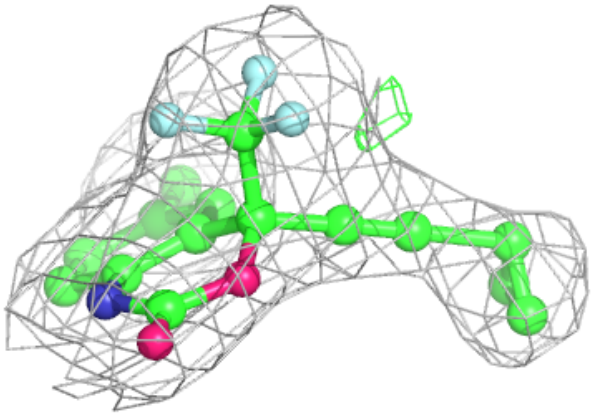
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists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

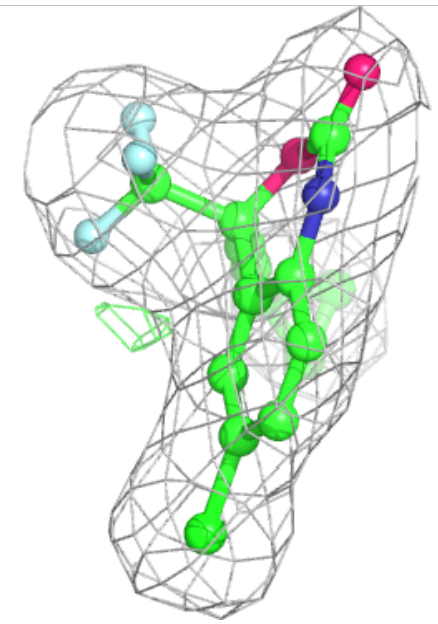
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	EFZ	A	999	21/21	0.96	0.14	31,55,65,70	0

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 999:

$2mF_o$-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.