wwPDB X-ray Structure Validation Summary Report (i)

Oct 17, 2021 – 08:39 AM EDT

PDB ID	:	1IKV
Title	:	K103N Mutant HIV-1 Reverse Transcriptase in Complex with Efivarenz
Authors	:	Lindberg, J.; Unge, T.
Deposited on	:	2001-05-07
Resolution	:	3.00 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:

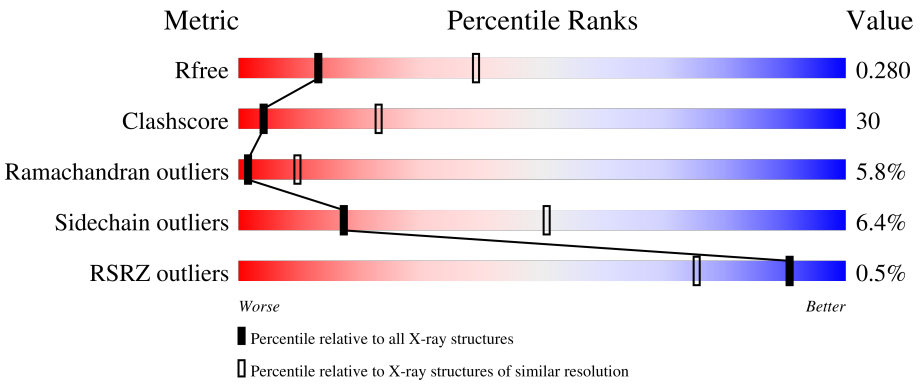
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 3.00 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	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)

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	46% 47% 6% ••
2	B	427	% 44% 44% 6% 5%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 7872 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 POL POLYPROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	556	4514	2921	752	833	8	2	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	103	ASN	LYS	engineered mutation	UNP P03366
A	478	GLN	GLU	engineered mutation	UNP P03366

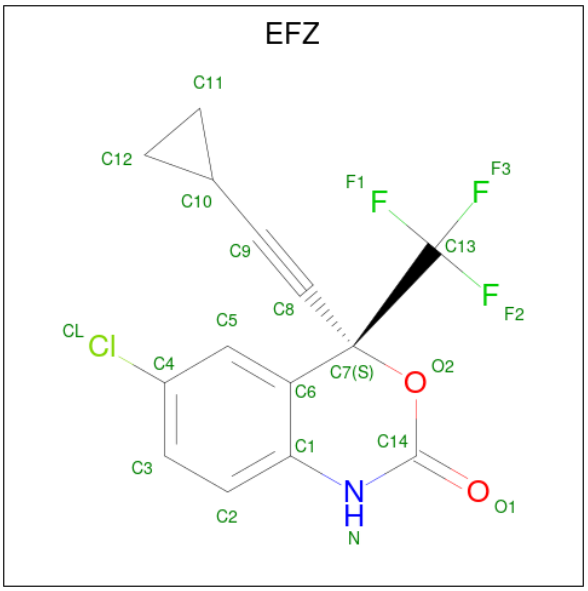
- Molecule 2 is a protein called POL POLYPROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	404	3337	2174	548	609	6	30	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1103	ASN	LYS	engineered mutation	UNP P03366

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



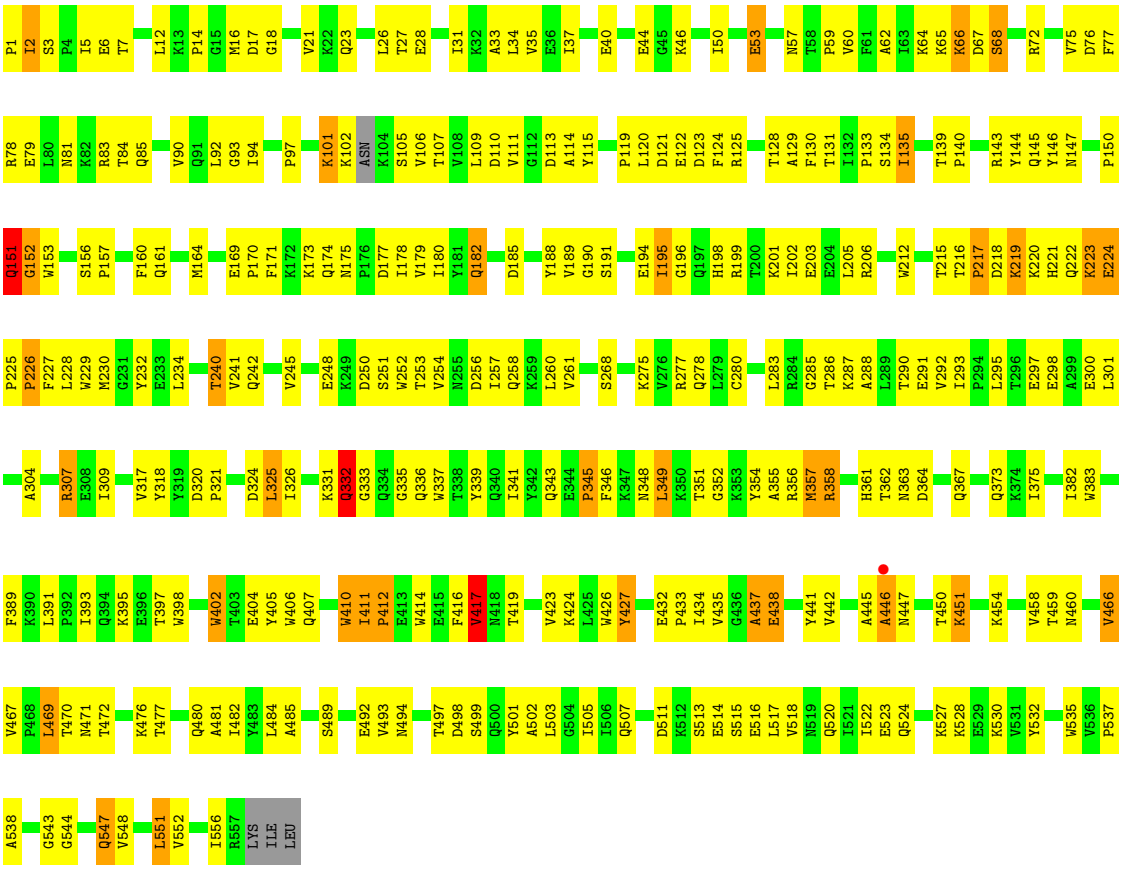
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
			Total	C	Cl	F	N	O		
3	A	1	21	14	1	3	1	2	0	0

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.

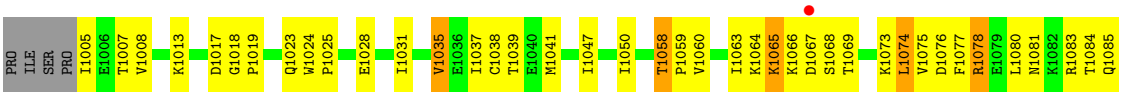
• Molecule 1: POL POLYPROTEIN

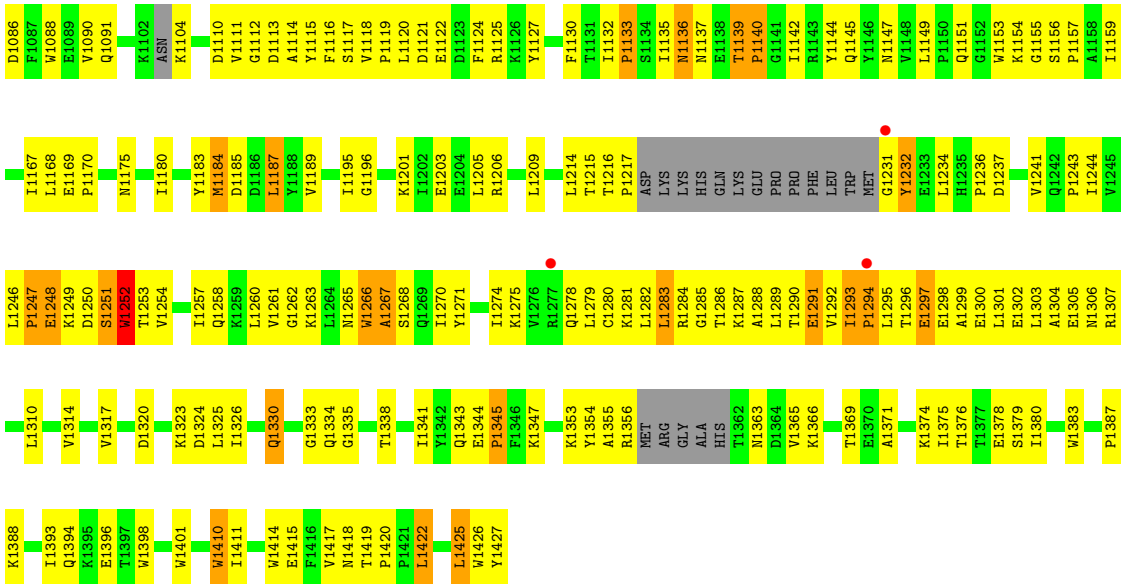
Chain A: 46% 47% 6% ••



• Molecule 2: POL POLYPROTEIN

Chain B: % 44% 44% 6% 5%




D1086 F1087 W1088 E1089 V1090 Q1091 K1102 ASN K1104 D1110 V1111 G1112 D1113 A1114 Y1115 F1116 S1117 V1118 P1119 L1120 D1121 E1122 D1123 F1124 R1125 K1126 Y1127 F1130 T1131 I1132 P1133 S1134 I1135 N1136 N1137 E1138 T1139 P1140 G1141 I1142 R1143 Y1144 Q1145 Y1146 N1147 V1148 L1149 P1150 Q1151 G1152 W1153 K1154 G1155 S1156 P1157 A1158 I1159
I1167 L1168 E1169 P1170 N1175 I1180 Y1183 M1184 D1185 D1186 L1187 Y1188 V1189 I1195 G1196 K1201 I1202 E1203 E1204 L1205 R1206 L1209 L1214 T1215 T1216 P1217 ASP LYS LYS HIS GLN LYS GLU PRO PRO PHE LEU TRP MET G1231 Y1232 E1233 L1234 H1235 P1236 D1237 V1241 Q1242 P1243 I1244 V1245
L1246 P1247 E1248 K1249 D1250 S1251 W1252 T1253 V1254 I1257 Q1258 K1259 L1260 V1261 G1262 K1263 L1264 N1265 W1266 A1267 S1268 Q1269 I1270 Y1271 I1274 K1275 V1276 R1277 Q1278 L1279 C1280 K1281 L1282 L1283 R1284 G1285 T1286 K1287 A1288 L1289 T1290 E1291 V1292 I1293 P1294 L1295 T1296 E1297 E1298 A1299 E1300 L1301 E1302 L1303 A1304 E1305 N1306 R1307
L1310 V1314 V1317 D1320 K1323 D1324 L1325 I1326 Q1330 G1333 Q1334 G1335 T1338 I1341 Y1342 Q1343 E1344 P1345 F1346 K1347 K1353 Y1354 A1355 R1356 MET ARG GLY ALA HIS T1362 N1363 D1364 V1365 K1366 T1369 E1370 A1371 K1374 I1375 T1376 T1377 E1378 S1379 I1380 W1383 P1387
K1388 I1393 Q1394 K1395 E1396 T1397 W1398 W1401 W1410 I1411 W1414 E1415 F1416 V1417 N1418 T1419 P1420 P1421 L1422 L1425 W1426 Y1427

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	119.63Å 157.17Å 156.15Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	24.84 – 3.00 24.83 – 2.80	Depositor EDS
% Data completeness (in resolution range)	87.3 (24.84-3.00) 83.9 (24.83-2.80)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.17 (at 2.80Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.228 , 0.294 0.219 , 0.280	Depositor DCC
R_{free} test set	1524 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å^2)	61.5	Xtriage
Anisotropy	0.161	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 43.8	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	0.89	EDS
Total number of atoms	7872	wwPDB-VP
Average B, all atoms (Å^2)	52.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.60% 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

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	0.41	0/4631	0.65	0/6291
2	B	0.42	0/3430	0.65	0/4660
All	All	0.42	0/8061	0.65	0/10951

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	4514	0	4565	273	0
2	B	3337	0	3365	212	0
3	A	21	0	8	0	0
All	All	7872	0	7938	468	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 30.

The worst 5 of 468 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:223:LYS:HG3	1:A:224:GLU:H	1.05	1.09
2:B:1283:LEU:HD13	2:B:1293:ILE:HB	1.37	1.05
1:A:278:GLN:HE21	1:A:298:GLU:HB2	1.20	1.03
1:A:139:THR:HG22	1:A:140:PRO:HD2	1.48	0.94
2:B:1246:LEU:HD12	2:B:1307:ARG:HB3	1.47	0.93

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	552/560 (99%)	438 (79%)	78 (14%)	36 (6%)	1 7
2	B	396/427 (93%)	333 (84%)	44 (11%)	19 (5%)	2 13
All	All	948/987 (96%)	771 (81%)	122 (13%)	55 (6%)	1 10

5 of 55 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	135	ILE
2	B	1065	LYS
2	B	1252	TRP
2	B	1267	ALA
2	B	1294	PRO

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	494/500 (99%)	457 (92%)	37 (8%)	13	43
2	B	368/389 (95%)	350 (95%)	18 (5%)	25	61
All	All	862/889 (97%)	807 (94%)	55 (6%)	17	51

5 of 55 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	402	TRP
1	A	547	GLN
2	B	1426	TRP
2	B	1345	PRO
1	A	404	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	1161	GLN
2	B	1306	ASN
2	B	1418	ASN
2	B	1258	GLN
1	A	475	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)

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

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	2000	-	23,23,23	4.53	12 (52%)	36,36,36	1.77	10 (27%)

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	2000	-	-	4/10/32/32	0/3/3/3

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	2000	EFZ	C4-CL	-17.96	1.35	1.74
3	A	2000	EFZ	C9-C8	5.29	1.35	1.19
3	A	2000	EFZ	C7-C6	5.03	1.58	1.51
3	A	2000	EFZ	C5-C6	3.79	1.45	1.39
3	A	2000	EFZ	C1-C6	3.43	1.44	1.40

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	2000	EFZ	C12-C10-C9	-3.55	109.42	119.06
3	A	2000	EFZ	O2-C7-C13	3.54	111.29	104.76
3	A	2000	EFZ	C7-O2-C14	-3.42	115.21	121.44
3	A	2000	EFZ	O2-C14-N	3.12	122.41	117.07
3	A	2000	EFZ	O2-C7-C6	3.11	113.84	111.64

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	2000	EFZ	F1-C13-C7-O2

Continued on next page...

Continued from previous page...

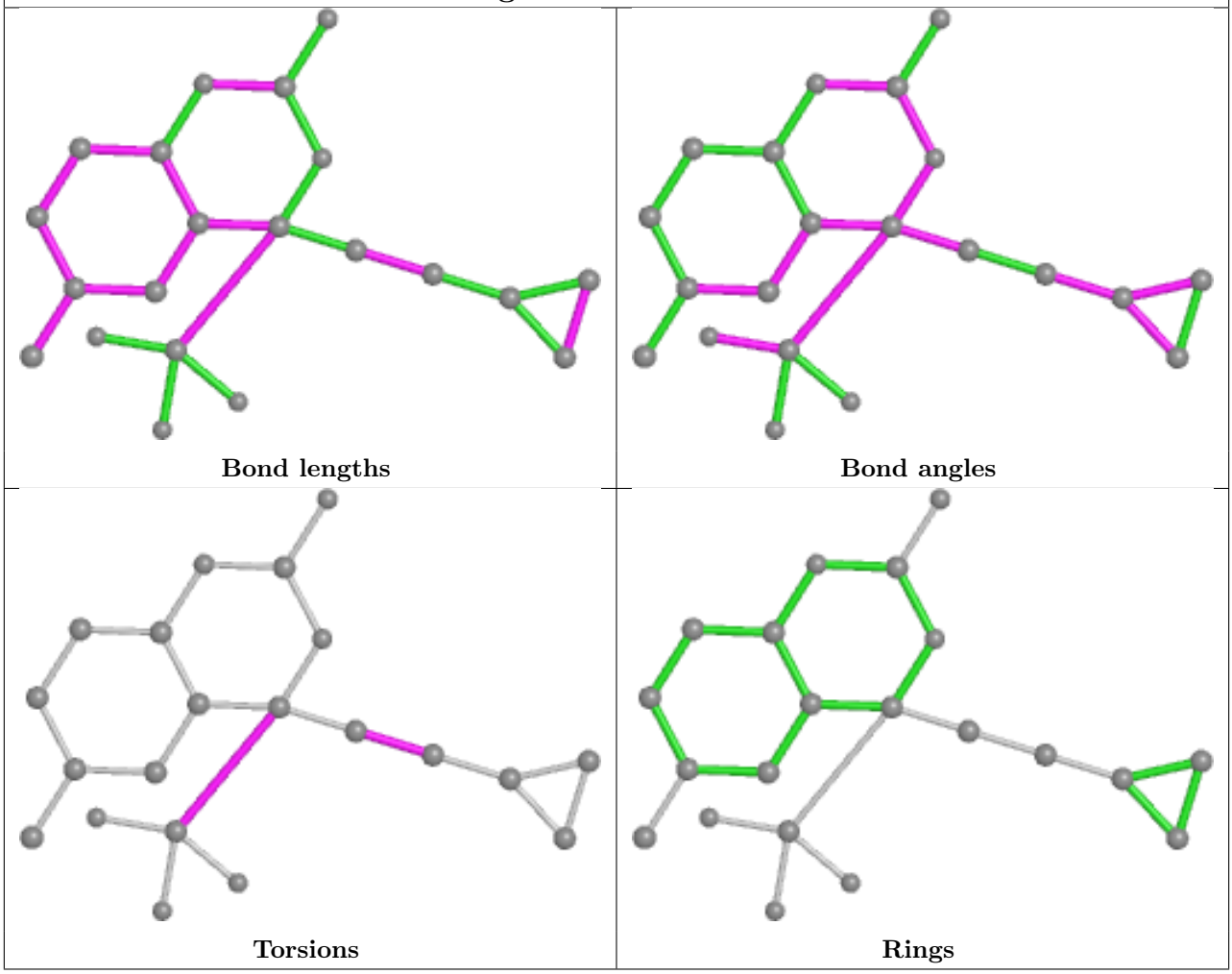
Mol	Chain	Res	Type	Atoms
3	A	2000	EFZ	F3-C13-C7-O2
3	A	2000	EFZ	C7-C8-C9-C10
3	A	2000	EFZ	F2-C13-C7-O2

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 EFZ A 2000



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	556/560 (99%)	-0.23	1 (0%)	95	87	13, 52, 89, 125	2 (0%)
2	B	404/427 (94%)	-0.25	4 (0%)	82	59	14, 43, 107, 122	9 (2%)
All	All	960/987 (97%)	-0.24	5 (0%)	91	75	13, 48, 100, 125	11 (1%)

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	1067	ASP	3.8
2	B	1277	ARG	2.2
2	B	1231	GLY	2.1
2	B	1294	PRO	2.1
1	A	446	ALA	2.0

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 monosaccharides in this entry.

6.4 Ligands (i)

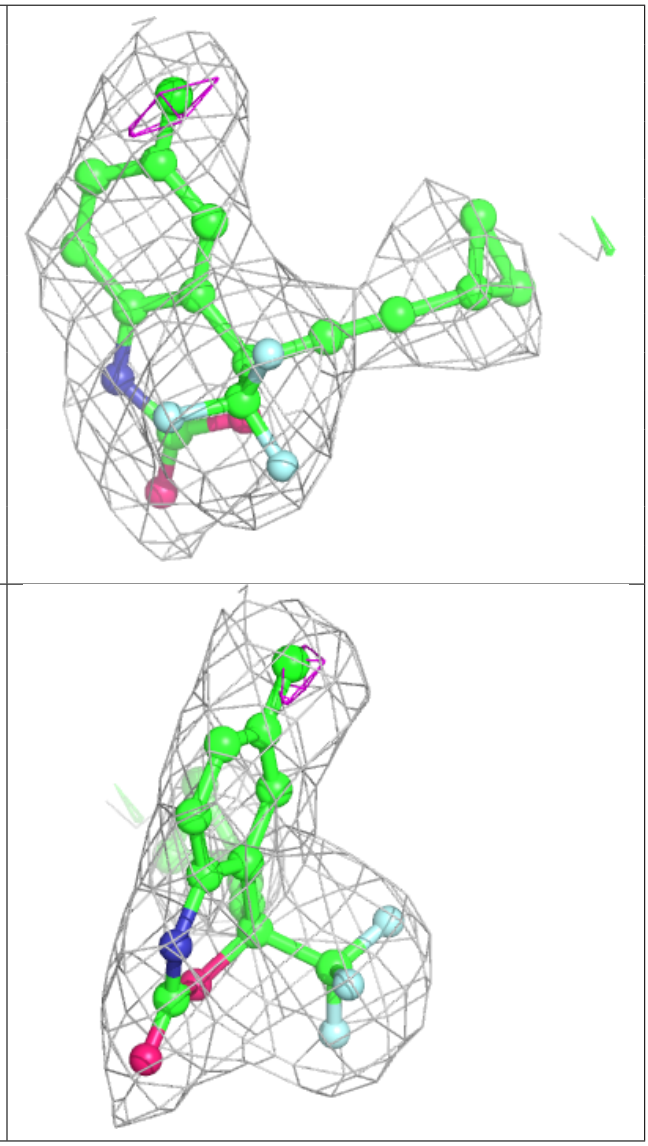
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists 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	2000	21/21	0.92	0.19	31,34,36,36	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 2000:

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.