wwPDB X-ray Structure Validation Summary Report (i)

Jun 12, 2024 – 08:44 PM EDT

PDB ID	:	1C0U
Title	:	CRYSTAL STRUCTURE OF HIV-1 REVERSE TRANSCRIPTASE IN COMPLEX WITH BM+50.0934
Authors	:	Ren, J.; Esnouf, R.M.; Hopkins, A.L.; Stuart, D.I.; Stammers, D.K.
Deposited on	:	1999-07-19
Resolution	:	2.52 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:

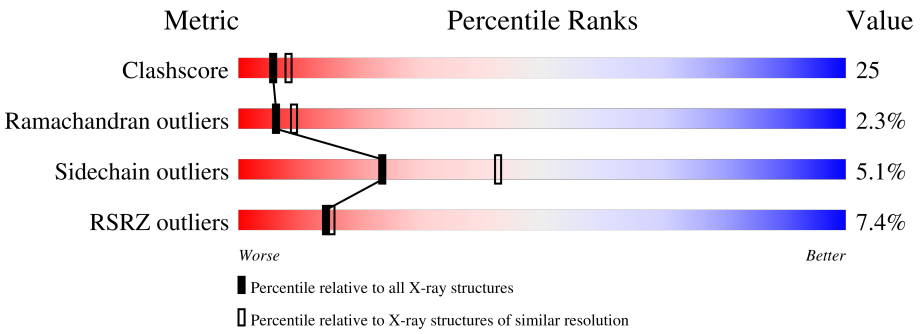
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36.2
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 2.52 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	6463 (2.54-2.50)
Ramachandran outliers	138981	6335 (2.54-2.50)
Sidechain outliers	138945	6337 (2.54-2.50)
RSRZ outliers	127900	5630 (2.54-2.50)

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% 49% 42% 5% •
2	B	440	6% 50% 40% • 6%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 7938 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.

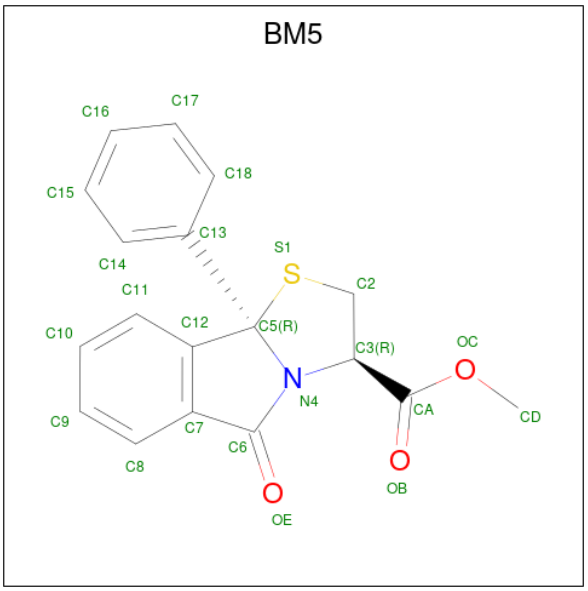
- Molecule 1 is a protein called HIV-1 REVERSE TRANSCRIPTASE (A-CHAIN).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	539	4410	2853	734	815	8	0	0	0

- Molecule 2 is a protein called HIV-1 REVERSE TRANSCRIPTASE (B-CHAIN).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	414	3415	2219	566	623	7	0	0	0

- Molecule 3 is (R)-(+) 5(9BH)-OXO-9B-PHENYL-2,3-DIHYDROTHIAZOLO[2,3-A]ISOINDOL-3-CARBOXYLIC ACID METHYL ESTER (three-letter code: BM5) (formula: $C_{18}H_{15}NO_3S$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
3	A	1	23	18	1	3	1	0	0

- Molecule 4 is water.

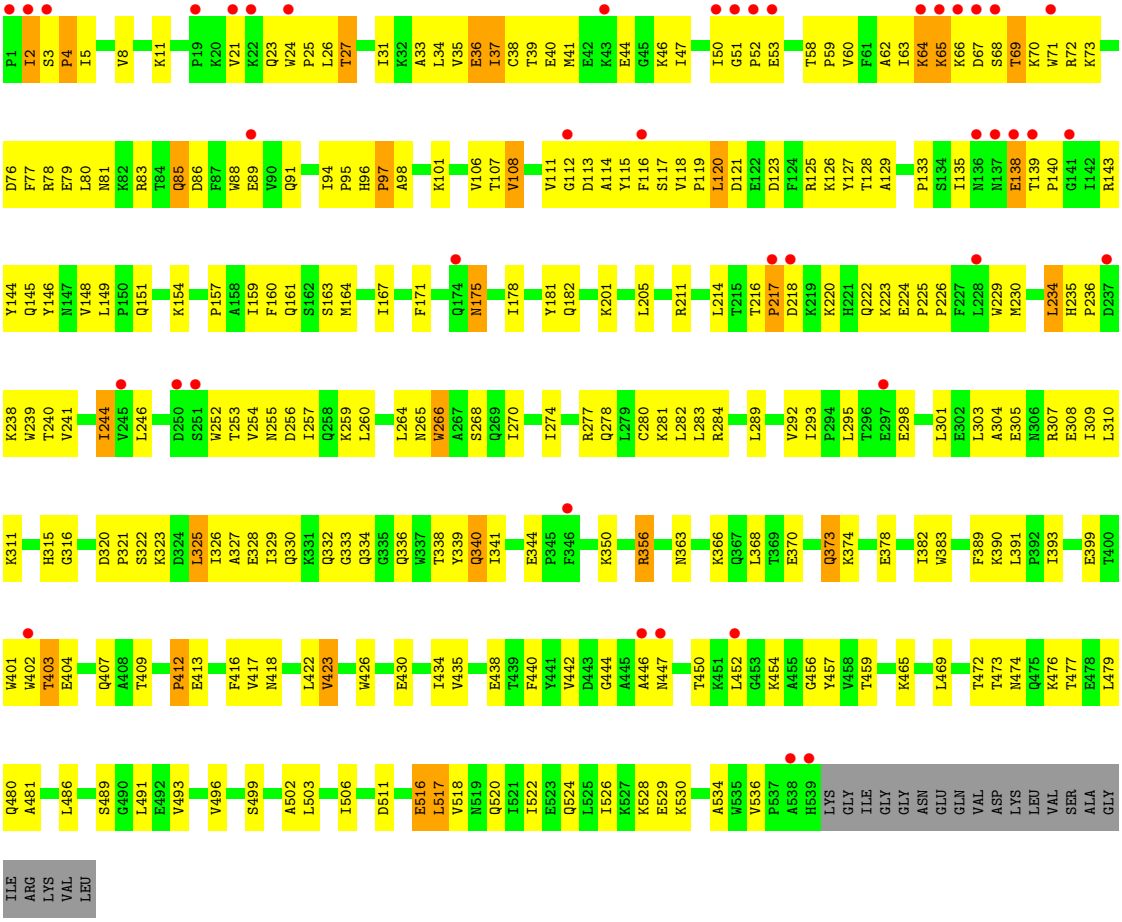
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	57	57	57	0	0
			Total	O		
4	B	33	33	33	0	0

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.

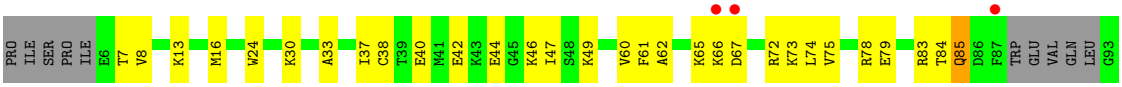
• Molecule 1: HIV-1 REVERSE TRANSCRIPTASE (A-CHAIN)

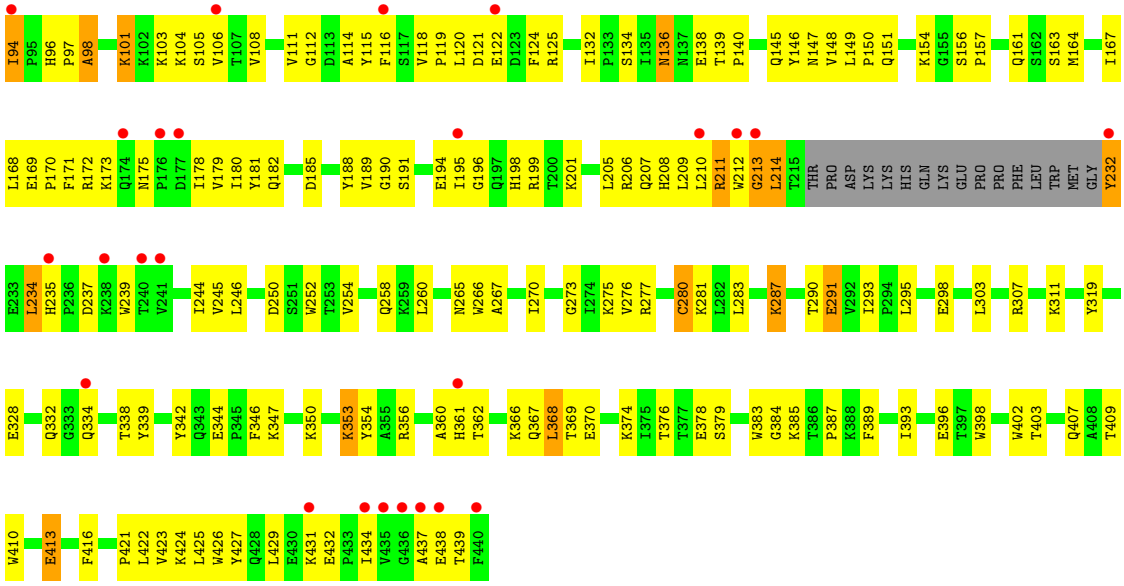
Chain A:



• Molecule 2: HIV-1 REVERSE TRANSCRIPTASE (B-CHAIN)

Chain B:




I94 P95 H96 P97 A98 K101 K102 K103 K104 S105 V106 T107 V108 V111 G112 D113 A114 Y115 F116 S117 V118 P119 L120 D121 E122 D123 F124 R125 I132 P133 S134 I135 N136 N137 E138 T139 P140 Q145 Y146 N147 V148 L149 P150 Q151 K154 G155 S156 P157 Q161 S162 S163 M164 I167
L168 E169 P170 F171 R172 K173 Q174 N175 P176 D177 I178 V179 I180 Y181 Q182 D185 Y188 V189 G190 S191 E194 I195 G196 Q197 H198 R199 T200 K201 L205 R206 Q207 H208 L209 L210 R211 W212 G213 L214 T215 THR PRO ASP LYS LYS HIS GLN LYS GLU PRO PRO PHE LEU TRP MET GLY Y232
E233 L234 H235 P236 D237 K238 W239 T240 V241 I244 V245 L246 D250 S251 W252 T253 V254 Q258 K259 L260 N265 W266 A267 I270 G273 I274 K275 V276 R277 C280 K281 L282 L283 K287 T290 E291 V292 I293 P294 L295 E298 L303 R307 K311 Y319
E328 Q332 G333 Q334 T338 Y339 Y342 Q343 E344 P345 F346 K347 K350 K353 Y354 A355 R356 A360 H361 T362 K366 Q367 L368 T369 E370 K374 I375 T376 T377 E378 S379 W383 G384 K385 T386 P387 K388 F389 I393 E396 T397 W398 W402 T403 Q407 A408 T409
W410 E413 F416 P421 L422 V423 K424 L425 W426 Y427 Q428 L429 E430 K431 E432 P433 I434 V435 G436 A437 E438 T439 F440

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	137.30Å 108.50Å 72.00Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	15.00 – 2.52 14.91 – 2.52	Depositor EDS
% Data completeness (in resolution range)	88.8 (15.00-2.52) 88.7 (14.91-2.52)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.78 (at 2.51Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.232 , 0.298 0.223 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	46.8	Xtriage
Anisotropy	0.387	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 58.8	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	0.92	EDS
Total number of atoms	7938	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 3.16% 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: CSD, BM5

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	0.38	0/4518	0.65	0/6140
2	B	0.40	0/3509	0.63	0/4763
All	All	0.39	0/8027	0.64	0/10903

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	4410	0	4452	239	0
2	B	3415	0	3445	171	0
3	A	23	0	14	1	0
4	A	57	0	0	4	0
4	B	33	0	0	4	0
All	All	7938	0	7911	401	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 25.

The worst 5 of 401 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:503:LEU:HA	1:A:506:ILE:HD12	1.43	0.99
2:B:125:ARG:HE	2:B:147:ASN:HA	1.33	0.94
1:A:244:ILE:HD13	1:A:244:ILE:H	1.32	0.93
1:A:244:ILE:HD11	1:A:310:LEU:HD22	1.50	0.93
2:B:111:VAL:HB	2:B:214:LEU:HD23	1.50	0.91

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	535/560 (96%)	474 (89%)	48 (9%)	13 (2%)	6 8
2	B	408/440 (93%)	360 (88%)	39 (10%)	9 (2%)	6 10
All	All	943/1000 (94%)	834 (88%)	87 (9%)	22 (2%)	6 9

5 of 22 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	2	ILE
1	A	138	GLU
2	B	360	ALA
1	A	36	GLU
1	A	37	ILE

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	483/499 (97%)	457 (95%)	26 (5%)	22	40
2	B	375/400 (94%)	357 (95%)	18 (5%)	25	45
All	All	858/899 (95%)	814 (95%)	44 (5%)	24	43

5 of 44 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	85	GLN
2	B	280	CYS
2	B	94	ILE
2	B	234	LEU
2	B	287	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 26 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	500	GLN
2	B	85	GLN
2	B	332	GLN
1	A	520	GLN
2	B	147	ASN

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	0.72	0	1,8,10	1.77	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
1	CSD	A	280	1	-	2/2/6/8	-

There are no bond length outliers.

There are no bond angle outliers.

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.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	A	280	CSD	2	0

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	BM5	A	999	-	24,26,26	2.46	5 (20%)	30,39,39	2.11	3 (10%)

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	BM5	A	999	-	-	2/12/42/42	0/4/4/4

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	999	BM5	C2-C3	-8.71	1.38	1.53
3	A	999	BM5	C3-N4	4.46	1.52	1.46
3	A	999	BM5	C5-C13	3.18	1.58	1.53
3	A	999	BM5	C2-S1	-3.02	1.76	1.80
3	A	999	BM5	C14-C13	2.08	1.42	1.39

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	999	BM5	C2-C3-CA	10.34	128.54	111.42
3	A	999	BM5	CA-C3-N4	2.56	117.39	111.61
3	A	999	BM5	C2-C3-N4	-2.55	103.02	106.19

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	999	BM5	OB-CA-OC-CD
3	A	999	BM5	C3-CA-OC-CD

There are no ring outliers.

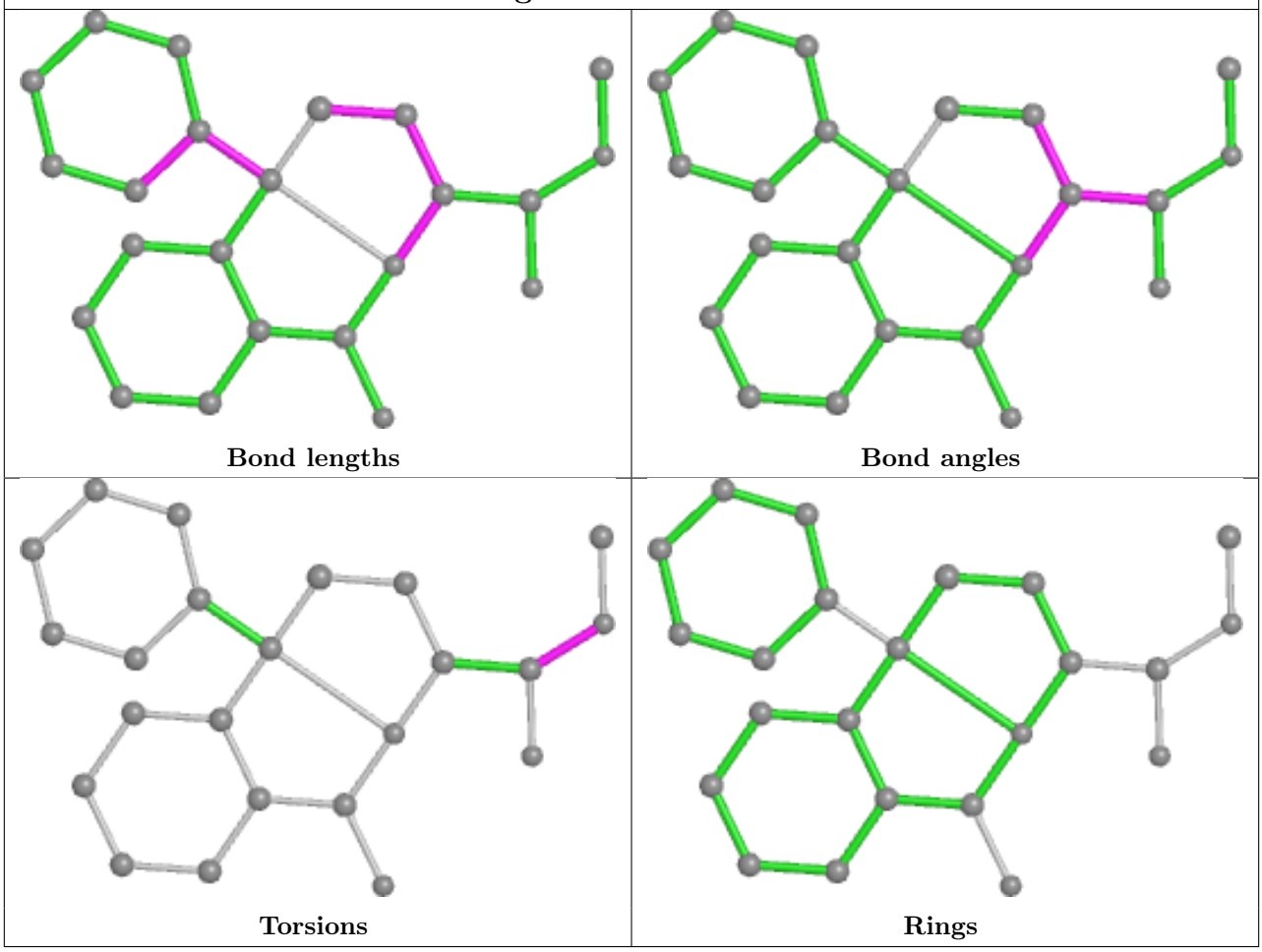
1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	999	BM5	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 BM5 A 999



5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	A	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	280:CSD	C	281:LYS	N	2.65

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	538/560 (96%)	0.23	42 (7%)	13	13	19, 51, 97, 150	0
2	B	414/440 (94%)	0.20	28 (6%)	17	18	20, 47, 102, 148	0
All	All	952/1000 (95%)	0.22	70 (7%)	14	15	19, 49, 101, 150	0

The worst 5 of 70 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	67	ASP	8.7
2	B	440	PHE	7.6
1	A	68	SER	7.4
2	B	436	GLY	7.3
2	B	437	ALA	6.6

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
1	CSD	A	280	8/9	0.94	0.15	40,46,64,64	0

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

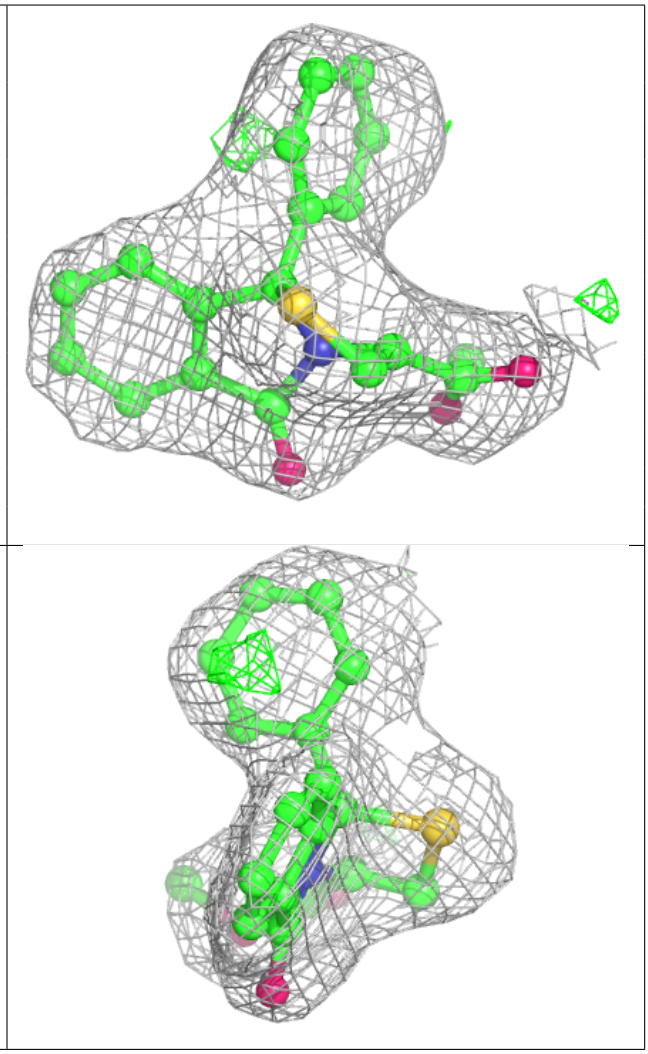
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists 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	BM5	A	999	23/23	0.88	0.25	25,47,57,61	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 BM5 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.