wwPDB X-ray Structure Validation Summary Report (i)

Aug 22, 2023 – 10:30 PM EDT

PDB ID	:	3DOK
Title	:	Crystal structure of K103N mutant HIV-1 reverse transcriptase in complex with GW678248.
Authors	:	Chamberlain, P.P.; Ren, J.; Stammers, D.K.
Deposited on	:	2008-07-04
Resolution	:	2.90 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:

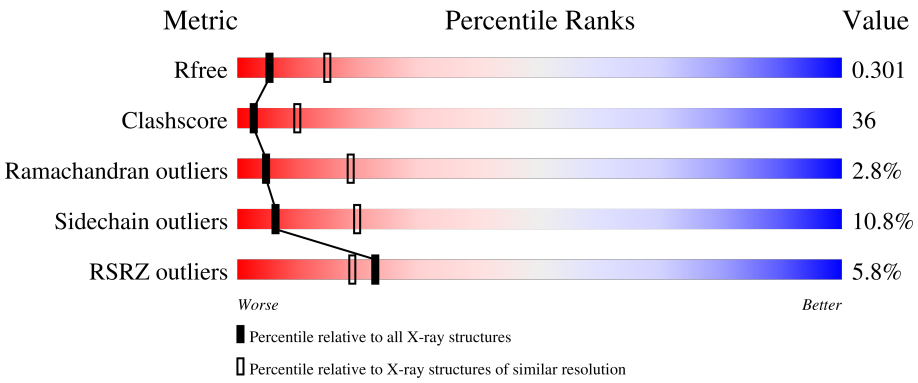
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 2.90 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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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	6% 36% 49% 9% · ·
2	B	440	5% 46% 40% 5% 9%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 7750 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/ribonuclease H.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	536	4383	2835	729	811	8	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	103	ASN	LYS	engineered mutation	UNP P04585

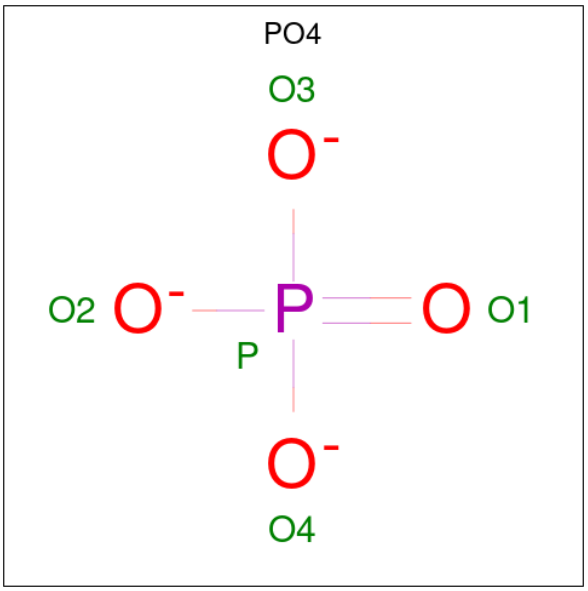
- Molecule 2 is a protein called p51 RT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	402	3320	2158	553	602	7	0	0	0

There is a discrepancy between the modelled and reference sequences:

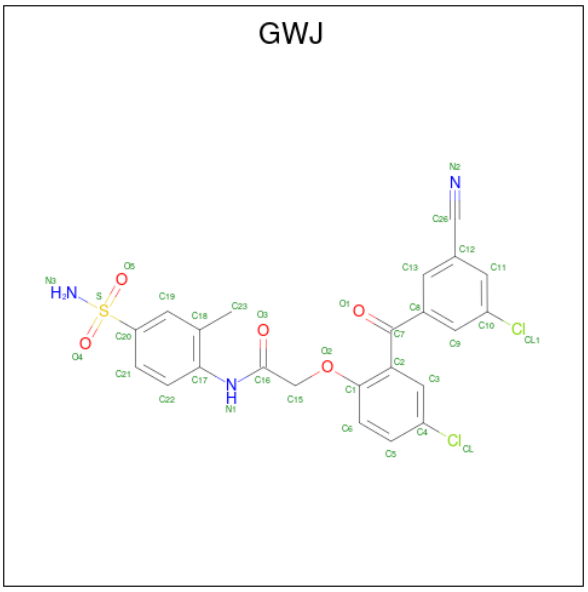
Chain	Residue	Modelled	Actual	Comment	Reference
B	103	ASN	LYS	engineered mutation	UNP P04585

- Molecule 3 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	P		
3	A	1	5	4	1	0	0

- Molecule 4 is 2-{4-chloro-2-[(3-chloro-5-cyanophenyl)carbonyl]phenoxy}-N-(2-methyl-4-sulfamoylphenyl)acetamide (three-letter code: GWJ) (formula: $C_{23}H_{17}Cl_2N_3O_5S$).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
			Total	C	Cl	N	O	S		
4	A	1	34	23	2	3	5	1	0	0

- Molecule 5 is water.

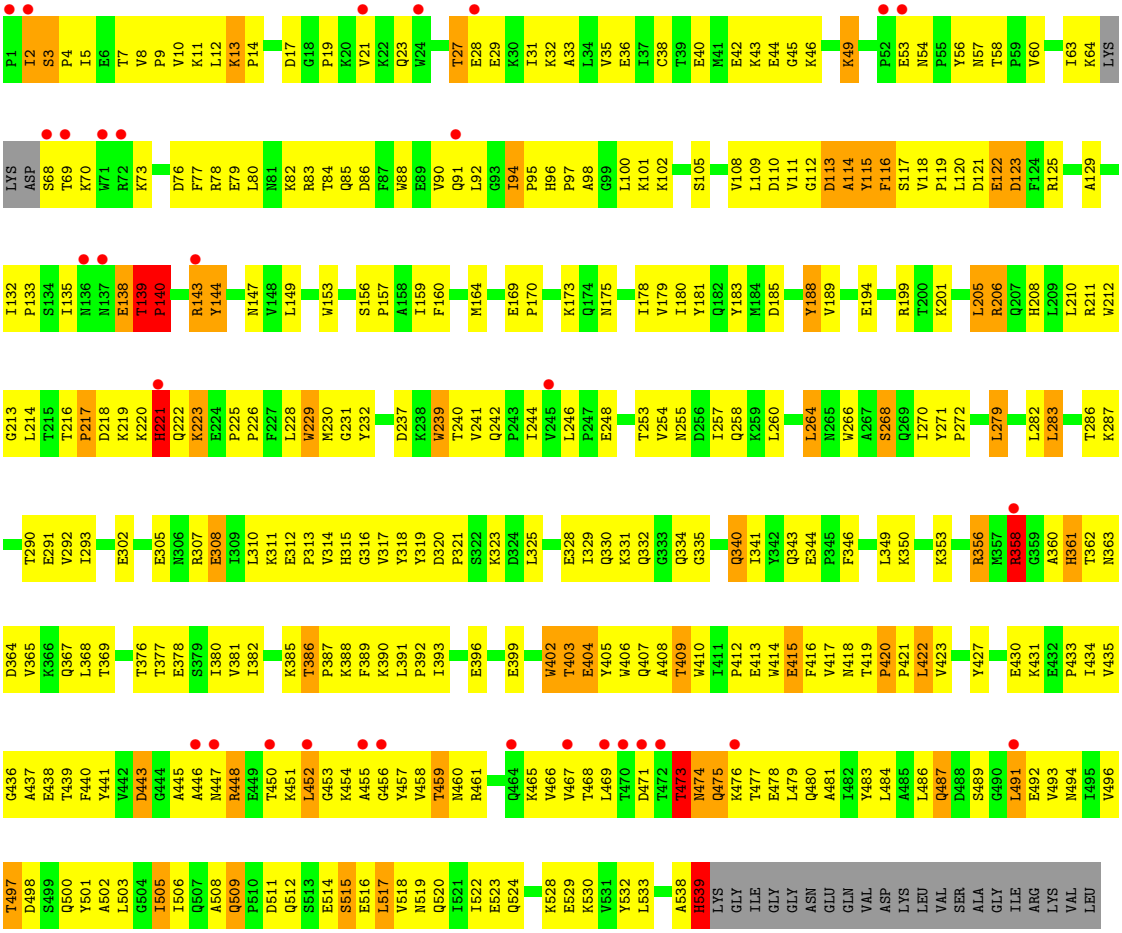
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	6	6	6	0	0
			Total	O		
5	B	2	2	2	0	0

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.

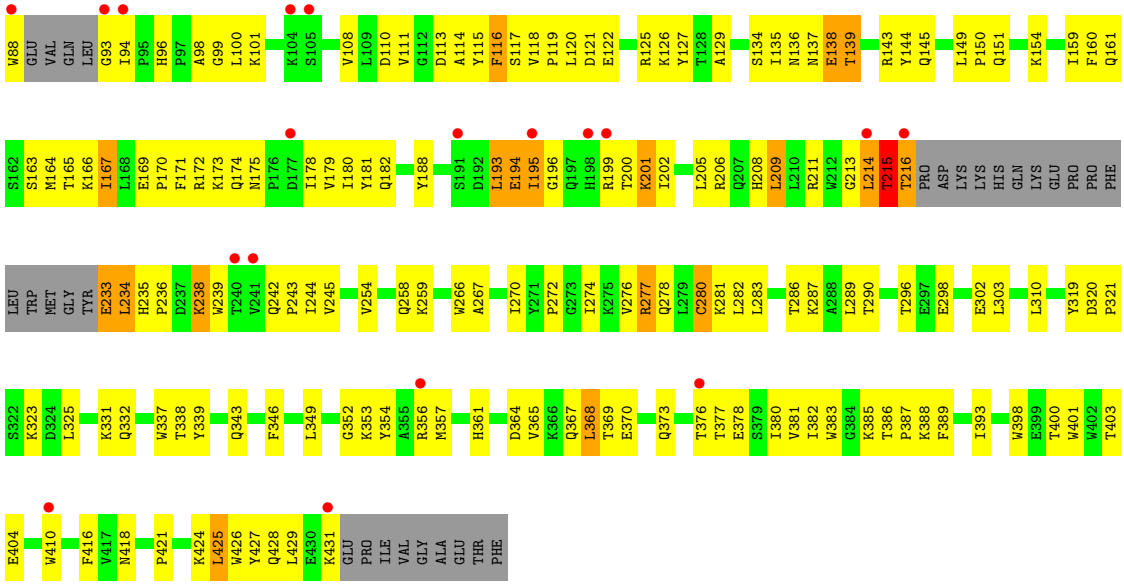
- Molecule 1: Reverse transcriptase/ribonuclease H

Chain A:



- Molecule 2: p51 RT

Chain B:


W88 GLU VAL GLN LEU G93 I94 P95 H96 P97 A98 G99 L100 K101 K104 S105 V108 L109 D110 V111 G112 D113 A114 Y115 F116 S117 V118 P119 L120 D121 E122 R125 K126 Y127 T128 A129 S134 I135 N136 N137 E138 T139 R143 Y144 Q145 L149 P150 Q151 K154 I159 F160 Q161
S162 S163 M164 T165 K166 I167 L168 E169 P170 F171 R172 K173 Q174 N175 P176 D177 I178 V179 I180 Y181 Q182 Y188 S191 D192 L193 E194 I195 G196 Q197 H198 R199 T200 K201 I202 L205 R206 Q207 H208 L209 L210 R211 W212 G213 L214 T215 T216 PRO ASP LYS LYS HIS GLN LYS GLU PRO PRO PHE
LEU TRP MET GLY TYR E233 L234 H235 P236 D237 K238 W239 T240 V241 Q242 P243 I244 V245 V254 Q258 K259 W266 A267 I270 Y271 P272 G273 I274 K275 V276 R277 Q278 L279 C280 K281 L282 L283 T286 K287 A288 L289 T290 T296 E297 E298 E302 L303 L310 Y319 D320 P321
S322 K323 D324 L325 K331 Q332 W337 T338 Y339 Q343 F346 L349 G352 K353 Y354 A355 R356 M357 H361 D364 V365 K366 Q367 L368 T369 E370 Q373 T376 T377 E378 S379 I380 V381 I382 W383 G384 K385 T386 P387 K388 F389 I393 W398 E399 T400 W401 W402 T403
E404 W410 F416 Y417 N418 P421 K424 L425 W426 Y427 Q428 L429 E430 K431 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	136.94Å 109.77Å 72.48Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.79 – 2.90 29.79 – 2.90	Depositor EDS
% Data completeness (in resolution range)	99.5 (29.79-2.90) 99.6 (29.79-2.90)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.46 (at 2.90Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.218 , 0.312 0.209 , 0.301	Depositor DCC
R_{free} test set	1203 reflections (4.86%)	wwPDB-VP
Wilson B-factor (Å^2)	80.4	Xtriage
Anisotropy	0.333	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 69.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	7750	wwPDB-VP
Average B, all atoms (Å^2)	75.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.91% 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: PO4, GWJ, 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	0.59	0/4491	0.85	7/6107 (0.1%)
2	B	0.57	0/3412	0.81	3/4635 (0.1%)
All	All	0.58	0/7903	0.83	10/10742 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	2

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	215	THR	N-CA-C	8.64	134.32	111.00
1	A	539	HIS	N-CA-C	-7.76	90.03	111.00
1	A	505	ILE	CB-CA-C	-6.26	99.07	111.60
1	A	140	PRO	N-CA-C	6.13	128.04	112.10
2	B	214	LEU	CA-CB-CG	-6.06	101.36	115.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	144	TYR	Sidechain
1	A	188	TYR	Sidechain

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	4383	0	4416	367	0
2	B	3320	0	3346	207	0
3	A	5	0	0	0	0
4	A	34	0	17	5	0
5	A	6	0	0	1	0
5	B	2	0	0	0	0
All	All	7750	0	7779	559	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 36.

The worst 5 of 559 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:139:THR:HB	1:A:140:PRO:HD2	1.28	1.14
1:A:448:ARG:H	1:A:448:ARG:HD3	0.93	1.09
1:A:49:LYS:H	1:A:49:LYS:HD3	0.97	1.05
1:A:49:LYS:H	1:A:49:LYS:CD	1.64	1.03
1:A:448:ARG:HD3	1:A:448:ARG:N	1.75	0.99

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	531/560 (95%)	441 (83%)	72 (14%)	18 (3%)	3 15

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
2	B	394/440 (90%)	347 (88%)	39 (10%)	8 (2%)	7 27
All	All	925/1000 (92%)	788 (85%)	111 (12%)	26 (3%)	5 19

5 of 26 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	2	ILE
1	A	114	ALA
1	A	116	PHE
1	A	140	PRO
1	A	223	LYS

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	480/499 (96%)	414 (86%)	66 (14%)	3 10
2	B	365/400 (91%)	340 (93%)	25 (7%)	16 42
All	All	845/899 (94%)	754 (89%)	91 (11%)	6 20

5 of 91 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	491	LEU
2	B	139	THR
1	A	497	THR
1	A	539	HIS
2	B	194	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 28 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	57	ASN
2	B	428	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	B	182	GLN
2	B	336	GLN
2	B	175	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	0.76	0	1,8,10	0.82	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	-	1/2/6/8	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) torsion outliers are listed below:

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

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)

2 ligands are 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	GWJ	A	999	-	36,36,36	2.63	10 (27%)	52,52,52	1.53	10 (19%)
3	PO4	A	1300	-	4,4,4	1.67	0	6,6,6	0.43	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
4	GWJ	A	999	-	-	2/25/25/25	0/3/3/3

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	999	GWJ	C17-C18	6.44	1.53	1.40
4	A	999	GWJ	C5-C6	6.35	1.50	1.38
4	A	999	GWJ	C2-C1	6.10	1.52	1.40
4	A	999	GWJ	C19-C20	5.57	1.49	1.39
4	A	999	GWJ	C22-C21	4.99	1.47	1.38

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	999	GWJ	O5-S-O4	-4.23	111.80	118.76
4	A	999	GWJ	O5-S-N3	3.81	113.02	107.36
4	A	999	GWJ	C20-S-N3	-3.21	103.84	108.38
4	A	999	GWJ	C22-C17-C18	-2.88	117.37	120.77
4	A	999	GWJ	C15-O2-C1	2.61	122.99	117.60

There are no chirality outliers.

All (2) torsion outliers are listed below:

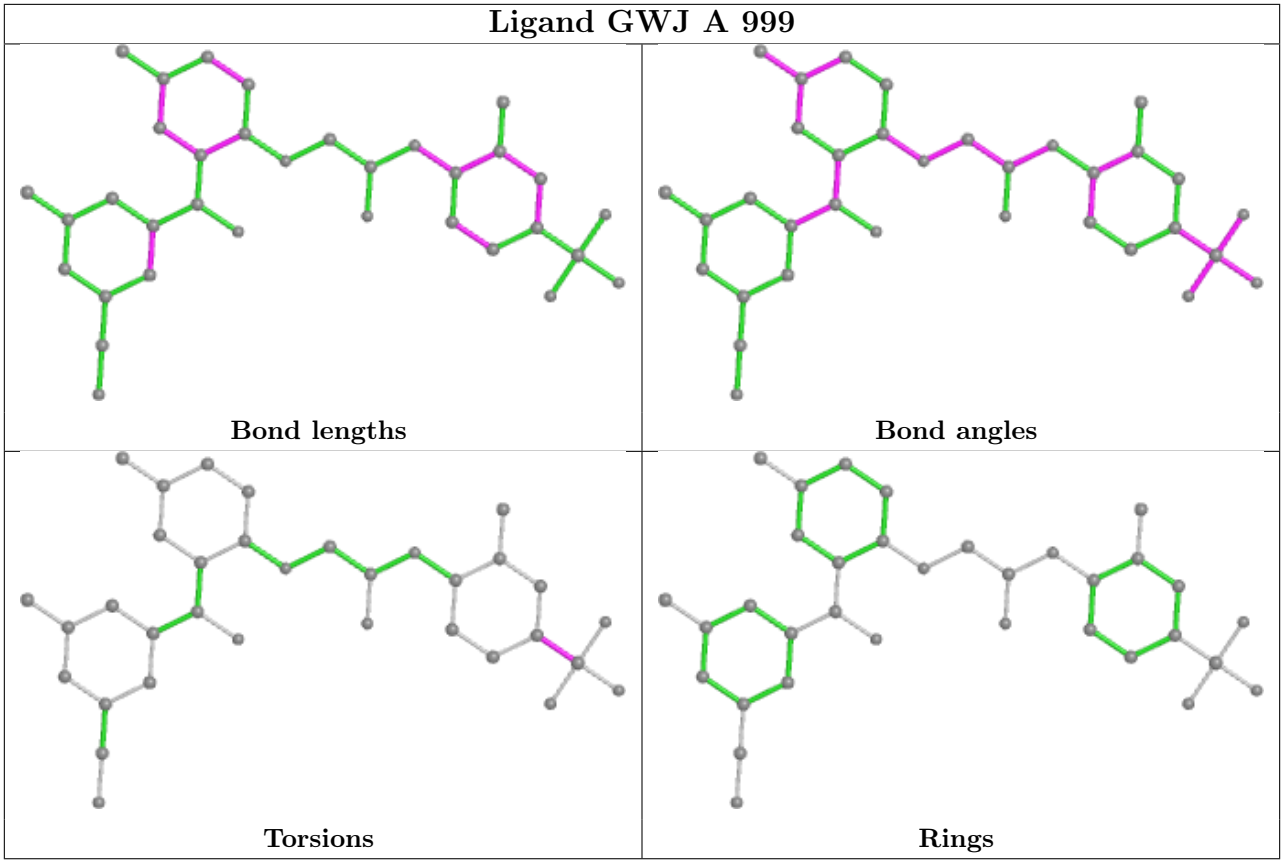
Mol	Chain	Res	Type	Atoms
4	A	999	GWJ	C19-C20-S-O5
4	A	999	GWJ	C21-C20-S-O5

There are no ring outliers.

1 monomer is involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	999	GWJ	5	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.



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	535/560 (95%)	0.02	32 (5%)	21	18	37, 72, 116, 174	0
2	B	402/440 (91%)	0.02	22 (5%)	25	21	32, 69, 120, 148	0
All	All	937/1000 (93%)	0.02	54 (5%)	23	19	32, 71, 119, 174	0

The worst 5 of 54 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	68	SER	5.8
1	A	471	ASP	5.6
2	B	431	LYS	5.6
1	A	2	ILE	5.3
1	A	470	THR	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	58,64,81,94	0

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

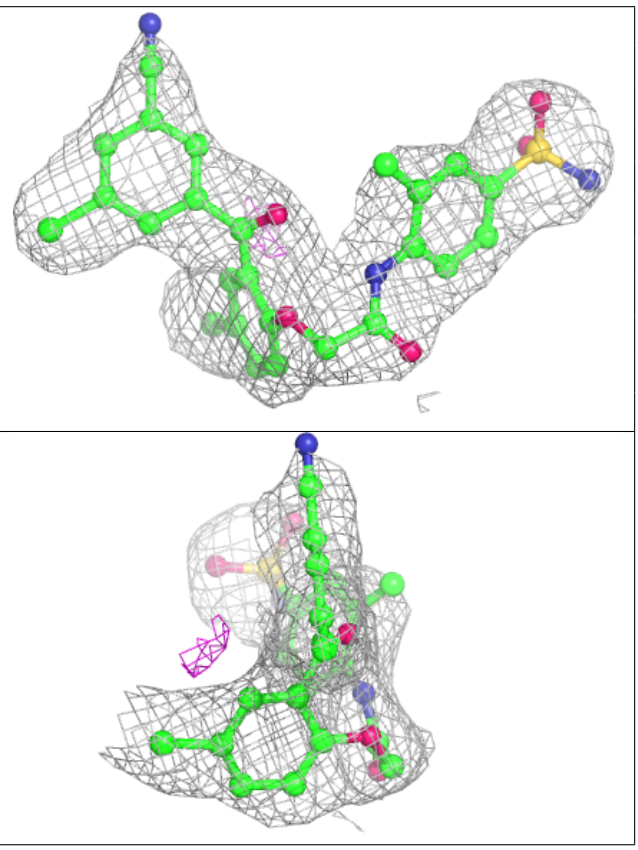
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists 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	PO4	A	1300	5/5	0.85	0.21	132,137,152,154	0
4	GWJ	A	999	34/34	0.97	0.17	35,59,75,78	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 GWJ 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.