wwPDB X-ray Structure Validation Summary Report (i)

Jun 26, 2024 – 05:10 AM EDT

PDB ID	:	6VKA
Title	:	HIV Integrase Core domain (IN) in complex with dimer-spanning ligand
Authors	:	Gorman, M.A.; Parker, M.W.
Deposited on	:	2020-01-19
Resolution	:	1.86 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:

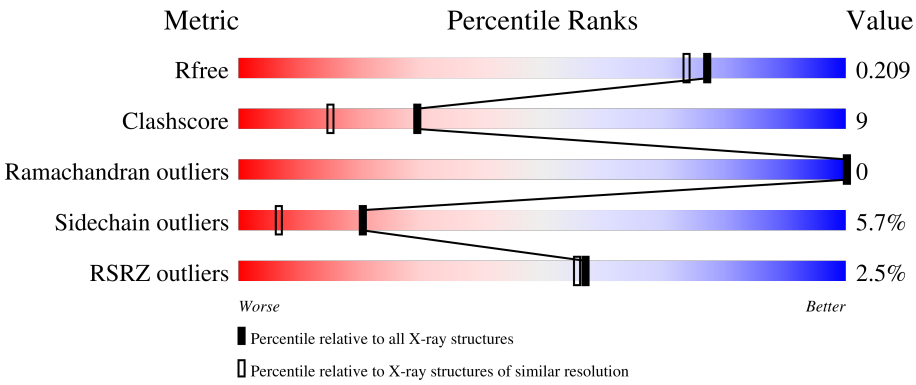
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 1.86 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	2469 (1.86-1.86)
Clashscore	141614	2625 (1.86-1.86)
Ramachandran outliers	138981	2592 (1.86-1.86)
Sidechain outliers	138945	2592 (1.86-1.86)
RSRZ outliers	127900	2436 (1.86-1.86)

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	165	2% 70% 13% • 15%
1	B	165	2% 73% 13% • 13%

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
2	IOD	B	301	-	-	X	-

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 4638 atoms, of which 2230 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 Integrase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	141	2164	692	1075	187	206	4	0	0	0
1	B	143	2230	711	1117	191	207	4	0	2	0

There are 14 discrepancies between the modelled and reference sequences:

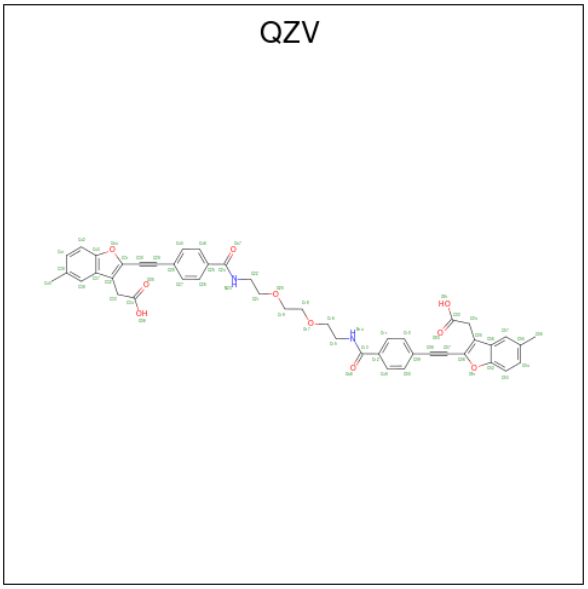
Chain	Residue	Modelled	Actual	Comment	Reference
A	48	GLY	-	expression tag	UNP F2WR39
A	49	SER	-	expression tag	UNP F2WR39
A	53	GLU	GLN	engineered mutation	UNP F2WR39
A	56	SER	CYS	engineered mutation	UNP F2WR39
A	131	GLU	TRP	engineered mutation	UNP F2WR39
A	185	LYS	PHE	engineered mutation	UNP F2WR39
A	209	GLU	GLN	engineered mutation	UNP F2WR39
B	48	GLY	-	expression tag	UNP F2WR39
B	49	SER	-	expression tag	UNP F2WR39
B	53	GLU	GLN	engineered mutation	UNP F2WR39
B	56	SER	CYS	engineered mutation	UNP F2WR39
B	131	GLU	TRP	engineered mutation	UNP F2WR39
B	185	LYS	PHE	engineered mutation	UNP F2WR39
B	209	GLU	GLN	engineered mutation	UNP F2WR39

- Molecule 2 is IODIDE ION (three-letter code: IOD) (formula: I).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	I		
2	A	3	3	3	0	0
2	B	3	3	3	0	0

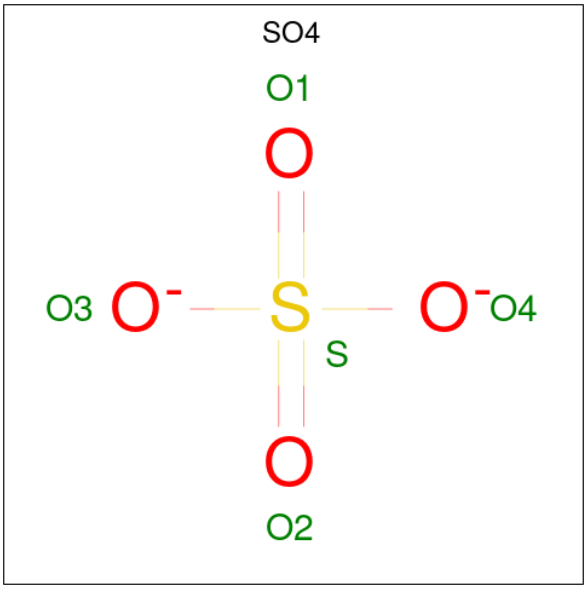
- Molecule 3 is 2,2'-{ethane-1,2-diylbis[oxyethane-2,1-diylcarbamoyl-4,1-phenyleneethyne-

2,1-diyl(5-methyl-1-benzofuran-2,3-diyl)]}diacetic acid (three-letter code: QZV) (formula: $C_{46}H_{40}N_2O_{10}$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	H	N	O		
3	A	1	96	46	38	2	10	0	0

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



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

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total O S 5 4 1	0	0
4	A	1	Total O S 5 4 1	0	0
4	A	1	Total O S 5 4 1	0	0
4	A	1	Total O S 5 4 1	0	0
4	B	1	Total O S 5 4 1	0	0
4	B	1	Total O S 5 4 1	0	0
4	B	1	Total O S 5 4 1	0	0
4	B	1	Total O S 5 4 1	0	0
4	B	1	Total O S 5 4 1	0	0

- Molecule 5 is water.

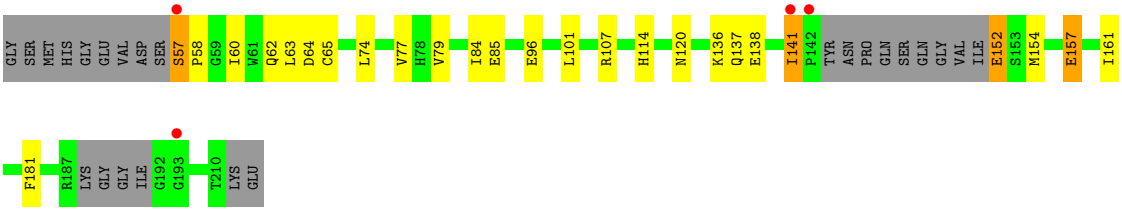
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	51	Total O 51 51	0	0
5	B	41	Total O 41 41	0	0

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.

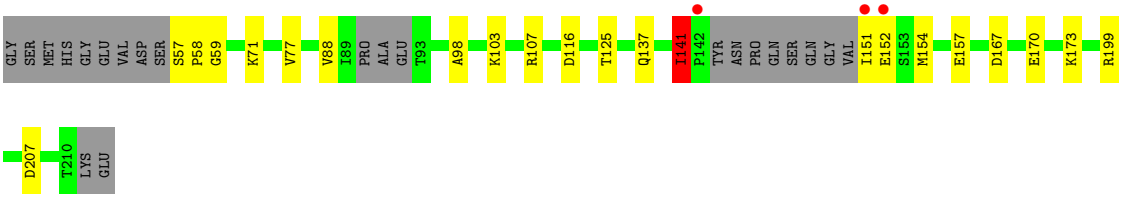
- Molecule 1: Integrase

Chain A:



- Molecule 1: Integrase

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43	Depositor
Cell constants a, b, c, α, β, γ	46.30Å 46.30Å 138.05Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.30 – 1.86 46.30 – 1.86	Depositor EDS
% Data completeness (in resolution range)	98.4 (46.30-1.86) 95.1 (46.30-1.86)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.51 (at 1.87Å)	Xtriage
Refinement program	PHENIX (1.13_2998: ???)	Depositor
R, R_{free}	0.145 , 0.212 0.150 , 0.209	Depositor DCC
R_{free} test set	2024 reflections (8.51%)	wwPDB-VP
Wilson B-factor (Å^2)	21.1	Xtriage
Anisotropy	0.214	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.41 , 34.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.40$, $< L^2 > = 0.23$	Xtriage
Estimated twinning fraction	0.398 for h,-k,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	4638	wwPDB-VP
Average B, all atoms (Å^2)	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.72% 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: SO4, CSO, IOD, QZV

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	0.54	0/1100	0.66	0/1485
1	B	0.56	0/1126	0.73	2/1518 (0.1%)
All	All	0.55	0/2226	0.69	2/3003 (0.1%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	141	ILE	C-N-CD	-11.80	94.63	120.60
1	B	141	ILE	C-N-CA	5.91	146.82	122.00

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	1089	1075	1091	28	0
1	B	1113	1117	1133	17	0
2	A	3	0	0	0	0
2	B	3	0	0	3	0
3	A	58	38	0	0	0
4	A	25	0	0	1	0
4	B	25	0	0	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	A	51	0	0	5	3
5	B	41	0	0	1	3
All	All	2408	2230	2224	43	4

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 9.

The worst 5 of 43 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:77:VAL:HG22	1:A:84:ILE:HG22	1.49	0.94
1:B:151:ILE:HA	1:B:154:MET:HB2	1.52	0.89
1:A:63:LEU:HD11	1:A:74:LEU:HD11	1.62	0.80
1:A:57:SER:HA	1:A:60:ILE:HD13	1.65	0.78
1:A:57:SER:O	5:A:401:HOH:O	2.05	0.73

All (4) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
5:A:434:HOH:O	5:B:436:HOH:O[1_565]	1.82	0.38
5:B:412:HOH:O	5:B:431:HOH:O[3_554]	1.87	0.33
5:A:408:HOH:O	5:A:411:HOH:O[3_654]	2.04	0.16
5:A:434:HOH:O	5:B:428:HOH:O[1_565]	2.17	0.03

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	134/165 (81%)	132 (98%)	2 (2%)	0	100 100
1	B	138/165 (84%)	137 (99%)	1 (1%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
All	All	272/330 (82%)	269 (99%)	3 (1%)	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	114/133 (86%)	108 (95%)	6 (5%)	22	8
1	B	117/133 (88%)	110 (94%)	7 (6%)	19	5
All	All	231/266 (87%)	218 (94%)	13 (6%)	20	7

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	141	ILE
1	B	152	GLU
1	B	199	ARG
1	B	167	ASP
1	B	170	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

2 non-standard protein/DNA/RNA residues 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	$\#\|Z\| > 2$
1	CSO	A	65	1	3,6,7	0.92	0	0,6,8	-	-
1	CSO	B	65	1	3,6,7	0.71	0	0,6,8	-	-

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	CSO	A	65	1	-	0/1/5/7	-
1	CSO	B	65	1	-	0/1/5/7	-

There are no bond length outliers.

There are no bond angle outliers.

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
1	A	65	CSO	1	0

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 17 ligands modelled in this entry, 6 are monoatomic - leaving 11 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	$\#\|Z\| > 2$
4	SO4	A	304	-	4,4,4	0.20	0	6,6,6	0.22	0
4	SO4	B	306	-	4,4,4	0.15	0	6,6,6	0.21	0
3	QZV	A	303	-	53,63,63	1.63	8 (15%)	64,86,86	1.99	10 (15%)
4	SO4	A	308	-	4,4,4	0.26	0	6,6,6	0.38	0
4	SO4	A	306	-	4,4,4	0.20	0	6,6,6	0.24	0
4	SO4	B	303	-	4,4,4	0.16	0	6,6,6	0.44	0
4	SO4	A	307	-	4,4,4	0.26	0	6,6,6	0.58	0
4	SO4	B	305	-	4,4,4	0.23	0	6,6,6	0.28	0
4	SO4	B	304	-	4,4,4	0.18	0	6,6,6	0.24	0
4	SO4	A	305	-	4,4,4	0.20	0	6,6,6	0.18	0
4	SO4	B	307	-	4,4,4	0.10	0	6,6,6	0.13	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
3	QZV	A	303	-	-	7/33/39/39	0/6/6/6

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	303	QZV	C24-N23	5.33	1.45	1.33
3	A	303	QZV	C13-N14	5.12	1.45	1.33
3	A	303	QZV	C37-C43	-5.11	1.32	1.43
3	A	303	QZV	C58-C52	-3.01	1.37	1.43
3	A	303	QZV	C28-C29	2.97	1.51	1.44

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	303	QZV	C39-C38-C37	-6.53	114.26	121.64
3	A	303	QZV	C31-C32-C37	-5.93	102.13	109.56
3	A	303	QZV	C25-C24-N23	5.71	129.35	117.09
3	A	303	QZV	C32-C33-C34	-4.77	108.44	114.18
3	A	303	QZV	C06-C05-C58	-4.15	104.36	109.56

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	303	QZV	O47-C24-N23-C22
3	A	303	QZV	C16-C15-N14-C13
3	A	303	QZV	C25-C24-N23-C22
3	A	303	QZV	O20-C21-C22-N23
3	A	303	QZV	N14-C15-C16-O17

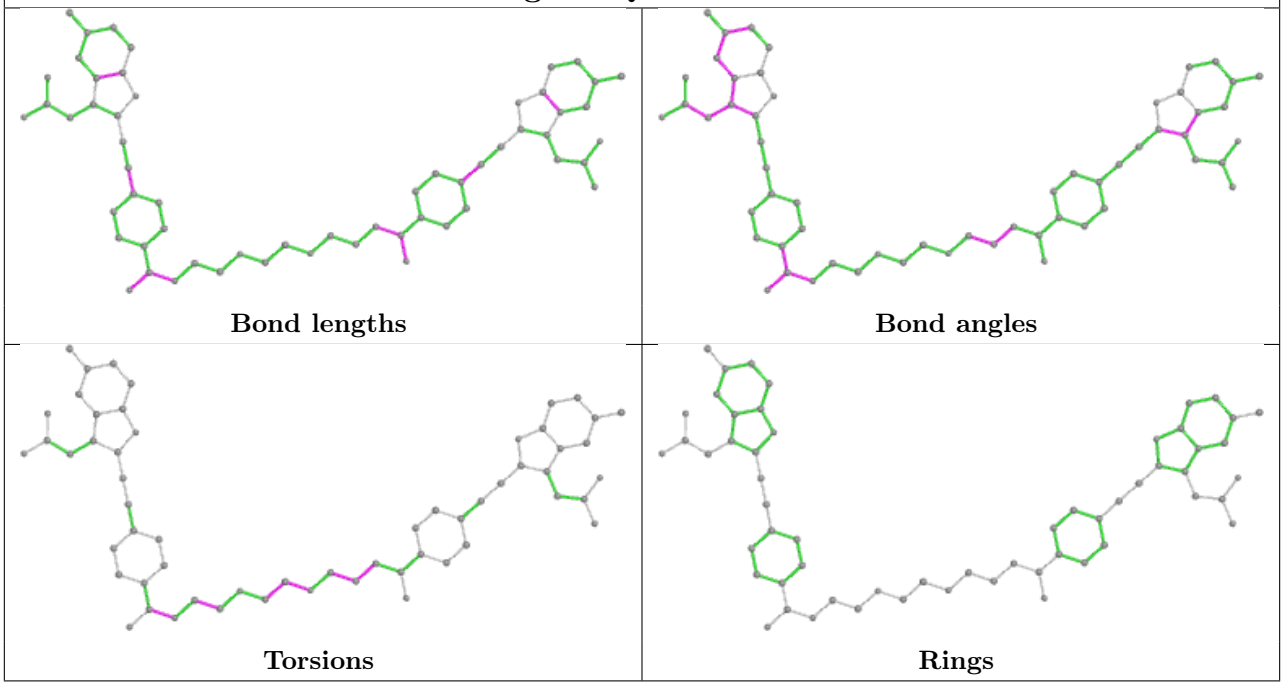
There are no ring outliers.

3 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	307	SO4	1	0
4	B	305	SO4	1	0
4	B	307	SO4	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 QZV A 303



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	140/165 (84%)	0.04	4 (2%)	51	50	12, 21, 42, 54	0
1	B	142/165 (86%)	-0.00	3 (2%)	63	63	12, 23, 40, 62	0
All	All	282/330 (85%)	0.02	7 (2%)	57	56	12, 22, 41, 62	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	142	PRO	4.7
1	B	151	ILE	3.0
1	A	141	ILE	2.8
1	B	152	GLU	2.6
1	A	193	GLY	2.4

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	CSO	B	65	7/8	0.91	0.09	32,36,47,47	0
1	CSO	A	65	7/8	0.94	0.12	23,27,30,36	0

6.3 Carbohydrates (i)

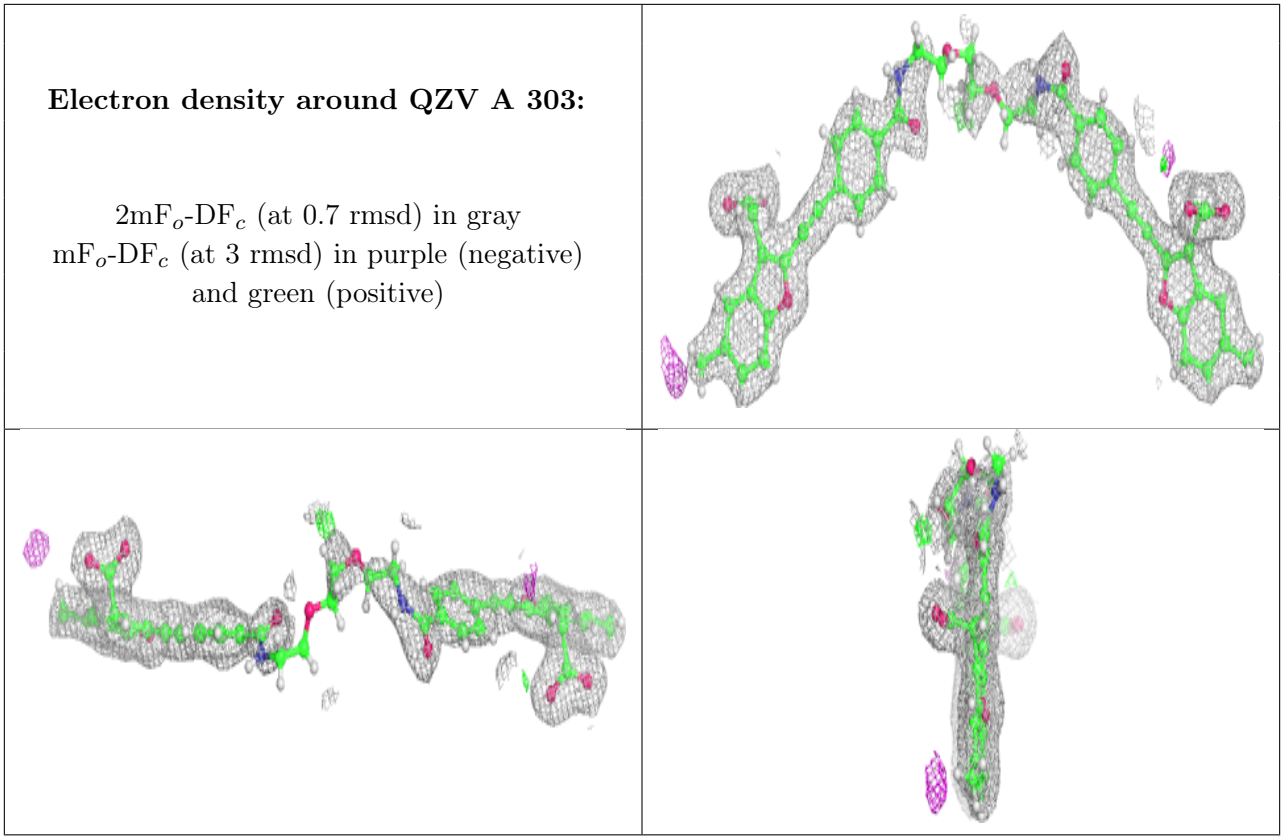
There are no monosaccharides in this entry.

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	QZV	A	303	58/58	0.94	0.14	11,26,75,83	0
4	SO4	B	304	5/5	0.94	0.22	35,36,47,48	0
4	SO4	B	303	5/5	0.95	0.11	37,47,59,59	0
2	IOD	B	308	1/1	0.97	0.04	43,43,43,43	0
4	SO4	A	308	5/5	0.97	0.24	22,33,38,55	0
4	SO4	B	305	5/5	0.97	0.14	31,34,38,45	0
2	IOD	B	301	1/1	0.98	0.04	36,36,36,36	0
4	SO4	A	304	5/5	0.98	0.07	28,32,34,38	0
2	IOD	A	301	1/1	0.98	0.04	38,38,38,38	0
4	SO4	B	306	5/5	0.98	0.06	35,39,42,44	0
4	SO4	B	307	5/5	0.98	0.25	31,36,44,47	0
2	IOD	A	309	1/1	0.99	0.07	30,30,30,30	0
4	SO4	A	305	5/5	0.99	0.08	25,31,34,35	0
4	SO4	A	306	5/5	0.99	0.15	18,30,37,39	0
4	SO4	A	307	5/5	0.99	0.15	15,34,40,43	0
2	IOD	A	302	1/1	0.99	0.10	22,22,22,22	0
2	IOD	B	302	1/1	1.00	0.05	31,31,31,31	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.



6.5 Other polymers (i)

There are no such residues in this entry.