



# Full wwPDB X-ray Structure Validation Report ⓘ

May 22, 2020 – 08:00 am BST

PDB ID : 4YDF  
Title : Crystal structure of compound 9 in complex with HTLV-1 Protease  
Authors : Kuhnert, M.; Blum, A.; Steuber, H.; Diederich, W.E.  
Deposited on : 2015-02-22  
Resolution : 2.80 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto: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 ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) 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 : 2.11  
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) : 2.11

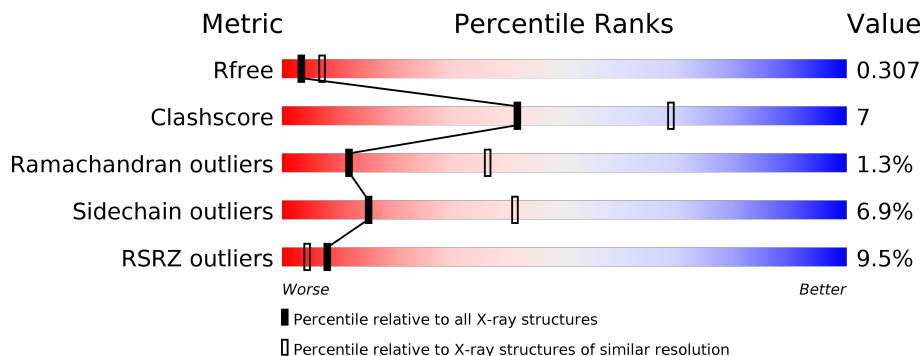
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

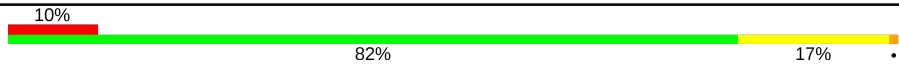
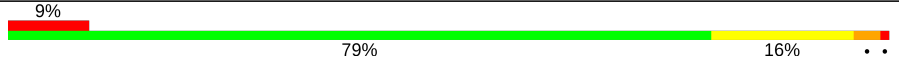
The reported resolution of this entry is 2.80 Å.

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	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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  $\leq 5\%$ . 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	116	
1	B	116	

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 1823 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.

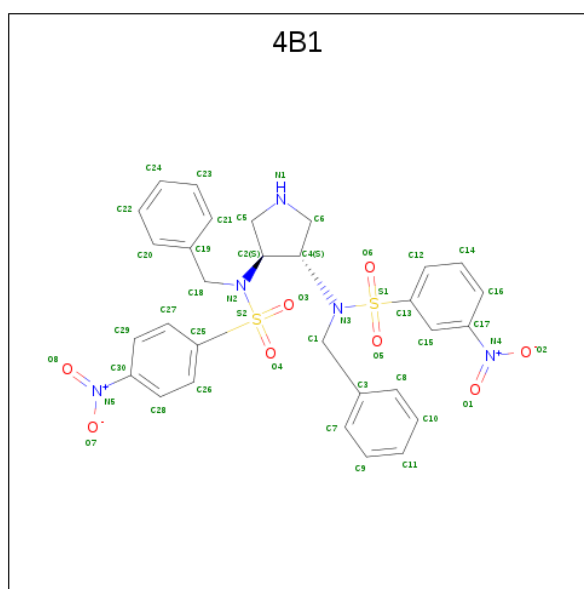
- Molecule 1 is a protein called HTLV-1 Protease.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	116	886	566	152	165	3	0	1	0
1	B	116	883	564	152	164	3	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

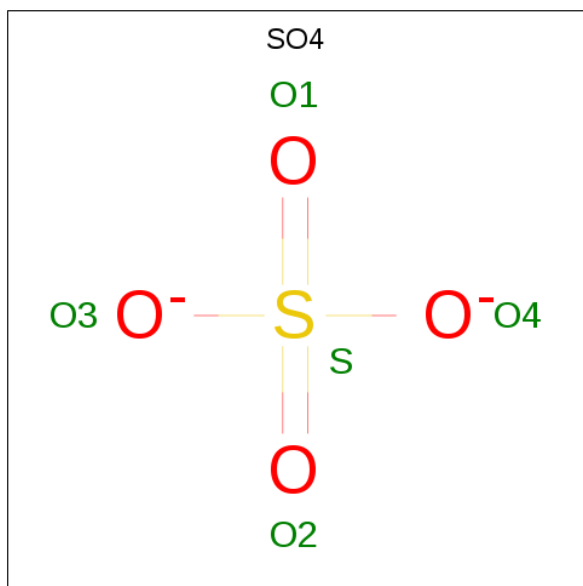
Chain	Residue	Modelled	Actual	Comment	Reference
A	40	ILE	LEU	engineered mutation	UNP Q82134
B	40	ILE	LEU	engineered mutation	UNP Q82134

- Molecule 2 is N-benzyl-N-[(3S,4S)-4-{benzyl[(4-nitrophenyl)sulfonyl]amino}pyrrolidin-3-yl]-3-nitrobenzenesulfonamide (three-letter code: 4B1) (formula: C<sub>30</sub>H<sub>29</sub>N<sub>5</sub>O<sub>8</sub>S<sub>2</sub>).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
2	A	1	45	30	5	8	2	0	0

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



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

- Molecule 4 is water.

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



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 63 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	77.35Å 77.35Å 160.05Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	38.67 – 2.80 38.68 – 2.80	Depositor EDS
% Data completeness (in resolution range)	94.6 (38.67-2.80) 94.7 (38.68-2.80)	Depositor EDS
$R_{merge}$	0.15	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.10 (at 2.81Å)	Xtrriage
Refinement program	PHENIX (phenix.refine: 1.8.1_1168)	Depositor
R, $R_{free}$	0.242 , 0.297 0.245 , 0.307	Depositor DCC
$R_{free}$ test set	355 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	46.6	Xtrriage
Anisotropy	0.604	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 45.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	1823	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	26.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 33.81 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 7.5357e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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: 4B1, SO4

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	0.20	0/907	0.41	0/1241
1	B	0.53	1/901 (0.1%)	0.55	1/1233 (0.1%)
All	All	0.40	1/1808 (0.1%)	0.49	1/2474 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	50	PRO	N-CD	11.82	1.64	1.47

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	50	PRO	CA-N-CD	-7.25	101.35	111.50

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	886	0	931	14	0
1	B	883	0	926	18	0
2	A	45	0	28	3	0
3	A	5	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	1	0	0	0	0
4	B	3	0	0	0	0
All	All	1823	0	1885	26	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 7.

All (26) 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:58:GLY:HA3	1:B:59:ALA:HB2	1.42	0.99
1:A:58:GLY:HA3	1:B:59:ALA:CB	2.00	0.91
1:A:110:GLN:HB3	1:B:116:PRO:HG2	1.71	0.73
1:B:65:ASP:N	1:B:65:ASP:OD1	2.27	0.68
1:B:18:ASP:HB3	1:B:75:LEU:HB2	1.76	0.66
1:A:27:ILE:HD13	1:A:44:LEU:HD11	1.84	0.59
1:A:36:ASP:OD1	1:A:103:ARG:NH2	2.37	0.58
1:B:48:ASN:N	1:B:48:ASN:OD1	2.25	0.57
1:B:42:ILE:HD12	1:B:94:THR:HG22	1.89	0.55
2:A:201:4B1:H7	2:A:201:4B1:C8	2.38	0.54
1:A:42:ILE:HD12	1:A:94:THR:HG22	1.89	0.53
1:A:58:GLY:HA3	1:B:59:ALA:HB3	1.88	0.53
1:B:49:THR:O	1:B:49:THR:HG22	2.09	0.53
1:B:36:ASP:OD1	1:B:103:ARG:NH2	2.39	0.52
1:A:59:ALA:HB3	1:B:58:GLY:HA3	1.91	0.52
1:A:52:LYS:HB2	1:A:69:LEU:HB3	1.93	0.51
1:A:18:ASP:HB3	1:A:75:LEU:HB2	1.93	0.50
1:A:58:GLY:HA2	2:A:201:4B1:H22	1.92	0.49
1:B:46:SER:OG	1:B:47:SER:N	2.48	0.47
2:A:201:4B1:H12	1:B:58:GLY:HA2	2.00	0.43
1:A:58:GLY:CA	1:B:59:ALA:CB	2.85	0.42
1:B:39:VAL:HG22	1:B:91:LEU:HB2	2.00	0.42
1:B:41:PRO:HB3	1:B:97:ASN:HB3	2.01	0.42
1:A:103:ARG:HA	1:A:106:LEU:HB2	2.01	0.42
1:B:75:LEU:HD13	1:B:86:VAL:HG22	2.01	0.42
1:A:10:ARG:NH1	1:B:36:ASP:OD2	2.53	0.41

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	115/116 (99%)	108 (94%)	6 (5%)	1 (1%)	17	46
1	B	114/116 (98%)	104 (91%)	8 (7%)	2 (2%)	8	28
All	All	229/232 (99%)	212 (93%)	14 (6%)	3 (1%)	12	36

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	48	ASN
1	B	59	ALA
1	A	79	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	103/102 (101%)	97 (94%)	6 (6%)	20	50
1	B	102/102 (100%)	94 (92%)	8 (8%)	12	35
All	All	205/204 (100%)	191 (93%)	14 (7%)	15	42

All (14) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	25	LYS
1	A	48	ASN
1	A	64	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	81	ARG
1	A	110	GLN
1	A	113	LEU
1	B	31	LEU
1	B	47	SER
1	B	48	ASN
1	B	50	PRO
1	B	57	LEU
1	B	65	ASP
1	B	80	PHE
1	B	113	LEU

Some 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](#)

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

## 5.5 Carbohydrates [i](#)

There are no carbohydrates 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
3	SO4	A	202	-	4,4,4	0.14	0	6,6,6	0.04	0
2	4B1	A	201	-	45,49,49	8.80	13 (28%)	55,71,71	12.61	17 (30%)

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
2	4B1	A	201	-	-	2/42/58/58	0/5/5/5

All (13) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	201	4B1	C13-S1	45.02	2.39	1.76
2	A	201	4B1	S2-N2	-20.80	1.31	1.63
2	A	201	4B1	S1-N3	-16.23	1.38	1.63
2	A	201	4B1	C25-S2	-14.73	1.55	1.76
2	A	201	4B1	O8-N5	10.83	1.41	1.22
2	A	201	4B1	O1-N4	10.71	1.41	1.22
2	A	201	4B1	C18-C19	-9.15	1.35	1.51
2	A	201	4B1	C1-C3	-7.89	1.37	1.51
2	A	201	4B1	C17-N4	-6.06	1.30	1.45
2	A	201	4B1	C30-N5	-5.78	1.31	1.45
2	A	201	4B1	O6-S1	4.68	1.48	1.43
2	A	201	4B1	O4-S2	4.44	1.48	1.43
2	A	201	4B1	C18-N2	-2.75	1.44	1.48

All (17) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	201	4B1	C12-C13-S1	-82.31	33.10	119.76
2	A	201	4B1	C15-C13-S1	28.18	150.71	119.34
2	A	201	4B1	O5-S1-C13	22.63	136.67	108.05
2	A	201	4B1	O6-S1-C13	-18.07	85.18	108.05
2	A	201	4B1	O6-S1-O5	-9.22	104.57	119.52
2	A	201	4B1	O4-S2-O3	-8.01	106.55	119.52
2	A	201	4B1	O6-S1-N3	4.74	115.62	106.97
2	A	201	4B1	C17-C15-C13	4.54	122.85	119.26
2	A	201	4B1	O4-S2-N2	4.47	115.12	106.97
2	A	201	4B1	C13-S1-N3	-4.31	99.69	107.36
2	A	201	4B1	C27-C25-S2	3.55	123.50	119.76

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	201	4B1	C29-C30-N5	3.34	121.89	119.38
2	A	201	4B1	O5-S1-N3	3.11	112.64	106.97
2	A	201	4B1	C16-C17-N4	2.82	121.50	119.38
2	A	201	4B1	C25-S2-N2	2.67	112.11	107.36
2	A	201	4B1	C18-N2-S2	2.54	122.05	117.89
2	A	201	4B1	C12-C13-C15	-2.45	117.61	120.62

There are no chirality outliers.

All (2) torsion outliers are listed below:

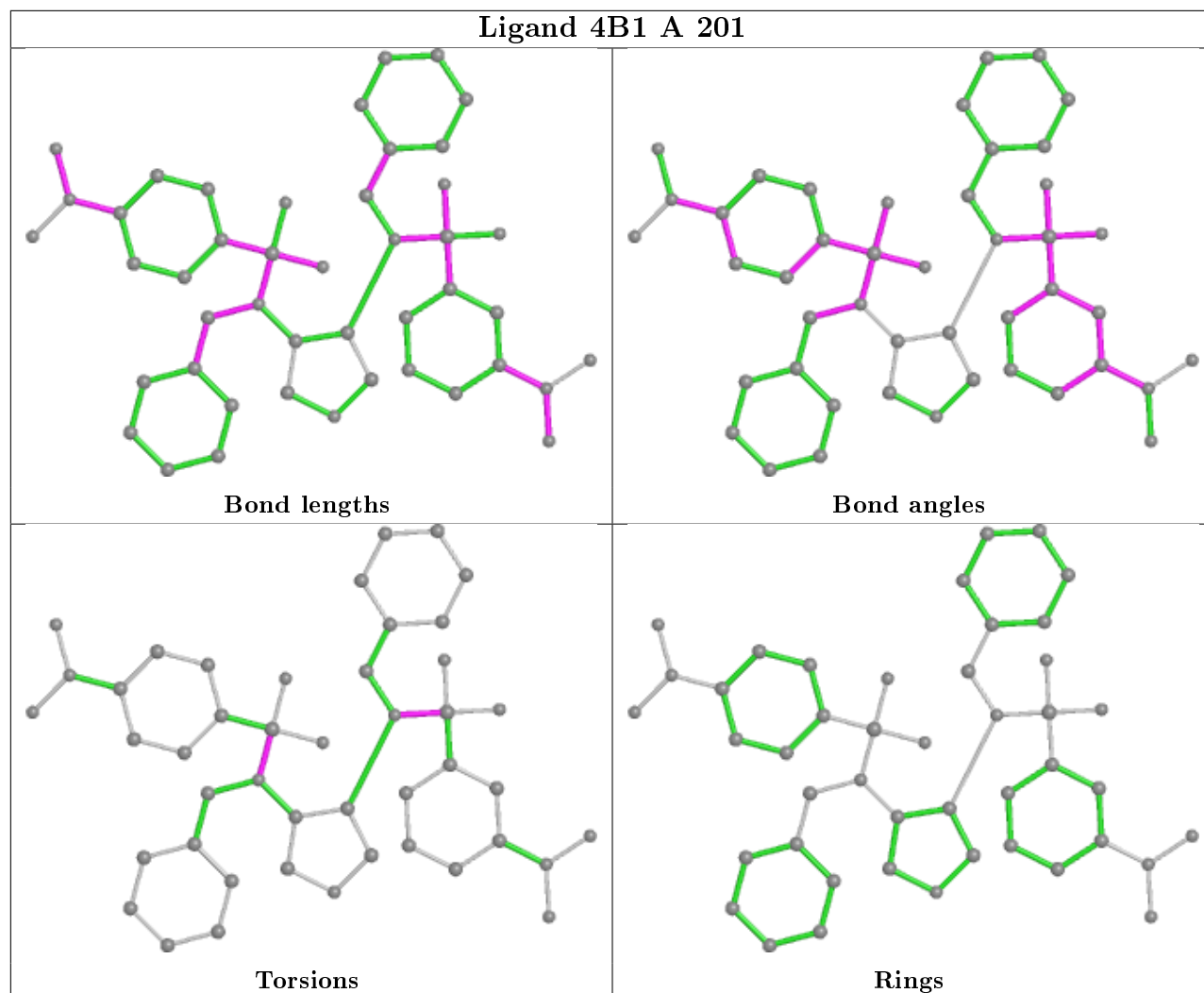
Mol	Chain	Res	Type	Atoms
2	A	201	4B1	C4-N3-S1-C13
2	A	201	4B1	C18-N2-S2-O3

There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	201	4B1	3	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 than 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

### 6.1 Protein, DNA and RNA chains

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	116/116 (100%)	0.45	12 (10%) <b>6</b> <b>3</b>	3, 19, 71, 83	0
1	B	116/116 (100%)	0.44	10 (8%) <b>10</b> <b>5</b>	4, 17, 76, 90	0
All	All	232/232 (100%)	0.45	22 (9%) <b>8</b> <b>4</b>	3, 17, 75, 90	0

All (22) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	60	GLY	5.1
1	A	58	GLY	5.1
1	A	62	GLN	5.0
1	B	59	ALA	4.9
1	A	61	GLY	4.9
1	B	58	GLY	4.8
1	B	62	GLN	4.2
1	A	59	ALA	4.1
1	A	57	LEU	3.5
1	B	57	LEU	3.5
1	B	55	SER	3.4
1	A	56	VAL	3.2
1	A	53	ASN	3.2
1	B	60	GLY	3.1
1	A	55	SER	3.0
1	B	48	ASN	2.8
1	B	82	THR	2.6
1	B	63	THR	2.2
1	A	80	PHE	2.2
1	B	53	ASN	2.2
1	A	50	PRO	2.1
1	A	97	ASN	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](#)

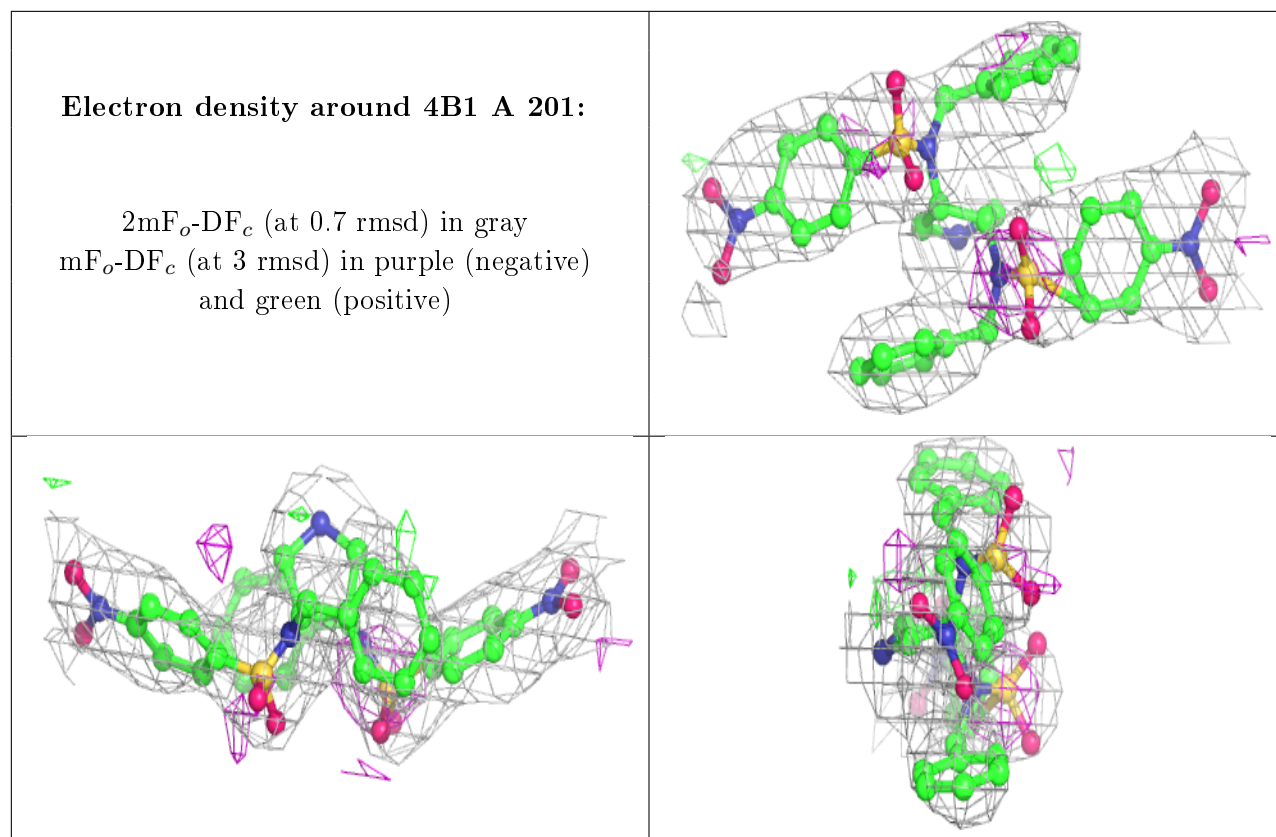
There are no carbohydrates in this entry.

## 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, 95<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
2	4B1	A	201	45/45	0.86	0.29	20,33,54,67	0
3	SO4	A	202	5/5	0.93	0.18	10,17,28,33	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.