



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 14, 2020 – 04:43 am BST

PDB ID : 1P9U  
Title : Coronavirus Main Proteinase (3CLpro) Structure: Basis for Design of anti-SARS Drugs  
Authors : Anand, K.; Ziebuhr, J.; Wadhwani, P.; Mesters, J.R.; Hilgenfeld, R.  
Deposited on : 2003-05-12  
Resolution : 2.37 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](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) : **NOT EXECUTED**  
EDS : **NOT EXECUTED**  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.11

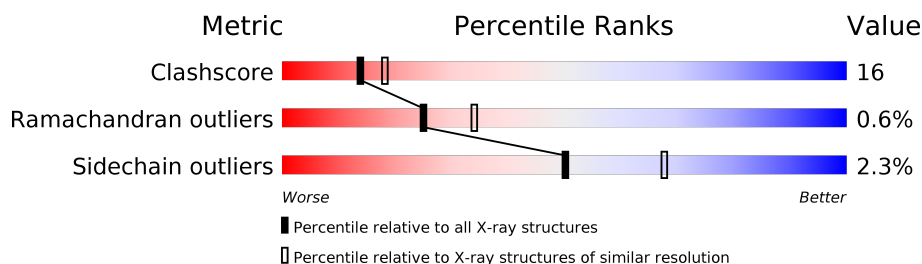
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.37 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	6082 (2.40-2.36)
Ramachandran outliers	138981	5973 (2.40-2.36)
Sidechain outliers	138945	5975 (2.40-2.36)

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  $\leq 5\%$ .

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	302	69% 29% ..
1	B	302	70% 28% ..
1	C	302	75% 24% .
1	D	302	76% 21% .
1	E	302	75% 22% ..
1	F	302	64% 33% ..
2	G	8	50% 25% 13% 13%
2	H	8	25% 38% 25% 13%

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 14981 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 putative coronavirus nsp2 (3CL-PRO).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	299	Total	C	N	O	S	0	0	0
			2297	1448	390	443	16			
1	B	300	Total	C	N	O	S	0	0	0
			2305	1452	392	445	16			
1	C	300	Total	C	N	O	S	0	0	0
			2305	1452	392	445	16			
1	D	301	Total	C	N	O	S	0	0	0
			2313	1458	393	446	16			
1	E	299	Total	C	N	O	S	0	0	0
			2297	1448	390	443	16			
1	F	300	Total	C	N	O	S	0	0	0
			2296	1445	390	445	16			

- Molecule 2 is a protein called PHQ-VNSTLQ-CHLOROMETHYLKETONE INHIBITOR.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	G	7	Total	C	N	O	0	0	1
			46	28	8	10			
2	H	7	Total	C	N	O	0	0	1
			46	28	8	10			

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



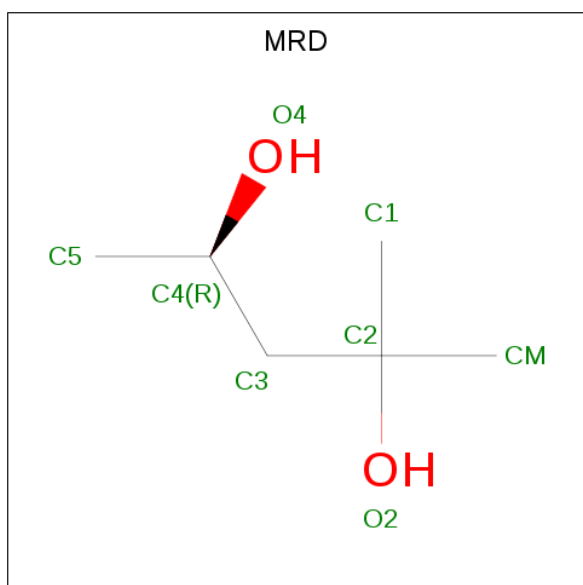
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	O	S	0	0
			5	4	1		
3	A	1	Total	O	S	0	0
			5	4	1		
3	A	1	Total	O	S	0	0
			5	4	1		
3	A	1	Total	O	S	0	0
			5	4	1		
3	B	1	Total	O	S	0	0
			5	4	1		
3	B	1	Total	O	S	0	0
			5	4	1		
3	C	1	Total	O	S	0	0
			5	4	1		
3	C	1	Total	O	S	0	0
			5	4	1		
3	C	1	Total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

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	D	1	Total	O	S	0	0
			5	4	1		
3	D	1	Total	O	S	0	0
			5	4	1		
3	E	1	Total	O	S	0	0
			5	4	1		
3	E	1	Total	O	S	0	0
			5	4	1		
3	E	1	Total	O	S	0	0
			5	4	1		
3	E	1	Total	O	S	0	0
			5	4	1		
3	F	1	Total	O	S	0	0
			5	4	1		
3	F	1	Total	O	S	0	0
			5	4	1		
3	F	1	Total	O	S	0	0
			5	4	1		

- Molecule 4 is (4R)-2-METHYLPENTANE-2,4-DIOL (three-letter code: MRD) (formula:  $C_6H_{14}O_2$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			8	6	2		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	C	1	Total	C	O	0	0
			8	6	2		
4	D	1	Total	C	O	0	0
			8	6	2		
4	E	1	Total	C	O	0	0
			8	6	2		

- Molecule 5 is water.

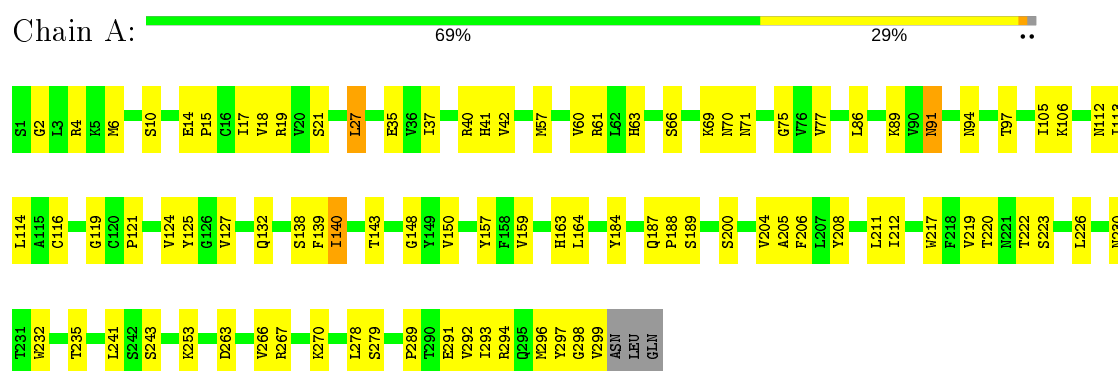
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	153	Total	O	0	0
			153	153		
5	B	193	Total	O	0	0
			193	193		
5	C	120	Total	O	0	0
			120	120		
5	D	150	Total	O	0	0
			150	150		
5	E	213	Total	O	0	0
			213	213		
5	F	95	Total	O	0	0
			95	95		

### 3 Residue-property plots

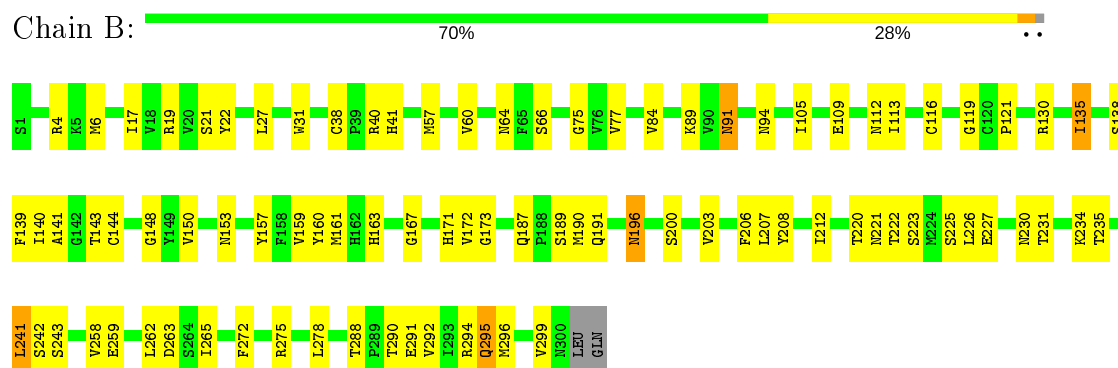
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according 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. 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.

Note EDS was not executed.

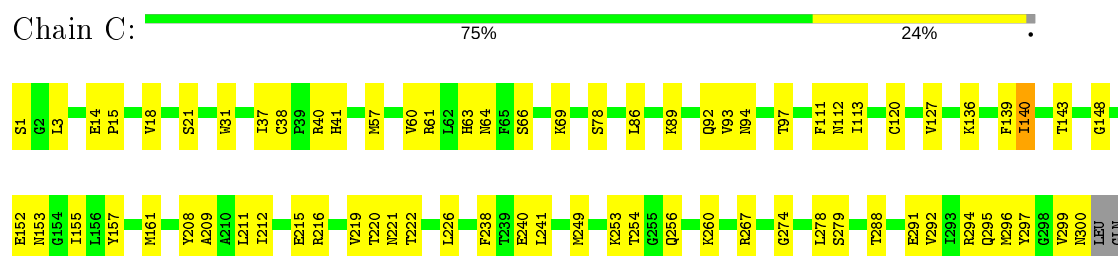
- Molecule 1: putative coronavirus nsp2 (3CL-PRO)




- Molecule 1: putative coronavirus nsp2 (3CL-PRO)

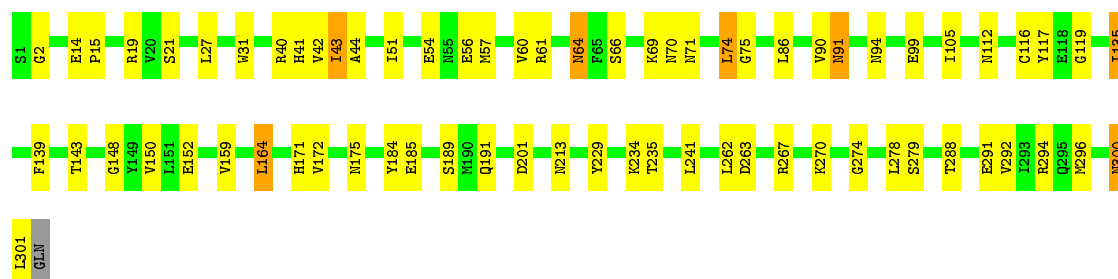


- Molecule 1: putative coronavirus nsp2 (3CL-PRO)




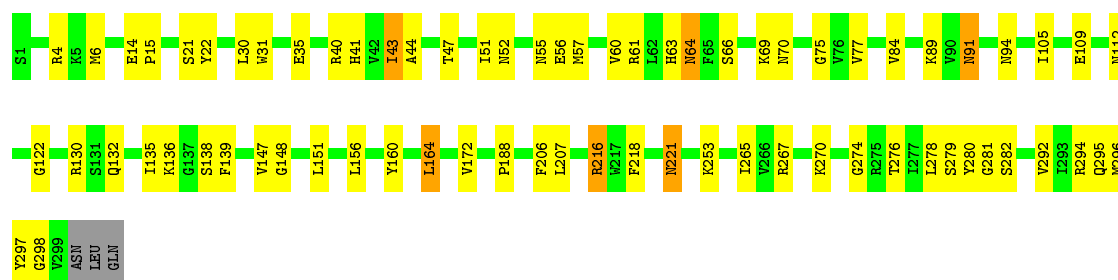
- Molecule 1: putative coronavirus nsp2 (3CL-PRO)

Chain D:  76% 21%



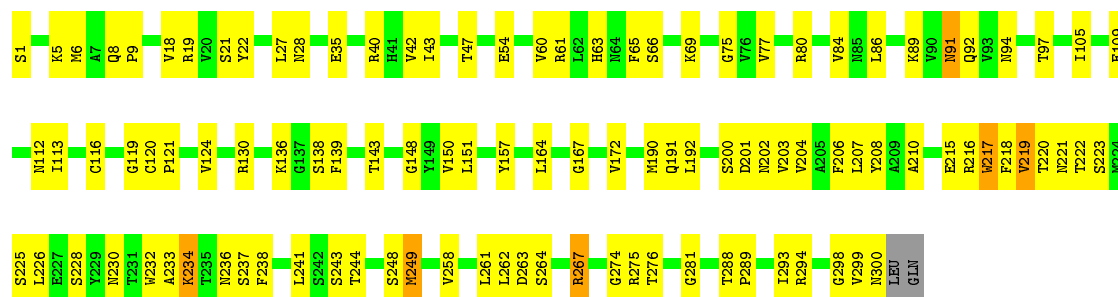
- Molecule 1: putative coronavirus nsp2 (3CL-PRO)

Chain E:  75% 22%



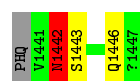
- Molecule 1: putative coronavirus nsp2 (3CL-PRO)

Chain F:  64% 33%




- Molecule 2: PHQ-VNSTLQ-CHLOROMETHYLKETONE INHIBITOR

Chain G:  50% 25% 13% 13%



- Molecule 2: PHQ-VNSTLQ-CHLOROMETHYLKETONE INHIBITOR

Chain H:  25% 38% 25% 13%



PHQ
V1441
D1442
S1443
T1444
L1445
Q1446
T1447

## 4 Data and refinement statistics

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	72.39 Å   158.55 Å   88.20 Å 90.00°   94.40°   90.00°	Depositor
Resolution (Å)	39.64 – 2.37	Depositor
% Data completeness (in resolution range)	98.5 (39.64-2.37)	Depositor
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.191 , 0.234	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	14981	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	42.0	wwPDB-VP

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: MRD, 0QE, 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	0.37	0/2343	0.63	0/3171
1	B	0.35	0/2351	0.64	0/3182
1	C	0.35	0/2351	0.62	0/3182
1	D	0.34	0/2359	0.62	0/3193
1	E	0.36	0/2343	0.64	0/3171
1	F	0.34	0/2342	0.61	0/3172
2	G	0.27	0/44	0.57	0/59
2	H	0.24	0/44	0.52	0/59
All	All	0.35	0/14177	0.62	0/19189

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
2	G	0	1
2	H	0	2
All	All	0	3

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	G	1442	ASN	Peptide
2	H	1442	ASN	Peptide
2	H	1445	LEU	Peptide

## 5.2 Too-close contacts ⓘ

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	2297	0	2247	81	0
1	B	2305	0	2252	91	0
1	C	2305	0	2253	55	0
1	D	2313	0	2264	58	0
1	E	2297	0	2247	64	0
1	F	2296	0	2229	110	0
2	G	46	0	45	14	0
2	H	46	0	45	15	0
3	A	25	0	0	1	0
3	B	10	0	0	0	0
3	C	20	0	0	1	0
3	D	25	0	0	0	0
3	E	25	0	0	0	0
3	F	15	0	0	1	0
4	A	8	0	13	3	0
4	C	8	0	13	0	0
4	D	8	0	13	0	0
4	E	8	0	13	2	0
5	A	153	0	0	5	1
5	B	193	0	0	2	0
5	C	120	0	0	3	0
5	D	150	0	0	7	0
5	E	213	0	0	5	0
5	F	95	0	0	3	1
All	All	14981	0	13634	437	1

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 16.

The worst 5 of 437 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:B:191:GLN:H	2:G:1442:ASN:ND2	1.61	0.98
1:F:191:GLN:H	2:H:1442:ASN:ND2	1.61	0.98
1:B:190:MET:HA	2:G:1442:ASN:HB3	1.47	0.94

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:F:47:THR:HG23	2:H:1445:LEU:HB2	1.51	0.90
1:D:185:GLU:H	1:D:191:GLN:HE22	1.20	0.89

All (1) 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:2081:HOH:O	5:F:2074:HOH:O[2_646]	2.12	0.08

## 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	297/302 (98%)	281 (95%)	15 (5%)	1 (0%)	41	53
1	B	298/302 (99%)	282 (95%)	15 (5%)	1 (0%)	41	53
1	C	298/302 (99%)	287 (96%)	11 (4%)	0	100	100
1	D	299/302 (99%)	286 (96%)	12 (4%)	1 (0%)	41	53
1	E	297/302 (98%)	285 (96%)	12 (4%)	0	100	100
1	F	298/302 (99%)	280 (94%)	14 (5%)	4 (1%)	12	15
2	G	4/8 (50%)	2 (50%)	1 (25%)	1 (25%)	0	0
2	H	4/8 (50%)	2 (50%)	0	2 (50%)	0	0
All	All	1795/1828 (98%)	1705 (95%)	80 (4%)	10 (1%)	25	34

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	298	GLY
1	F	217	TRP
1	D	300	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	F	298	GLY
2	G	1442	ASN

### 5.3.2 Protein sidechains ⓘ

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	255/258 (99%)	249 (98%)	6 (2%)	49	66
1	B	256/258 (99%)	247 (96%)	9 (4%)	36	52
1	C	256/258 (99%)	254 (99%)	2 (1%)	81	91
1	D	257/258 (100%)	250 (97%)	7 (3%)	44	62
1	E	255/258 (99%)	248 (97%)	7 (3%)	44	62
1	F	254/258 (98%)	250 (98%)	4 (2%)	62	78
2	G	6/6 (100%)	6 (100%)	0	100	100
2	H	6/6 (100%)	6 (100%)	0	100	100
All	All	1545/1560 (99%)	1510 (98%)	35 (2%)	50	68

5 of 35 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	140	ILE
1	D	74	LEU
1	F	221	ASN
1	C	221	ASN
1	D	43	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 43 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	64	ASN
1	D	213	ASN
1	F	256	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	D	70	ASN
1	D	91	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

28 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	MRD	E	1004	-	7,7,7	2.07	1 (14%)	9,10,10	0.51	0
3	SO4	B	2006	-	4,4,4	0.23	0	6,6,6	0.12	0
3	SO4	E	2007	-	4,4,4	0.26	0	6,6,6	0.10	0
3	SO4	A	2011	-	4,4,4	0.25	0	6,6,6	0.11	0
3	SO4	F	2024	-	4,4,4	0.26	0	6,6,6	0.05	0
3	SO4	A	2003	-	4,4,4	0.28	0	6,6,6	0.07	0
3	SO4	C	2002	-	4,4,4	0.25	0	6,6,6	0.08	0
3	SO4	C	2005	-	4,4,4	0.21	0	6,6,6	0.16	0
3	SO4	C	2001	-	4,4,4	0.22	0	6,6,6	0.07	0
3	SO4	E	2009	-	4,4,4	0.27	0	6,6,6	0.05	0
3	SO4	D	2012	-	4,4,4	0.28	0	6,6,6	0.07	0
3	SO4	D	2020	-	4,4,4	0.24	0	6,6,6	0.21	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
3	SO4	D	2016	-	4,4,4	0.26	0	6,6,6	0.07	0
3	SO4	C	2022	-	4,4,4	0.24	0	6,6,6	0.06	0
3	SO4	A	2004	-	4,4,4	0.26	0	6,6,6	0.09	0
3	SO4	E	2013	-	4,4,4	0.24	0	6,6,6	0.09	0
3	SO4	E	2017	-	4,4,4	0.27	0	6,6,6	0.05	0
3	SO4	B	2019	-	4,4,4	0.27	0	6,6,6	0.08	0
4	MRD	C	1002	-	7,7,7	2.13	1 (14%)	9,10,10	0.48	0
3	SO4	D	2008	-	4,4,4	0.26	0	6,6,6	0.08	0
3	SO4	E	2018	-	4,4,4	0.26	0	6,6,6	0.05	0
3	SO4	F	2010	-	4,4,4	0.19	0	6,6,6	0.06	0
4	MRD	A	1001	-	7,7,7	2.08	1 (14%)	9,10,10	0.52	0
4	MRD	D	1003	-	7,7,7	2.09	1 (14%)	9,10,10	0.57	0
3	SO4	D	2015	-	4,4,4	0.24	0	6,6,6	0.04	0
3	SO4	F	2023	-	4,4,4	0.24	0	6,6,6	0.05	0
3	SO4	A	2025	-	4,4,4	0.26	0	6,6,6	0.06	0
3	SO4	A	2021	-	4,4,4	0.26	0	6,6,6	0.15	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	MRD	A	1001	-	-	0/5/5/5	-
4	MRD	E	1004	-	-	0/5/5/5	-
4	MRD	D	1003	-	-	0/5/5/5	-
4	MRD	C	1002	-	-	0/5/5/5	-

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	C	1002	MRD	O2-C2	-5.58	1.30	1.44
4	D	1003	MRD	O2-C2	-5.49	1.30	1.44
4	A	1001	MRD	O2-C2	-5.43	1.31	1.44
4	E	1004	MRD	O2-C2	-5.43	1.31	1.44

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.



5 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	E	1004	MRD	2	0
3	A	2011	SO4	1	0
3	C	2001	SO4	1	0
3	F	2010	SO4	1	0
4	A	1001	MRD	3	0

## 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

EDS was not executed - this section is therefore empty.

### 6.2 Non-standard residues in protein, DNA, RNA chains

EDS was not executed - this section is therefore empty.

### 6.3 Carbohydrates

EDS was not executed - this section is therefore empty.

### 6.4 Ligands

EDS was not executed - this section is therefore empty.

### 6.5 Other polymers

EDS was not executed - this section is therefore empty.