wwPDB X-ray Structure Validation Summary Report (i)

Mar 5, 2026 – 07:54 AM UTC

PDB ID	:	7GKP / pdb_00007gkp
Title	:	Group deposition SARS-CoV-2 main protease in complex with inhibitors from the COVID Moonshot – Crystal Structure of SARS-CoV-2 main protease in complex with MAT-POS-4223bc15-3 (Mpro-P1007)
Authors	:	Fearon, D.; Aimon, A.; Aschenbrenner, J.C.; Balcomb, B.H.; Bertram, F.K.R.; Brandao-Neto, J.; Dias, A.; Douangamath, A.; Dunnett, L.; Godoy, A.S.; Gorrie-Stone, T.J.; Koekemoer, L.; Krojer, T.; Lithgo, R.M.; Lukacik, P.; Marples, P.G.; Mikolajek, H.; Nelson, E.; Owen, C.D.; Powell, A.J.; Rangel, V.L.; Skyner, R.; Strain-Damerell, C.M.; Thompson, W.; Tomlinson, C.W.E.; Wild, C.; Walsh, M.A.; von Delft, F.
Deposited on	:	2023-08-11
Resolution	:	1.79 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:

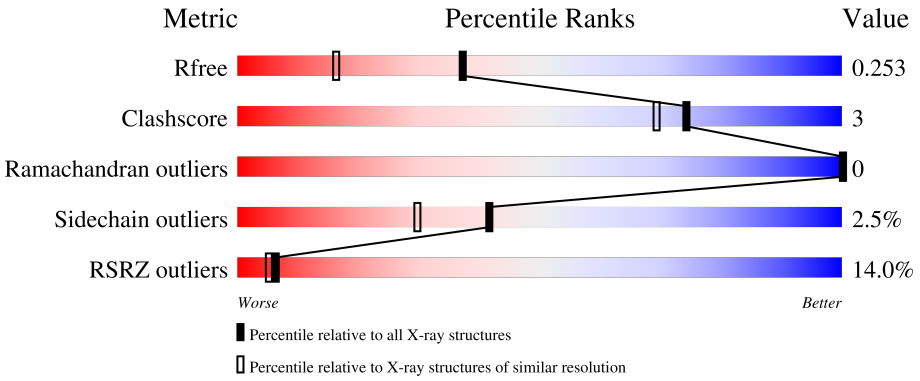
MolProbity	:	4-5-2 with Phenix2.0
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Buster-report	:	wwPDB partial adaption of 1.1.7 (2018)
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)

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 1.79 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}	180053	7662 (1.80-1.80)
Clashscore	190562	8479 (1.80-1.80)
Ramachandran outliers	187476	8391 (1.80-1.80)
Sidechain outliers	187428	8390 (1.80-1.80)
RSRZ outliers	180081	7663 (1.80-1.80)

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 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	306	7% 89% 10%
1	B	306	21% 90% 9%

CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

2 Entry composition (i)

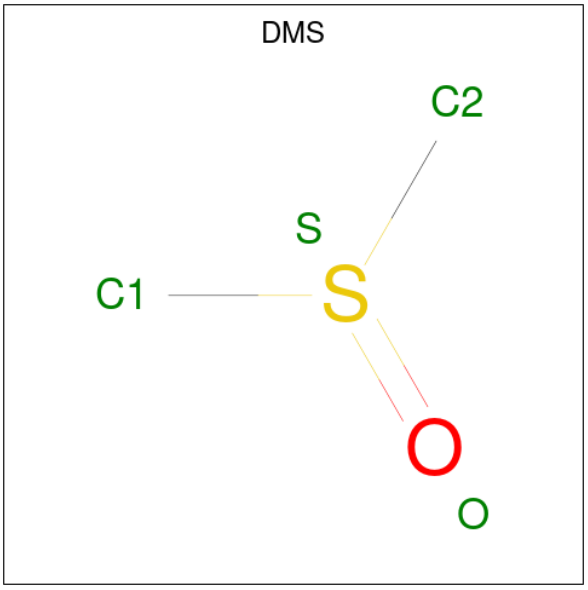
There are 5 unique types of molecules in this entry. The entry contains 5081 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 3C-like proteinase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	302	Total	C	N	O	S	0	0	0
			2333	1476	397	438	22			
1	B	305	Total	C	N	O	S	0	0	0
			2358	1494	400	442	22			

- Molecule 2 is DIMETHYL SULFOXIDE (CCD ID: DMS) (formula: C_2H_6OS).



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

Continued on next page...

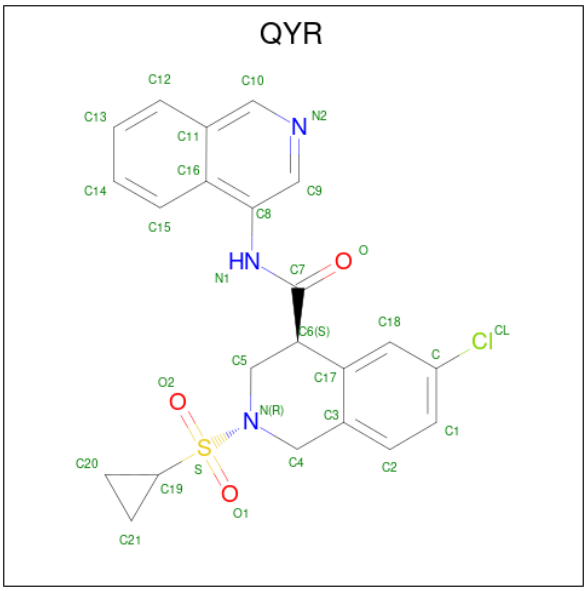
Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C O S 4 2 1 1	0	0
2	A	1	Total C O S 4 2 1 1	0	0
2	B	1	Total C O S 4 2 1 1	0	0
2	B	1	Total C O S 4 2 1 1	0	0
2	B	1	Total C O S 4 2 1 1	0	0
2	B	1	Total C O S 4 2 1 1	0	0

- Molecule 3 is CHLORIDE ION (CCD ID: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total Cl 1 1	0	0

- Molecule 4 is (4S)-6-chloro-2-(cyclopropanesulfonyl)-N-(isoquinolin-4-yl)-1,2,3,4-tetrahydroisoquinoline-4-carboxamide (CCD ID: QYR) (formula: $C_{22}H_{20}ClN_3O_3S$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C Cl N O S 30 22 1 3 3 1	0	0

- Molecule 5 is water.

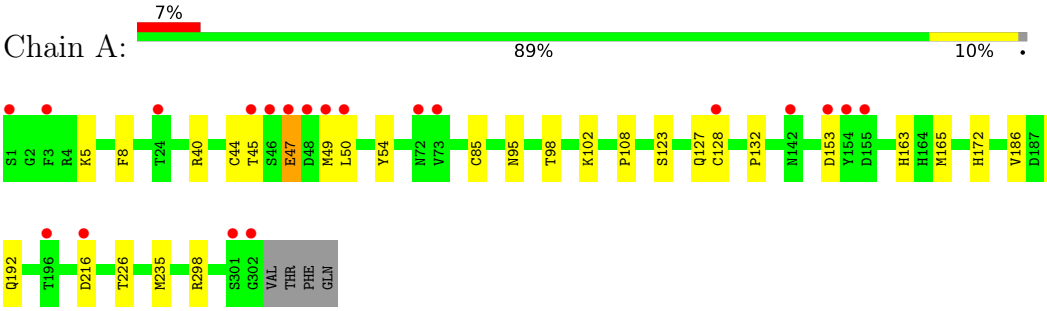
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	201	Total O 201 201	0	0
5	B	118	Total O 118 118	0	0

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.

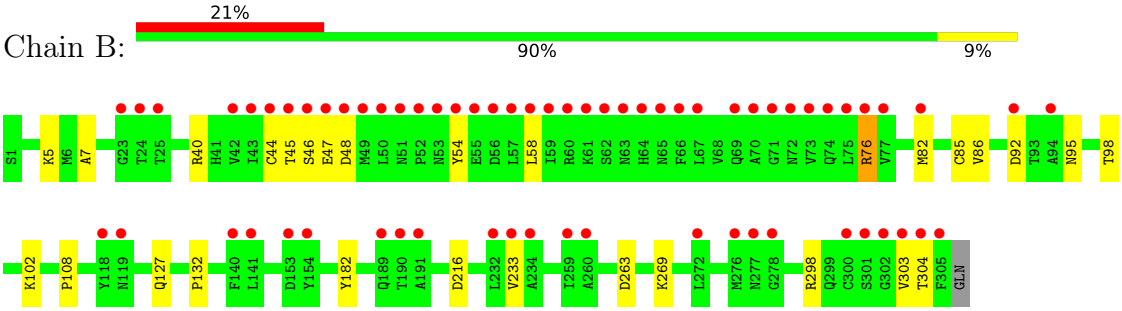
• Molecule 1: 3C-like proteinase

Chain A:



• Molecule 1: 3C-like proteinase

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	67.69Å 99.70Å 103.93Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	56.72 – 1.79 56.72 – 1.79	Depositor EDS
% Data completeness (in resolution range)	60.3 (56.72-1.79) 60.3 (56.72-1.79)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.91 (at 1.80Å)	Xtriage
Refinement program	BUSTER 2.10.4 (20-APR-2021)	Depositor
R, R_{free}	0.218 , 0.263 0.212 , 0.253	Depositor DCC
R_{free} test set	1922 reflections (2.89%)	wwPDB-VP
Wilson B-factor (Å^2)	25.8	Xtriage
Anisotropy	0.086	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 37.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.013 for -h,l,k	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	5081	wwPDB-VP
Average B, all atoms (Å^2)	33.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.04% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: DMS, CL, QYR

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	0.73	0/2385	1.00	2/3241 (0.1%)
1	B	0.74	0/2411	0.98	2/3277 (0.1%)
All	All	0.73	0/4796	0.99	4/6518 (0.1%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	263	ASP	CA-CB-CG	9.92	122.52	112.60
1	A	165	MET	N-CA-C	5.63	117.63	109.07
1	A	226	THR	N-CA-C	-5.45	103.19	110.55
1	B	182	TYR	N-CA-C	-5.01	100.99	108.96

There are no chirality outliers.

There are no planarity outliers.

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	2333	0	2281	15	0
1	B	2358	0	2306	16	0
2	A	24	0	36	4	0
2	B	16	0	24	3	0
3	A	1	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	30	0	0	0	0
5	A	201	0	0	0	0
5	B	118	0	0	0	0
All	All	5081	0	4647	29	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 3.

The worst 5 of 29 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:298:ARG:HD3	2:A:403:DMS:H12	1.56	0.85
1:B:298:ARG:HD2	2:B:402:DMS:H13	1.71	0.72
1:A:123:SER:HB2	1:B:303:VAL:O	1.93	0.69
1:A:123:SER:HB3	1:B:304:THR:HG22	1.83	0.61
1:B:298:ARG:HH11	2:B:402:DMS:H13	1.67	0.58

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	300/306 (98%)	289 (96%)	11 (4%)	0	100	100
1	B	303/306 (99%)	293 (97%)	10 (3%)	0	100	100
All	All	603/612 (98%)	582 (96%)	21 (4%)	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	259/263 (98%)	251 (97%)	8 (3%)	35 23
1	B	262/263 (100%)	257 (98%)	5 (2%)	50 41
All	All	521/526 (99%)	508 (98%)	13 (2%)	42 30

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

Mol	Chain	Res	Type
1	A	235	MET
1	B	47	GLU
1	B	216	ASP
1	B	86	VAL
1	B	102	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	244	GLN
1	B	119	ASN
1	B	164	HIS
1	A	119	ASN
1	A	64	HIS

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 oligosaccharides in this entry.

5.6 Ligand geometry (i)

Of 12 ligands modelled in this entry, 1 is 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	QYR	A	407	-	32,34,34	0.22	0	44,51,51	0.65	1 (2%)
2	DMS	A	408	-	3,3,3	0.32	0	3,3,3	0.26	0
2	DMS	A	404	-	3,3,3	0.27	0	3,3,3	0.18	0
2	DMS	A	403	-	3,3,3	0.23	0	3,3,3	0.79	0
2	DMS	B	401	-	3,3,3	0.32	0	3,3,3	0.22	0
2	DMS	A	405	-	3,3,3	0.35	0	3,3,3	0.27	0
2	DMS	B	403	-	3,3,3	0.28	0	3,3,3	0.45	0
2	DMS	A	402	-	3,3,3	0.31	0	3,3,3	0.38	0
2	DMS	B	404	-	3,3,3	0.32	0	3,3,3	0.62	0
2	DMS	A	401	-	3,3,3	0.30	0	3,3,3	0.41	0
2	DMS	B	402	-	3,3,3	0.08	0	3,3,3	0.38	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	QYR	A	407	-	-	6/20/34/34	0/5/5/5

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	407	QYR	O1-S-N	2.65	110.11	107.44

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	407	QYR	C9-C8-N1-C7
4	A	407	QYR	C20-C19-S-O1
4	A	407	QYR	C5-C6-C7-N1
4	A	407	QYR	C5-C6-C7-O
4	A	407	QYR	C16-C8-N1-C7

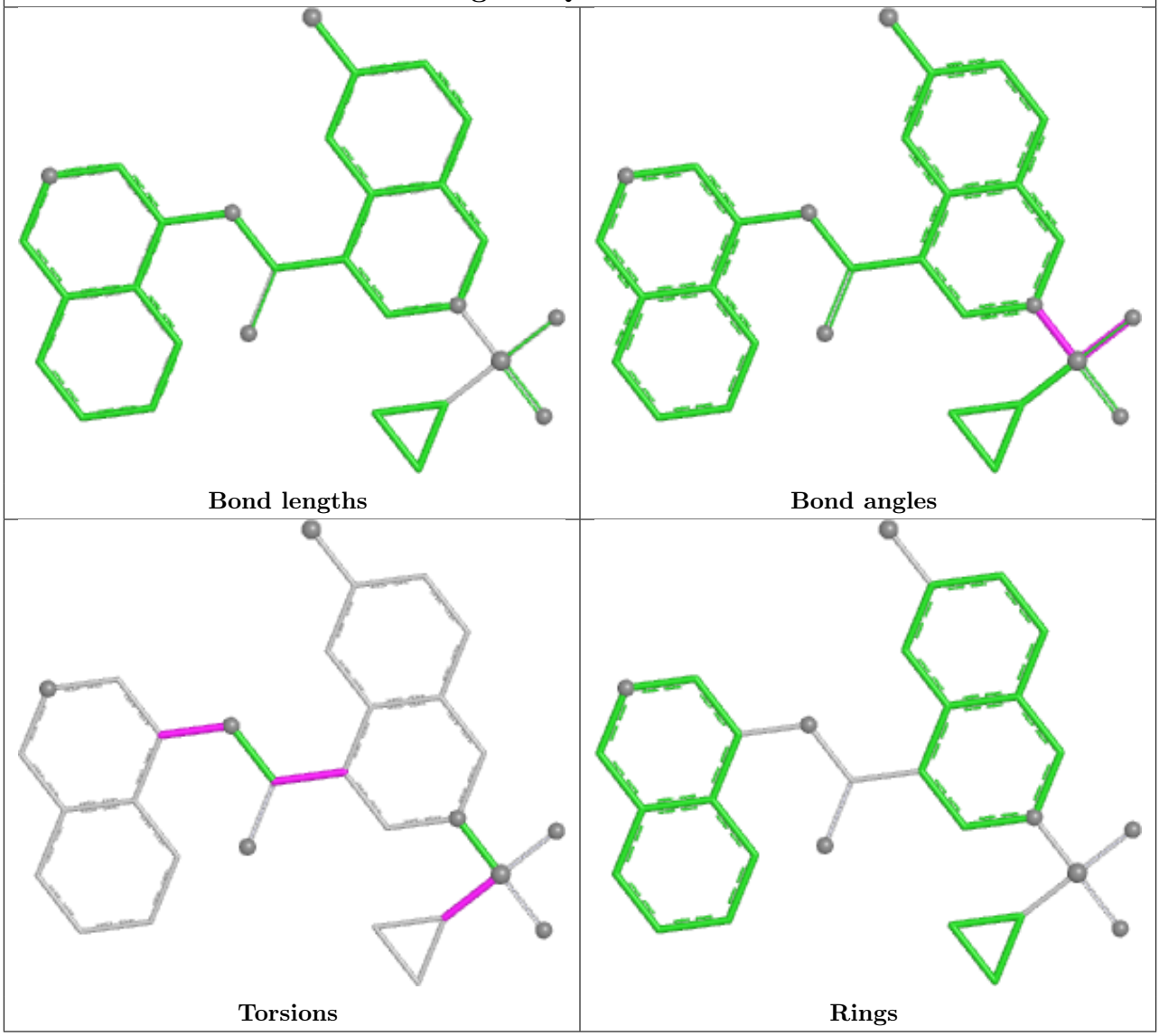
There are no ring outliers.

4 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	403	DMS	3	0
2	A	402	DMS	1	0
2	B	404	DMS	1	0
2	B	402	DMS	2	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 QYR A 407



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, 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	302/306 (98%)	0.46	20 (6%) 24 23	15, 27, 46, 70	0
1	B	305/306 (99%)	1.14	65 (21%) 2 2	17, 32, 73, 104	0
All	All	607/612 (99%)	0.80	85 (14%) 6 5	15, 29, 57, 104	0

The worst 5 of 85 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	50	LEU	9.6
1	B	49	MET	8.0
1	B	54	TYR	7.8
1	B	59	ILE	7.7
1	A	154	TYR	7.4

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

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

6.3 Carbohydrates

There are no oligosaccharides in this entry.

6.4 Ligands

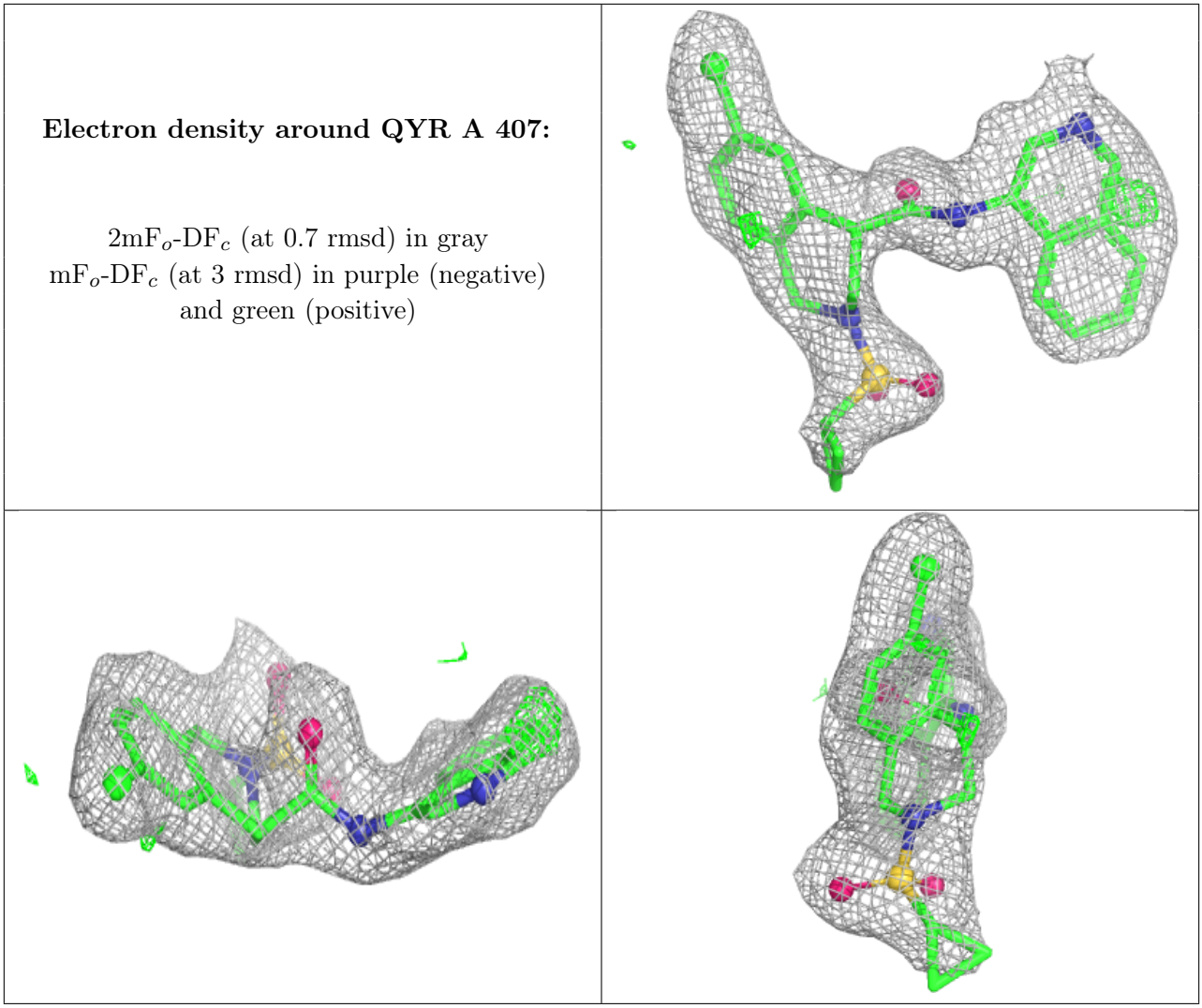
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	DMS	A	405	4/4	0.61	0.35	87,87,87,87	0
2	DMS	A	408	4/4	0.69	0.32	90,90,90,90	0
2	DMS	B	401	4/4	0.69	0.38	114,115,115,115	0
2	DMS	A	402	4/4	0.78	0.27	92,92,92,92	0
2	DMS	B	403	4/4	0.86	0.20	78,78,78,78	0
2	DMS	B	404	4/4	0.86	0.20	58,59,59,59	0
2	DMS	A	403	4/4	0.88	0.18	52,52,52,52	0
4	QYR	A	407	30/30	0.88	0.14	33,39,45,45	30
2	DMS	A	401	4/4	0.89	0.19	57,57,57,57	0
2	DMS	A	404	4/4	0.89	0.18	64,64,65,65	0
2	DMS	B	402	4/4	0.98	0.08	33,33,33,33	0
3	CL	A	406	1/1	0.99	0.05	21,21,21,21	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 QYR A 407:

$2mF_o$-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

There are no such residues in this entry.