



# Full wwPDB X-ray Structure Validation Report ⓘ

Mar 2, 2021 – 10:12 AM EST

PDB ID : 5R0E  
Title : PanDDA analysis group deposition – Aar2/RNaseH in complex with fragment F2X-Entry D02, DMSO-free  
Authors : Wollenhaupt, J.; Metz, A.; Barthel, T.; Lima, G.M.A.; Heine, A.; Mueller, U.; Klebe, G.; Weiss, M.S.  
Deposited on : 2020-02-12  
Resolution : 1.59 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 : 2.17.1  
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) : 2.17.1

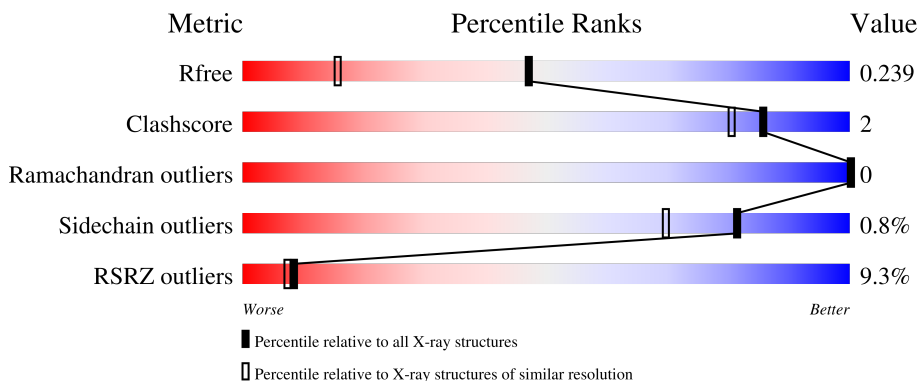
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*


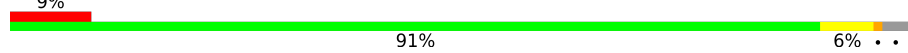
The reported resolution of this entry is 1.59 Å.

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	3398 (1.60-1.60)
Clashscore	141614	3665 (1.60-1.60)
Ramachandran outliers	138981	3564 (1.60-1.60)
Sidechain outliers	138945	3563 (1.60-1.60)
RSRZ outliers	127900	3321 (1.60-1.60)

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  $\geq 3$ , 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	258	
2	B	308	

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 4778 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 Pre-mRNA-splicing factor 8.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	237	2002	1283	335	372	12	0	12	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1833	GLY	-	expression tag	UNP P33334
A	1834	ALA	-	expression tag	UNP P33334
A	1835	MET	-	expression tag	UNP P33334

- Molecule 2 is a protein called A1 cistron-splicing factor AAR2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	300	2580	1654	421	485	20	0	9	0

There are 20 discrepancies between the modelled and reference sequences:

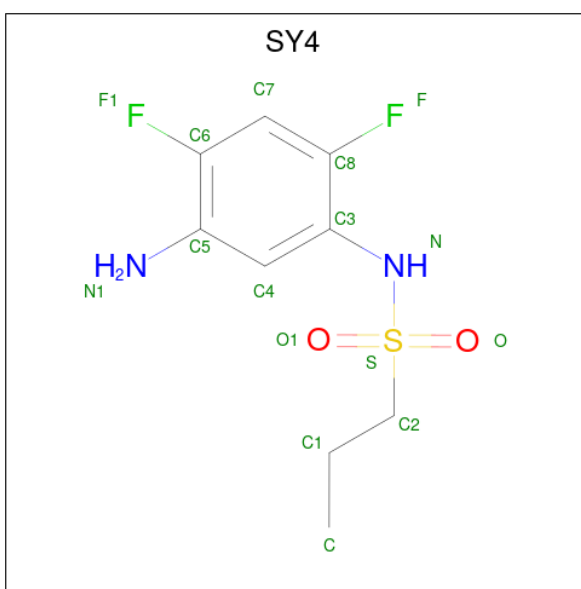
Chain	Residue	Modelled	Actual	Comment	Reference
B	-3	GLY	-	expression tag	UNP P32357
B	-2	ALA	-	expression tag	UNP P32357
B	-1	MET	-	expression tag	UNP P32357
B	0	ALA	-	expression tag	UNP P32357
B	166	SER	LEU	conflict	UNP P32357
B	167	SER	LYS	conflict	UNP P32357
B	170	SER	LEU	conflict	UNP P32357
B	?	-	GLN	deletion	UNP P32357
B	?	-	LYS	deletion	UNP P32357
B	?	-	ALA	deletion	UNP P32357
B	?	-	GLY	deletion	UNP P32357
B	?	-	SER	deletion	UNP P32357
B	?	-	LYS	deletion	UNP P32357

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	?	-	MET	deletion	UNP P32357
B	?	-	GLU	deletion	UNP P32357
B	?	-	ALA	deletion	UNP P32357
B	?	-	LYS	deletion	UNP P32357
B	?	-	ASN	deletion	UNP P32357
B	?	-	GLU	deletion	UNP P32357
B	?	-	ASP	deletion	UNP P32357

- Molecule 3 is {N}-[5-azanyl-2,4-bis(fluoranyl)phenyl]propane-1-sulfonamide (three-letter code: SY4) (formula: C<sub>9</sub>H<sub>12</sub>F<sub>2</sub>N<sub>2</sub>O<sub>2</sub>S) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	
			Total	C	F	N	O			S
3	A	1	Total	C	F	N	O	S	0	0
			16	9	2	2	2	1		
3	B	1	Total	C	F	N	O	S	0	0
			16	9	2	2	2	1		

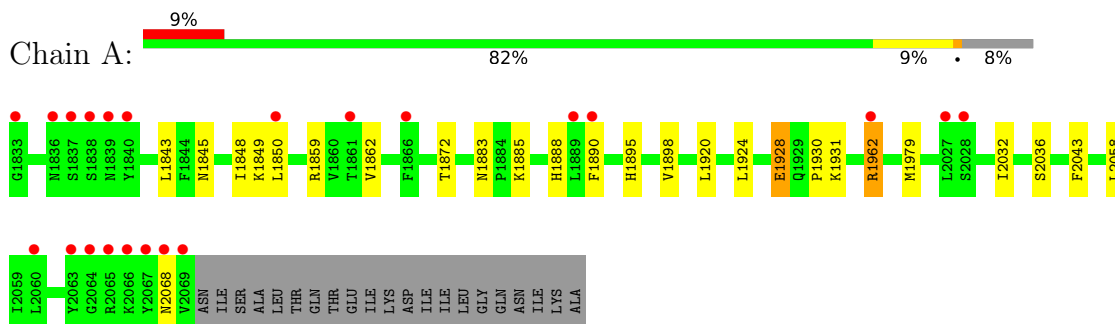
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	92	Total O 92 92	0	0
4	B	72	Total O 72 72	0	0

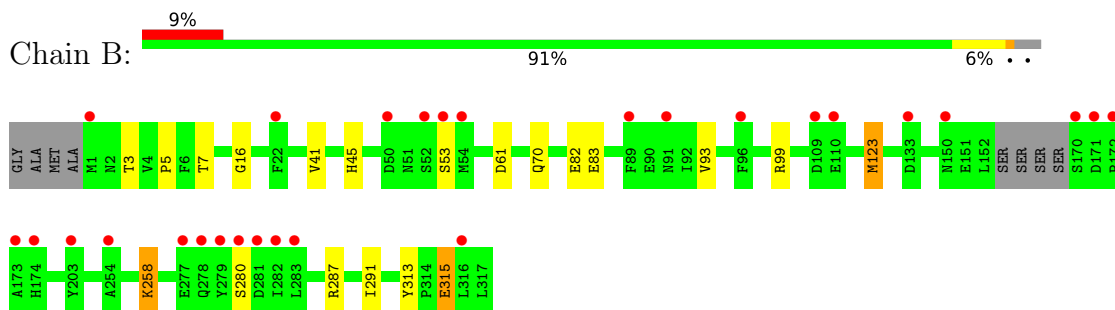
### 3 Residue-property plots [i](#)

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: Pre-mRNA-splicing factor 8



- Molecule 2: A1 cistron-splicing factor AAR2



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	88.91Å 82.05Å 94.36Å 90.00° 108.99° 90.00°	Depositor
Resolution (Å)	26.58 – 1.59 44.61 – 1.59	Depositor EDS
% Data completeness (in resolution range)	99.2 (26.58-1.59) 99.3 (44.61-1.59)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.96 (at 1.59Å)	Xtrriage
Refinement program	REFMAC 5.8.0238	Depositor
R, $R_{free}$	0.206 , 0.236 0.223 , 0.239	Depositor DCC
$R_{free}$ test set	2101 reflections (2.48%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	31.0	Xtrriage
Anisotropy	0.355	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 46.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	4778	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	41.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.65% of the height of the origin peak. No significant pseudotranslation is detected.*

<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: SY4

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	0.71	4/2049 (0.2%)	0.75	1/2775 (0.0%)
2	B	0.70	4/2651 (0.2%)	0.77	5/3581 (0.1%)
All	All	0.70	8/4700 (0.2%)	0.76	6/6356 (0.1%)

All (8) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	5	PRO	N-CA	11.33	1.66	1.47
1	A	1924	LEU	C-N	9.11	1.51	1.34
2	B	82	GLU	CD-OE1	-9.05	1.15	1.25
1	A	1883	ASN	C-N	8.80	1.50	1.34
2	B	83	GLU	CD-OE1	-6.65	1.18	1.25
1	A	1928	GLU	CD-OE2	-6.28	1.18	1.25
2	B	83	GLU	CD-OE2	-5.96	1.19	1.25
1	A	1888	HIS	CE1-NE2	-5.29	1.20	1.32

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	123	MET	CG-SD-CE	6.68	110.88	100.20
2	B	99	ARG	CB-CA-C	6.63	123.66	110.40
2	B	5	PRO	CA-N-CD	-6.61	102.24	111.50
2	B	5	PRO	CB-CA-C	6.13	127.32	112.00
2	B	70	GLN	CB-CA-C	-5.71	98.97	110.40
1	A	1859	ARG	NE-CZ-NH2	5.07	122.84	120.30

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	2002	0	2029	11	2
2	B	2580	0	2450	9	1
3	A	16	0	0	0	0
3	B	16	0	0	1	0
4	A	92	0	0	0	0
4	B	72	0	0	1	0
All	All	4778	0	4479	20	2

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 2.

All (20) 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)
2:B:3:THR:HG21	2:B:93:VAL:HG13	1.83	0.59
1:A:1848:ILE:H	1:A:1931[A]:LYS:HZ2	1.51	0.57
2:B:287:ARG:O	2:B:291:ILE:HD13	2.05	0.57
1:A:1962:ARG:H	1:A:1962:ARG:HD3	1.70	0.57
2:B:16:GLY:HA3	2:B:45:HIS:CE1	2.45	0.52
1:A:1895:HIS:O	1:A:1898[A]:VAL:HG22	2.10	0.50
1:A:2058:LEU:HD23	1:A:2058:LEU:C	2.31	0.50
1:A:1850:LEU:HD12	1:A:1930:PRO:HG3	1.95	0.48
1:A:1845:ASN:O	1:A:1885:LYS:HD3	2.14	0.48
2:B:258:LYS:HD2	2:B:258:LYS:H	1.80	0.47
2:B:315:GLU:CD	2:B:315:GLU:H	2.18	0.47
1:A:1862:VAL:HG22	1:A:1872:THR:HG22	1.98	0.45
2:B:61:ASP:HB2	2:B:123:MET:HG2	1.98	0.44
1:A:1890:PHE:CD2	1:A:1920:LEU:HD11	2.53	0.44
1:A:1962:ARG:H	1:A:1962:ARG:CD	2.31	0.43
2:B:280:SER:HB3	2:B:313:TYR:CE1	2.55	0.42
2:B:41[A]:VAL:HG12	4:B:555:HOH:O	2.20	0.41
1:A:2032:ILE:HD13	1:A:2043:PHE:CD1	2.56	0.41
1:A:1843:LEU:HA	1:A:1849:LYS:HD2	2.02	0.41
2:B:7:THR:OG1	3:B:401:SY4:N	2.54	0.41



All (2) 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)
1:A:1928:GLU:OE1	1:A:2036:SER:OG[2_656]	2.07	0.13
1:A:2068:ASN:O	2:B:53:SER:OG[2_655]	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	248/258 (96%)	243 (98%)	5 (2%)	0	100	100
2	B	306/308 (99%)	297 (97%)	9 (3%)	0	100	100
All	All	554/566 (98%)	540 (98%)	14 (2%)	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	226/233 (97%)	222 (98%)	4 (2%)	59	36
2	B	287/284 (101%)	285 (99%)	2 (1%)	84	73
All	All	513/517 (99%)	507 (99%)	6 (1%)	81	54

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1962	ARG
1	A	1979[A]	MET
1	A	1979[B]	MET
1	A	1979[C]	MET
2	B	258	LYS
2	B	315	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](#)

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

## 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
3	SY4	B	401	-	15,16,16	1.54	3 (20%)	18,23,23	1.12	2 (11%)
3	SY4	A	2501	-	15,16,16	1.24	2 (13%)	18,23,23	1.13	1 (5%)

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	SY4	B	401	-	-	3/9/9/9	0/1/1/1
3	SY4	A	2501	-	-	2/9/9/9	0/1/1/1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	401	SY4	O1-S	-3.10	1.39	1.43
3	A	2501	SY4	O1-S	-2.55	1.40	1.43
3	A	2501	SY4	C1-C2	-2.38	1.38	1.51
3	B	401	SY4	C4-C5	-2.33	1.36	1.40
3	B	401	SY4	S-N	-2.20	1.58	1.62

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	2501	SY4	C-C1-C2	-4.37	99.60	112.45
3	B	401	SY4	O-S-N	2.17	112.57	107.28
3	B	401	SY4	C4-C3-N	-2.04	116.18	120.70

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	2501	SY4	C3-N-S-C2
3	B	401	SY4	C1-C2-S-O
3	B	401	SY4	C1-C2-S-O1
3	A	2501	SY4	C3-N-S-O
3	B	401	SY4	C1-C2-S-N

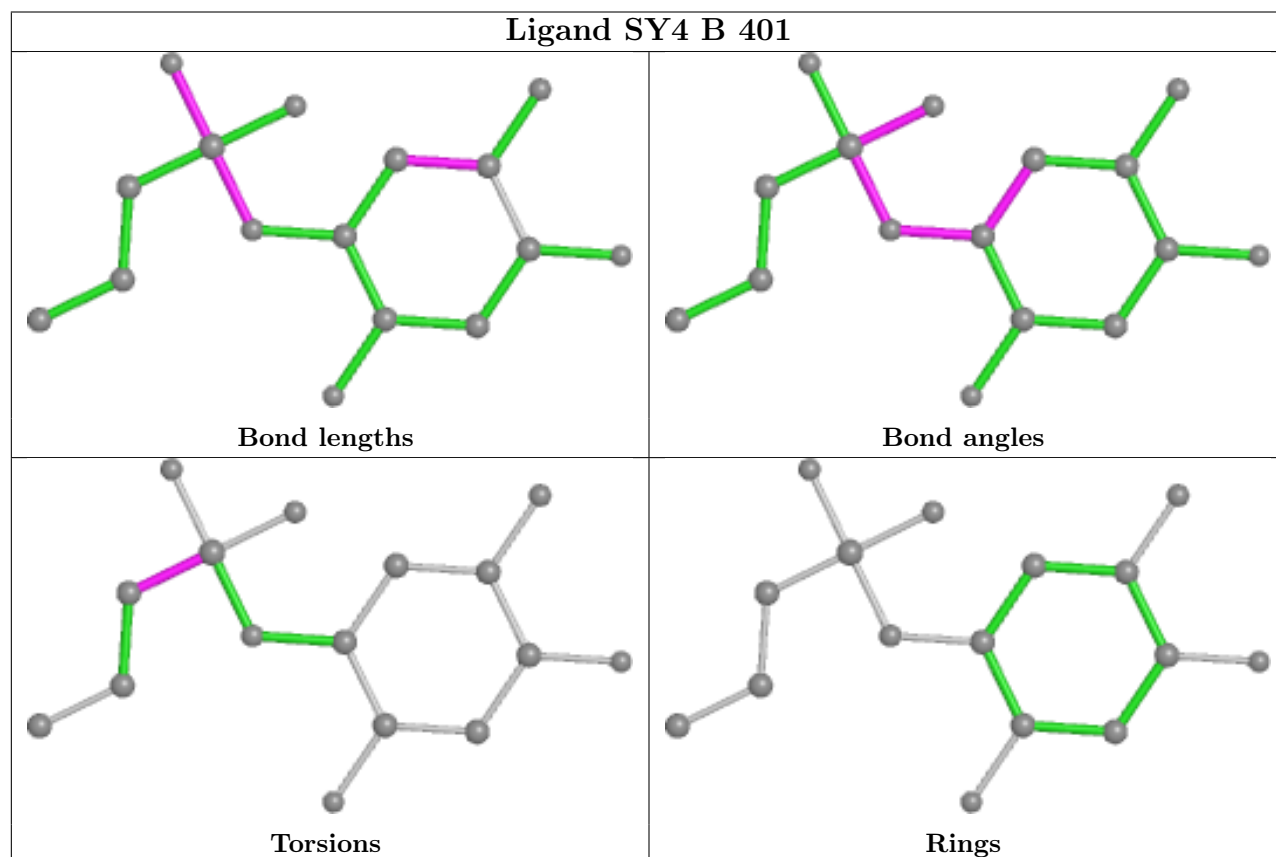
There are no ring outliers.

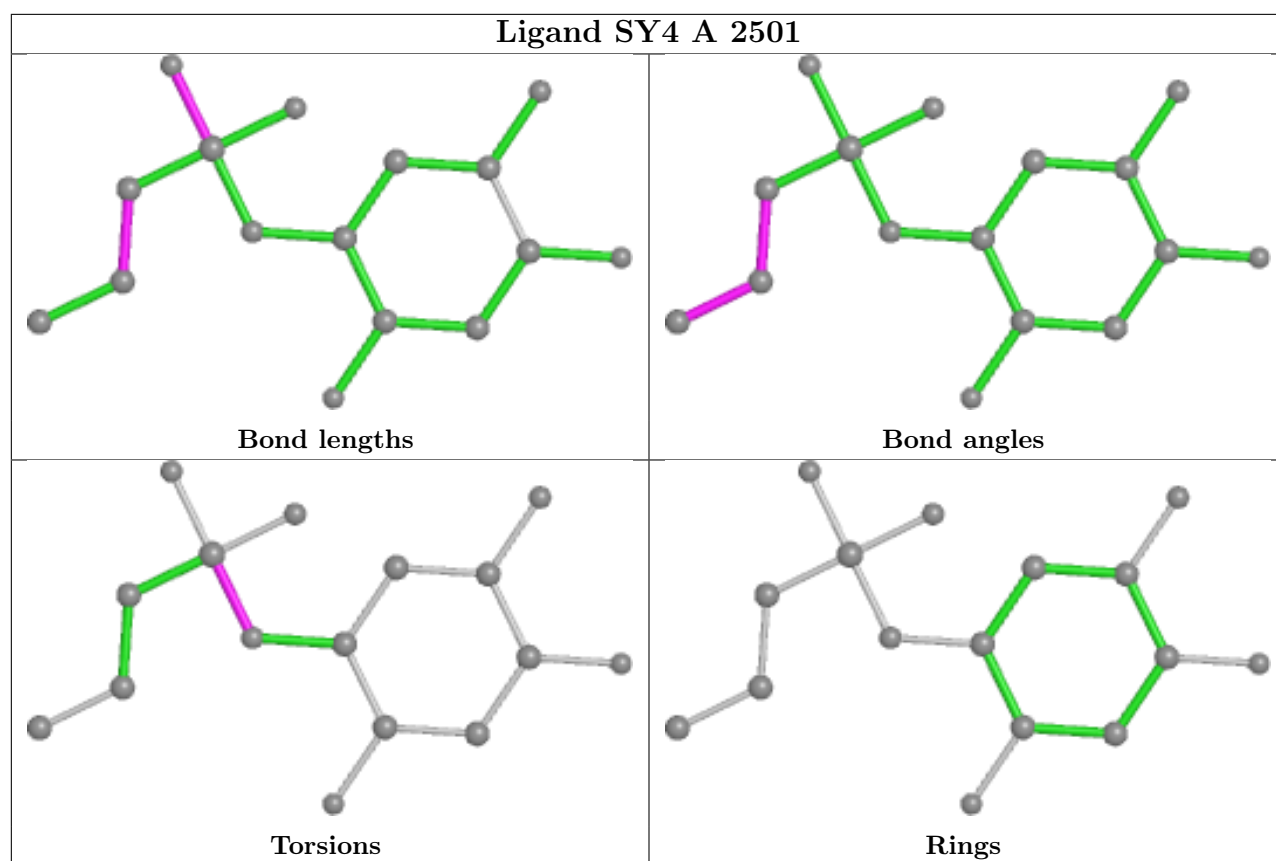
1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	401	SY4	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 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	237/258 (91%)	0.80	22 (9%) <b>8</b> <b>7</b>	24, 34, 67, 108	0
2	B	300/308 (97%)	0.75	28 (9%) <b>8</b> <b>7</b>	27, 39, 76, 103	0
All	All	537/566 (94%)	0.77	50 (9%) <b>8</b> <b>7</b>	24, 37, 73, 108	0

All (50) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	1	MET	7.5
1	A	1833	GLY	7.3
1	A	2063	TYR	7.1
2	B	279	TYR	6.6
1	A	2065	ARG	5.9
1	A	2064	GLY	5.1
1	A	1838	SER	5.0
1	A	2060	LEU	4.9
2	B	278	GLN	4.8
1	A	2069	VAL	4.8
2	B	280	SER	4.6
1	A	2067	TYR	4.6
1	A	2066	LYS	4.5
2	B	53	SER	4.5
1	A	2068	ASN	4.4
2	B	281	ASP	4.3
1	A	2027	LEU	4.0
2	B	172	PRO	4.0
2	B	109	ASP	4.0
2	B	52	SER	3.9
2	B	110	GLU	3.7
2	B	170	SER	3.4
1	A	1866	PHE	3.4
2	B	22	PHE	3.4

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	1837	SER	3.2
2	B	54[A]	MET	3.0
2	B	96	PHE	3.0
2	B	174	HIS	3.0
1	A	1890	PHE	3.0
2	B	171	ASP	2.8
2	B	316	LEU	2.7
1	A	2028	SER	2.7
2	B	173	ALA	2.7
2	B	150	ASN	2.7
1	A	1850	LEU	2.7
2	B	91	ASN	2.6
2	B	254	ALA	2.6
2	B	277	GLU	2.6
1	A	1962	ARG	2.6
1	A	1839	ASN	2.6
2	B	89	PHE	2.5
1	A	1836	ASN	2.5
2	B	283	LEU	2.4
1	A	1840	TYR	2.4
2	B	203[A]	TYR	2.2
2	B	133	ASP	2.2
2	B	50	ASP	2.1
1	A	1889	LEU	2.1
1	A	1861	THR	2.1
2	B	282	ILE	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 [\(i\)](#)

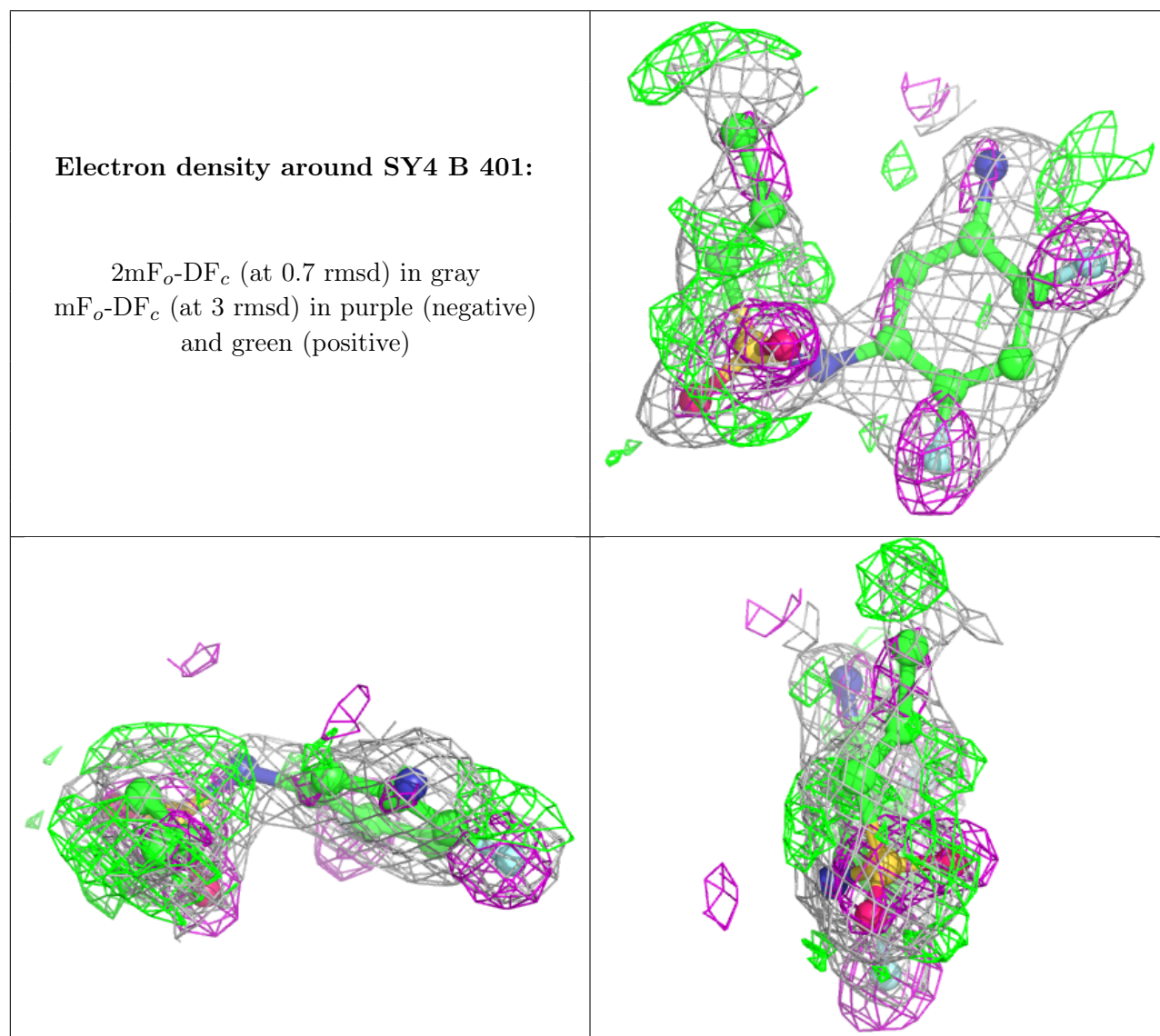
There are no monosaccharides in this entry.

## 6.4 Ligands [\(i\)](#)

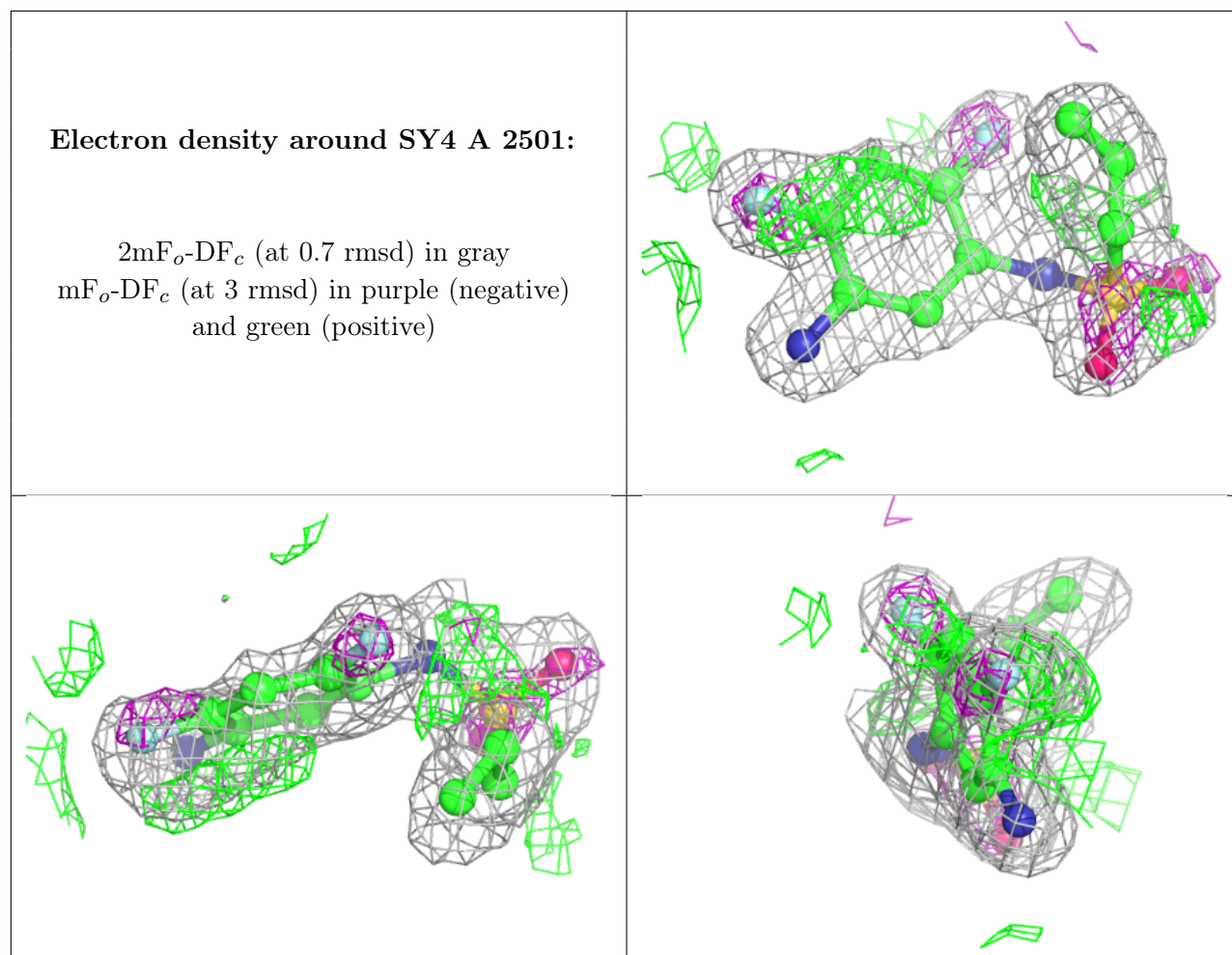
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 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
3	SY4	B	401	16/16	0.83	0.24	20,20,20,20	16
3	SY4	A	2501	16/16	0.89	0.14	20,20,20,20	16

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.