



wwPDB X-ray Structure Validation Summary Report ⓘ

May 25, 2020 – 10:11 am BST

PDB ID : 5YBB
Title : Structural basis underlying complex assembly and conformational transition of the type I R-M system
Authors : Liu, Y.P.; Tang, Q.; Zhang, J.Z.; Tian, L.F.; Gao, P.; Yan, X.X.
Deposited on : 2017-09-04
Resolution : 3.20 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 ⓘ 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

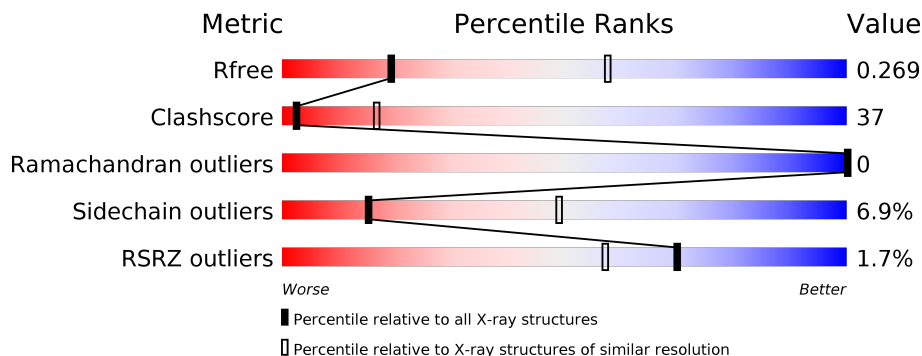
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.20 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	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)

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 ≥ 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	507	
1	B	507	
1	C	507	
1	E	507	
2	D	398	
2	G	398	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	H	22	
4	I	22	

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
5	SAM	A	601	-	-	X	-

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 15456 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 Type I restriction-modification system methyltransferase subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	487	3934	2513	686	719	16	0	0	0
1	C	30	229	145	35	48	1	0	0	0
1	B	487	3934	2513	686	719	16	0	0	0
1	E	30	229	145	35	48	1	0	0	0

- Molecule 2 is a protein called Restriction endonuclease S subunits.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	D	390	3090	1971	549	558	12	0	0	0
2	G	390	3090	1971	549	558	12	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	1	VAL	-	expression tag	UNP Q8R9Q6
G	1	VAL	-	expression tag	UNP Q8R9Q6

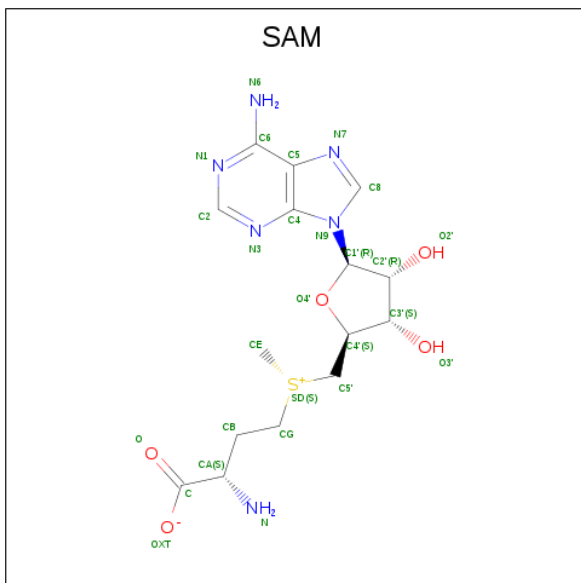
- Molecule 3 is a DNA chain called DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	H	22	454	215	88	130	21	0	0	0

- Molecule 4 is a DNA chain called DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	I	22	442	211	80	130	21	0	0	0

- Molecule 5 is S-ADENOSYLMETHIONINE (three-letter code: SAM) (formula: $C_{15}H_{22}N_6O_5S$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
5	A	1	27	15	6	5	1	0	0
5	B	1	27	15	6	5	1	0	0

4 Data and refinement statistics

Property	Value	Source
Space group	P 41	Depositor
Cell constants a, b, c, α , β , γ	121.60Å 121.60Å 280.36Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.00 – 3.20 47.00 – 3.20	Depositor EDS
% Data completeness (in resolution range)	95.3 (47.00-3.20) 95.4 (47.00-3.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.90 (at 3.19Å)	Xtrriage
Refinement program	PHENIX 1.8.4_1496	Depositor
R, R_{free}	0.236 , 0.272 0.242 , 0.269	Depositor DCC
R_{free} test set	3039 reflections (4.77%)	wwPDB-VP
Wilson B-factor (Å ²)	96.9	Xtrriage
Anisotropy	0.075	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 61.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.427 for h,-k,-l	Xtrriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	15456	wwPDB-VP
Average B, all atoms (Å ²)	104.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.66% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: SAM

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	0.57	3/4031 (0.1%)	0.81	12/5466 (0.2%)
1	B	0.53	4/4031 (0.1%)	0.72	6/5466 (0.1%)
1	C	0.63	0/229	0.89	2/308 (0.6%)
1	E	0.75	0/229	0.92	1/308 (0.3%)
2	D	0.61	2/3165 (0.1%)	0.89	19/4307 (0.4%)
2	G	0.47	2/3165 (0.1%)	0.74	7/4307 (0.2%)
3	H	0.55	0/510	0.81	0/787
4	I	0.59	0/494	0.88	0/759
All	All	0.55	11/15854 (0.1%)	0.80	47/21708 (0.2%)

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
1	A	0	2
1	B	0	2
1	E	0	1
2	D	0	4
2	G	0	3
All	All	0	12

The worst 5 of 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	158	PRO	N-CD	-19.13	1.21	1.47
1	A	106	PRO	N-CD	7.70	1.58	1.47
1	A	362	PRO	N-CD	5.38	1.55	1.47
1	B	464	PRO	N-CD	5.33	1.55	1.47
1	B	362	PRO	N-CD	5.31	1.55	1.47

The worst 5 of 47 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
2	D	141	SER	N-CA-CB	-18.16	83.27	110.50
2	D	158	PRO	N-CA-C	13.89	148.22	112.10
1	A	106	PRO	N-CA-C	13.13	146.25	112.10
1	A	107	LEU	N-CA-C	-12.56	77.09	111.00
2	D	158	PRO	CA-N-CD	12.09	128.63	111.70

There are no chirality outliers.

5 of 12 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	107	LEU	Peptide
1	A	375	ARG	Peptide
2	D	158	PRO	Peptide
2	D	159	PRO	Peptide
2	D	160	LEU	Peptide

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	3934	0	3858	358	0
1	B	3934	0	3857	305	1
1	C	229	0	233	27	0
1	E	229	0	233	24	0
2	D	3090	0	3103	263	1
2	G	3090	0	3106	159	2
3	H	454	0	248	5	1
4	I	442	0	248	4	1
5	A	27	0	22	9	0
5	B	27	0	22	8	0
All	All	15456	0	14930	1117	3

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 37.

The worst 5 of 1117 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:50:GLU:CG	1:B:67:ILE:HD11	1.19	1.64
2:D:37:PHE:CD2	2:D:99:ILE:HD11	1.30	1.63
1:B:50:GLU:HG3	1:B:67:ILE:CD1	1.10	1.54
1:A:68:ILE:HG21	1:A:98:TYR:CE1	1.47	1.49
2:D:37:PHE:CE2	2:D:99:ILE:HD11	1.51	1.45

All (3) 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)
2:D:271:ARG:NH1	1:B:85:ASP:OD2[2_764]	1.87	0.33
2:G:339:LYS:NZ	3:H:9:DT:OP1[3_645]	2.06	0.14
2:G:268:LYS:NZ	4:I:16:DT:OP2[3_645]	2.19	0.01

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	483/507 (95%)	460 (95%)	23 (5%)	0	100	100
1	B	483/507 (95%)	465 (96%)	18 (4%)	0	100	100
1	C	28/507 (6%)	26 (93%)	2 (7%)	0	100	100
1	E	28/507 (6%)	27 (96%)	1 (4%)	0	100	100
2	D	386/398 (97%)	368 (95%)	18 (5%)	0	100	100
2	G	386/398 (97%)	372 (96%)	14 (4%)	0	100	100
All	All	1794/2824 (64%)	1718 (96%)	76 (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	420/445 (94%)	390 (93%)	30 (7%)	14	47
1	B	420/445 (94%)	386 (92%)	34 (8%)	11	42
1	C	26/445 (6%)	23 (88%)	3 (12%)	5	24
1	E	26/445 (6%)	23 (88%)	3 (12%)	5	24
2	D	333/340 (98%)	309 (93%)	24 (7%)	14	47
2	G	333/340 (98%)	319 (96%)	14 (4%)	30	65
All	All	1558/2460 (63%)	1450 (93%)	108 (7%)	15	49

5 of 108 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	179	GLU
1	B	9	SER
2	G	164	ARG
2	D	209	ASP
2	D	305	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	5	GLN
1	B	56	GLN
2	G	121	HIS
1	A	456	ASN
1	B	456	ASN

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
5	SAM	A	601	-	21,29,29	1.11	2 (9%)	18,42,42	1.61	5 (27%)
5	SAM	B	601	-	21,29,29	1.15	2 (9%)	18,42,42	1.59	5 (27%)

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
5	SAM	A	601	-	-	3/8/33/33	0/3/3/3
5	SAM	B	601	-	-	3/8/33/33	0/3/3/3

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	B	601	SAM	C2'-C1'	-2.27	1.50	1.53
5	B	601	SAM	CE-SD	-2.11	1.65	1.78
5	A	601	SAM	C2'-C1'	-2.08	1.50	1.53
5	A	601	SAM	CE-SD	-2.02	1.66	1.78

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

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
5	B	601	SAM	N3-C2-N1	-3.29	123.53	128.68
5	A	601	SAM	N3-C2-N1	-3.17	123.72	128.68
5	A	601	SAM	C4-C5-N7	-2.77	106.52	109.40
5	B	601	SAM	C5'-SD-CG	2.56	109.94	103.40
5	B	601	SAM	C4-C5-N7	-2.51	106.78	109.40

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

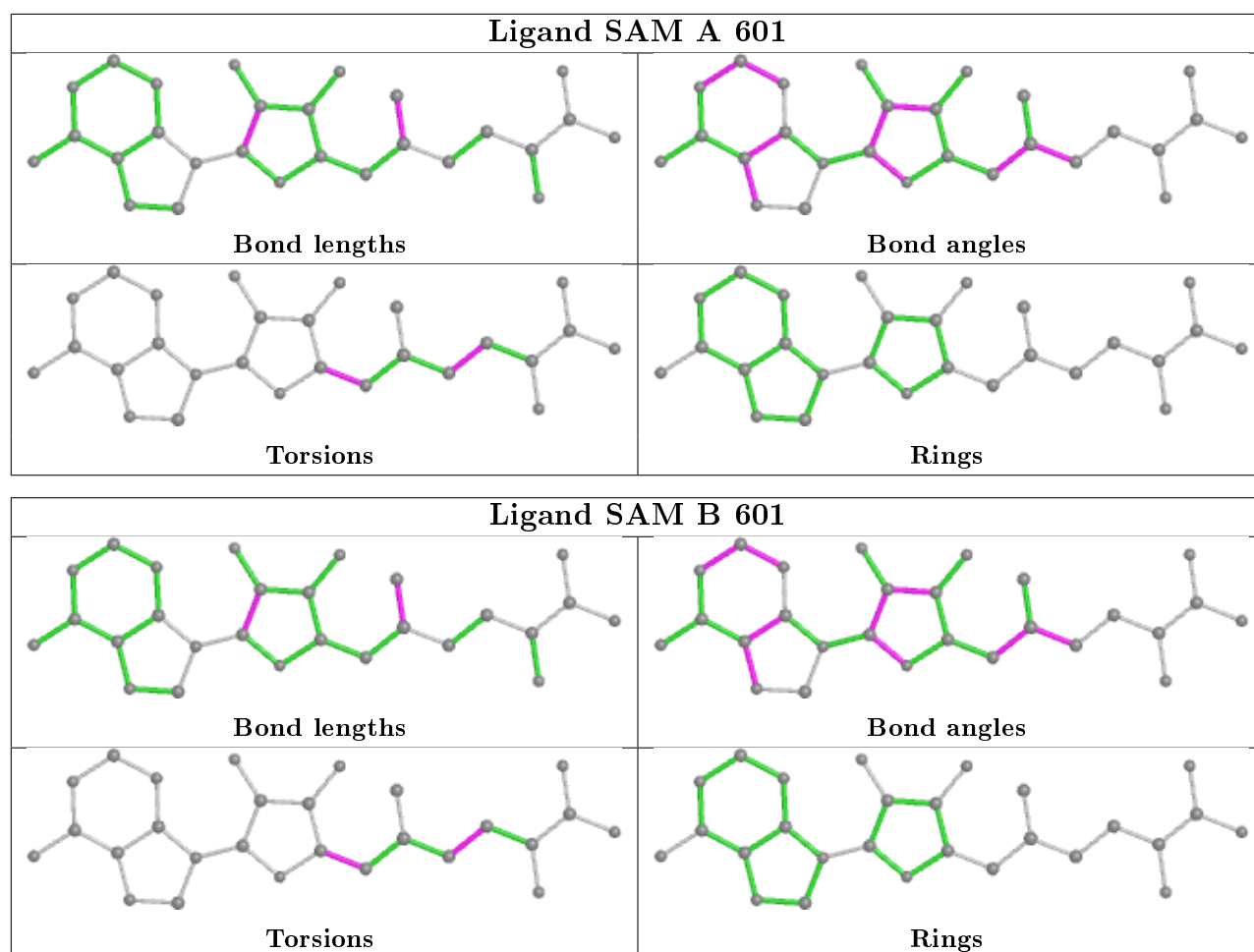
Mol	Chain	Res	Type	Atoms
5	A	601	SAM	O4'-C4'-C5'-SD
5	A	601	SAM	C3'-C4'-C5'-SD
5	B	601	SAM	CA-CB-CG-SD
5	B	601	SAM	O4'-C4'-C5'-SD
5	A	601	SAM	CA-CB-CG-SD

There are no ring outliers.

2 monomers are involved in 17 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	601	SAM	9	0
5	B	601	SAM	8	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 [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	487/507 (96%)	0.16	3 (0%) 89 83	42, 94, 167, 229	0
1	B	487/507 (96%)	0.17	6 (1%) 79 67	45, 97, 167, 238	0
1	C	30/507 (5%)	0.09	1 (3%) 46 30	87, 117, 154, 217	0
1	E	30/507 (5%)	0.28	3 (10%) 7 4	90, 117, 169, 233	0
2	D	390/398 (97%)	0.20	5 (1%) 77 65	63, 94, 164, 235	0
2	G	390/398 (97%)	0.27	12 (3%) 49 32	59, 95, 154, 218	0
3	H	22/22 (100%)	-0.09	0 100 100	84, 100, 183, 220	0
4	I	22/22 (100%)	-0.15	1 (4%) 33 21	78, 92, 170, 237	0
All	All	1858/2868 (64%)	0.19	31 (1%) 70 57	42, 96, 167, 238	0

The worst 5 of 31 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	496	GLY	8.1
2	G	160	LEU	5.7
2	G	138	ARG	3.7
1	A	496	GLY	3.7
2	G	135	SER	3.7

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](#)

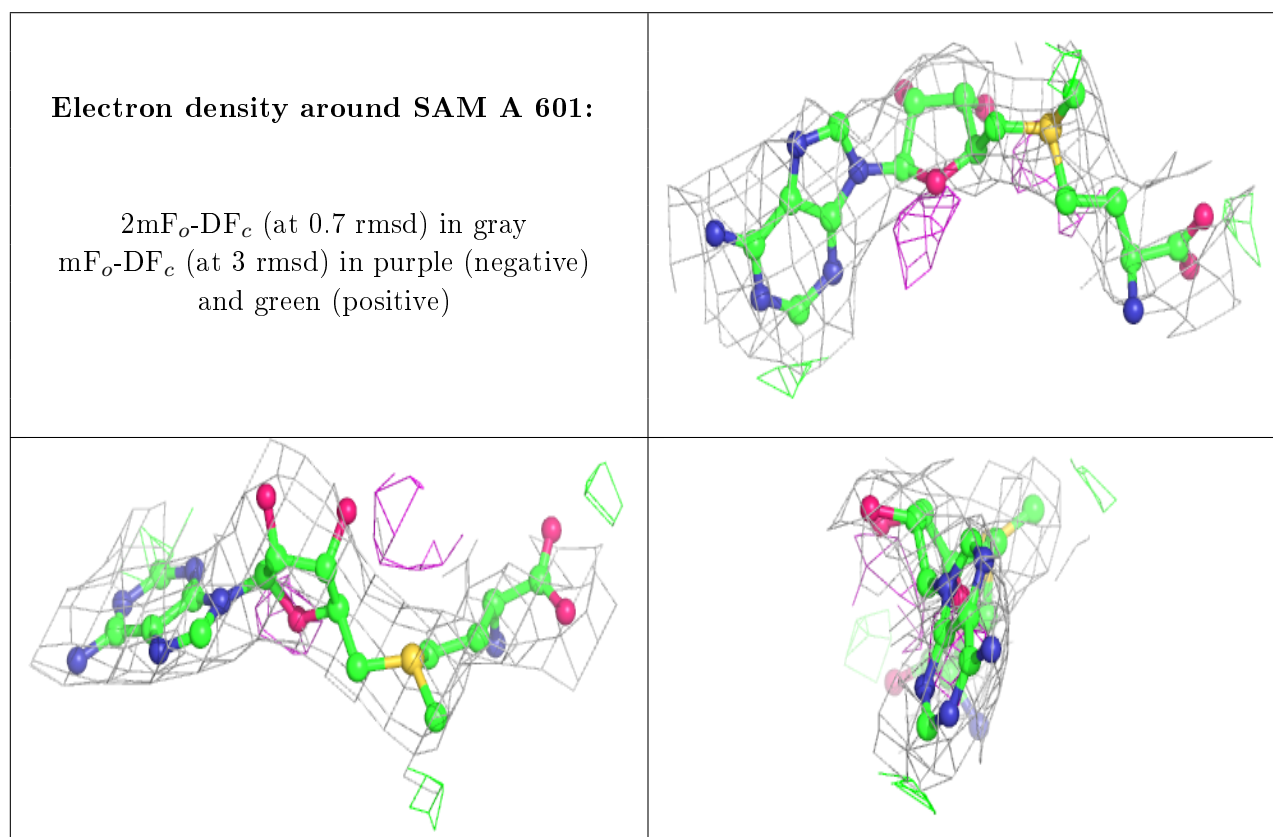
There are no carbohydrates in this entry.

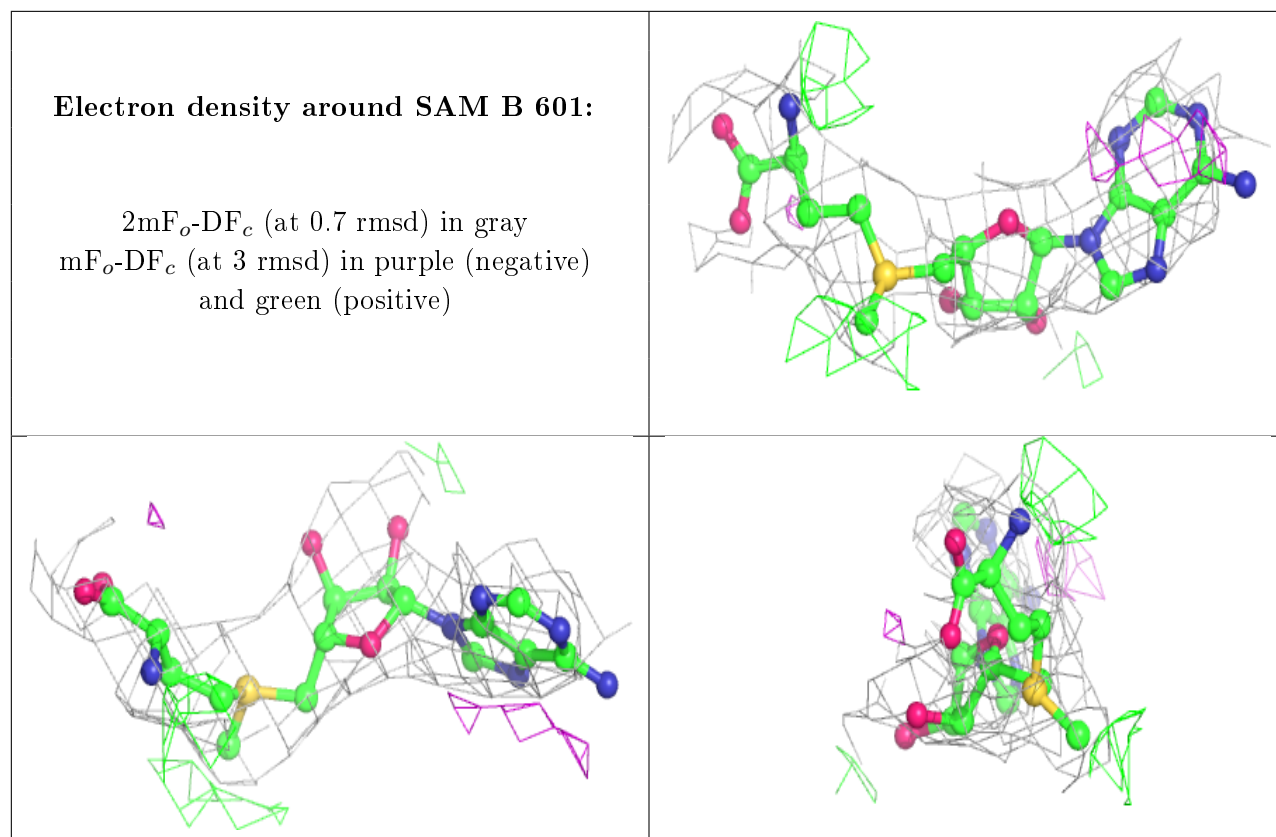
6.4 Ligands [i](#)

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, 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
5	SAM	A	601	27/27	0.78	0.39	106,142,155,271	0
5	SAM	B	601	27/27	0.82	0.33	93,137,149,189	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.