



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 21, 2016 – 01:59 PM EST

PDB ID : 5HFJ
Title : crystal structure of M1.HpyAVI-SAM complex
Authors : Ma, B.; Liu, W.; Zhang, H.
Deposited on : 2016-01-07
Resolution : 3.10 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

<http://wwpdb.org/validation/2016/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	:	1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20028320
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20028320

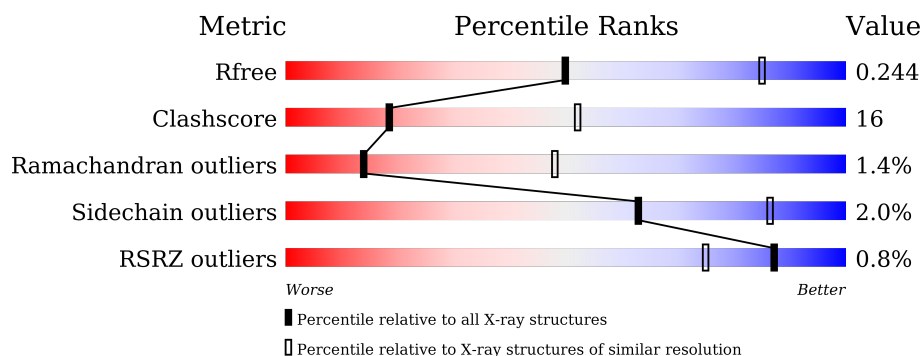
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.10 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}	91344	1114 (3.14-3.06)
Clashscore	102246	1222 (3.14-3.06)
Ramachandran outliers	100387	1174 (3.14-3.06)
Sidechain outliers	100360	1174 (3.14-3.06)
RSRZ outliers	91569	1119 (3.14-3.06)

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. 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	232	<div> <div>%</div> <div> <div></div> <div>59%</div> <div>23%</div> <div>• •</div> <div>13%</div> </div> </div>
1	B	232	<div> <div>64%</div> <div>18%</div> <div>• •</div> <div>16%</div> </div>
1	C	232	<div> <div>63%</div> <div>23%</div> <div>•</div> <div>12%</div> </div>
1	D	232	<div> <div>53%</div> <div>25%</div> <div>•</div> <div>19%</div> </div>
1	E	232	<div> <div>%</div> <div>52%</div> <div>26%</div> <div>• •</div> <div>17%</div> </div>
1	F	232	<div> <div>56%</div> <div>25%</div> <div>•</div> <div>16%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	232	<div><div><div>%</div><div><div></div><div></div><div></div></div><div>58%25%14%</div></div></div>
1	H	232	<div><div><div>%</div><div><div></div><div></div><div></div></div><div>55%26%17%</div></div></div>

2 Entry composition

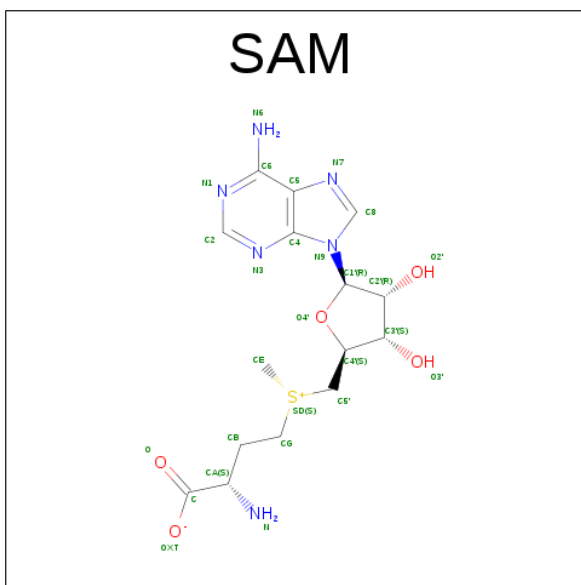
There are 2 unique types of molecules in this entry. The entry contains 13194 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Adenine specific DNA methyltransferase (DpnA).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	201	Total	C	N	O	S	0	0	0
			1661	1087	278	288	8			
1	B	194	Total	C	N	O	S	0	0	0
			1611	1054	272	277	8			
1	C	204	Total	C	N	O	S	0	0	0
			1684	1098	284	294	8			
1	D	188	Total	C	N	O	S	0	0	0
			1558	1021	260	269	8			
1	E	193	Total	C	N	O	S	0	0	0
			1600	1045	270	277	8			
1	F	194	Total	C	N	O	S	0	0	0
			1605	1050	269	278	8			
1	G	200	Total	C	N	O	S	0	0	0
			1662	1086	279	289	8			
1	H	193	Total	C	N	O	S	0	0	0
			1597	1044	268	277	8			

- Molecule 2 is S-ADENOSYLMETHIONINE (three-letter code: SAM) (formula: C₁₅H₂₂N₆O₅S).

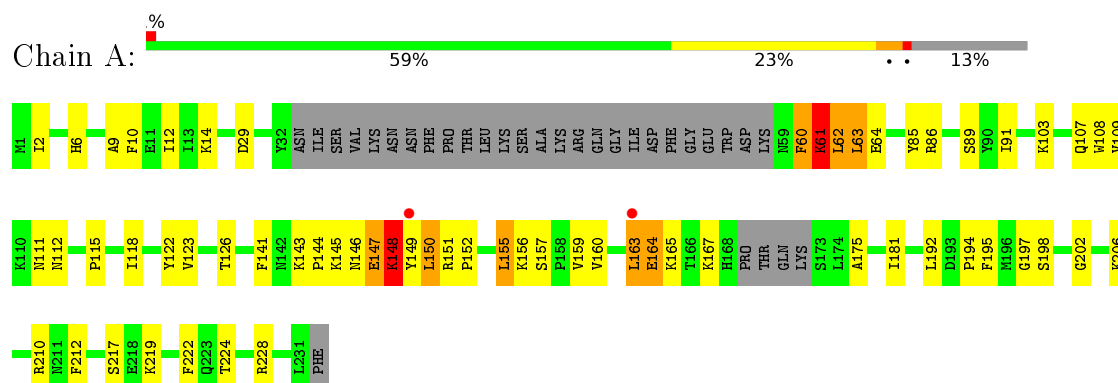


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total 27	C 15	N 6	O 5	S 1	0	0
2	B	1	Total 27	C 15	N 6	O 5	S 1	0	0
2	C	1	Total 27	C 15	N 6	O 5	S 1	0	0
2	D	1	Total 27	C 15	N 6	O 5	S 1	0	0
2	E	1	Total 27	C 15	N 6	O 5	S 1	0	0
2	F	1	Total 27	C 15	N 6	O 5	S 1	0	0
2	G	1	Total 27	C 15	N 6	O 5	S 1	0	0
2	H	1	Total 27	C 15	N 6	O 5	S 1	0	0

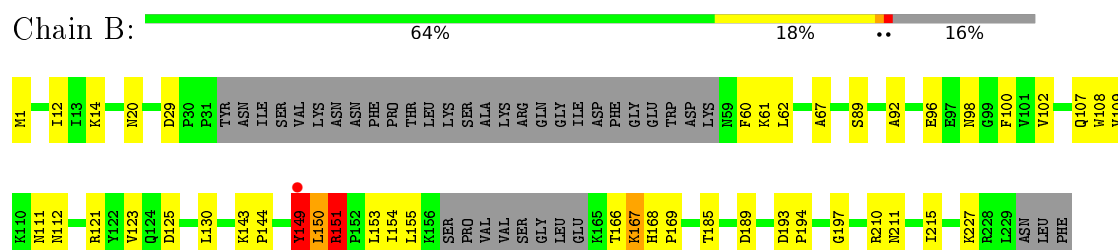
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

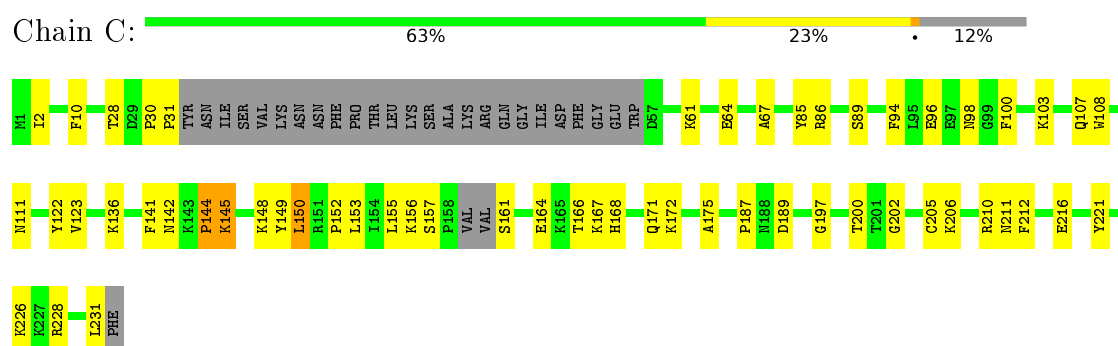
- Molecule 1: Adenine specific DNA methyltransferase (DpnA)



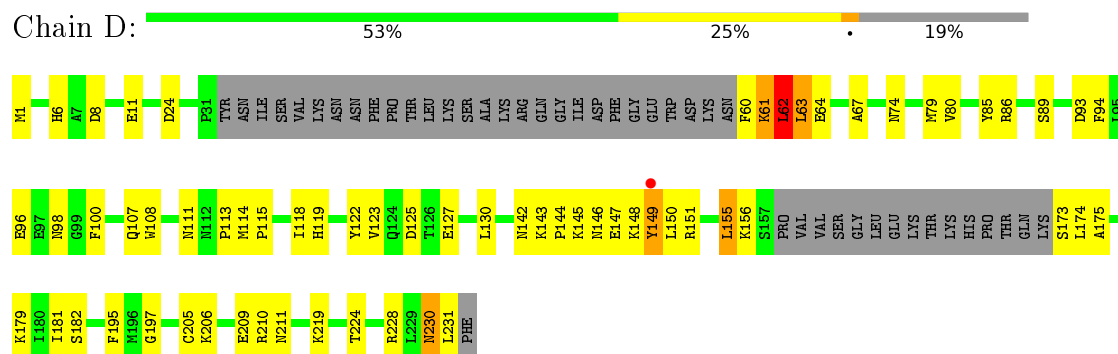
- Molecule 1: Adenine specific DNA methyltransferase (DpnA)



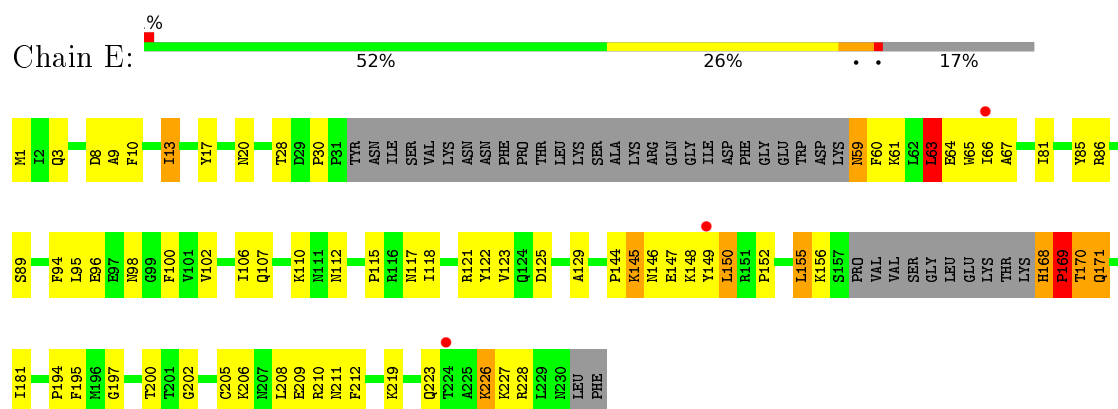
- Molecule 1: Adenine specific DNA methyltransferase (DpnA)



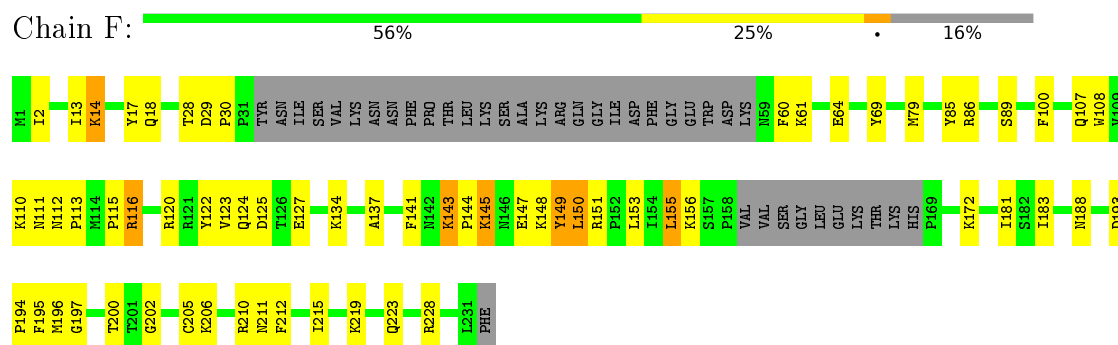
- Molecule 1: Adenine specific DNA methyltransferase (DpnA)



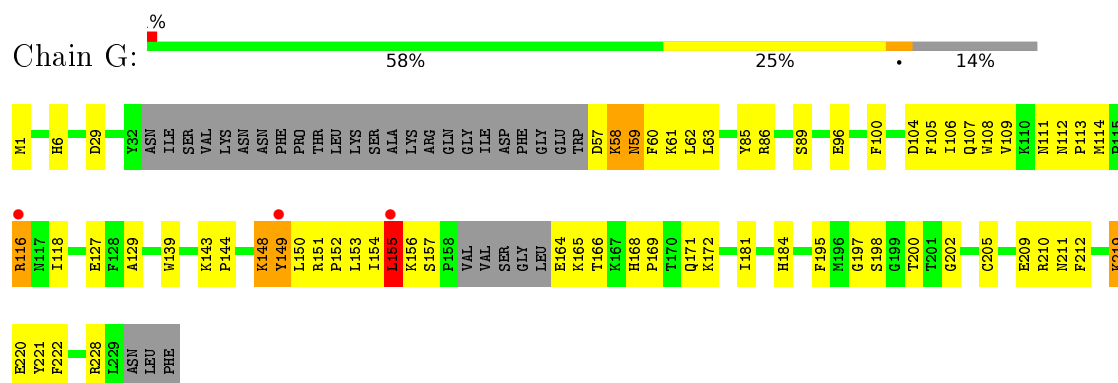
• Molecule 1: Adenine specific DNA methyltransferase (DpnA)



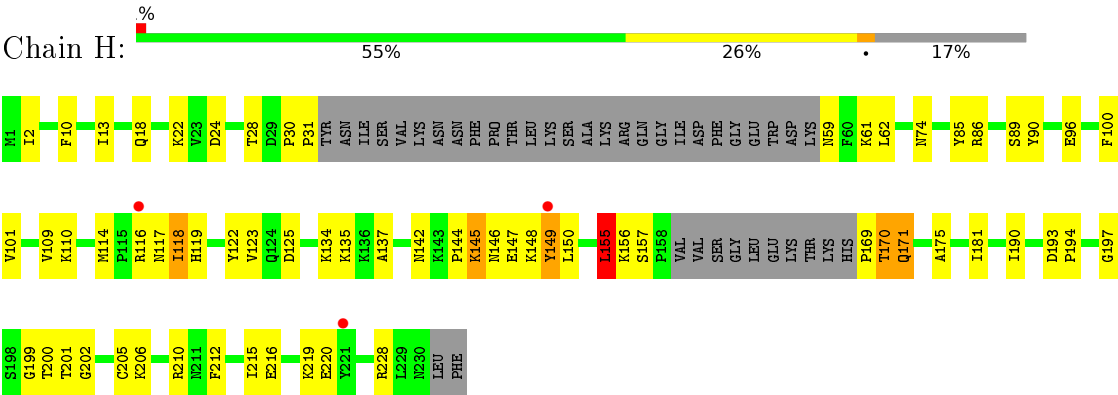
• Molecule 1: Adenine specific DNA methyltransferase (DpnA)



• Molecule 1: Adenine specific DNA methyltransferase (DpnA)



● Molecule 1: Adenine specific DNA methyltransferase (DpnA)



4 Data and refinement statistics

Property	Value	Source
Space group	P 65	Depositor
Cell constants a, b, c, α , β , γ	135.60Å 135.60Å 265.15Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	48.91 – 3.10 48.91 – 3.10	Depositor EDS
% Data completeness (in resolution range)	99.9 (48.91-3.10) 99.9 (48.91-3.10)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	0.11	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.53 (at 3.12Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, R_{free}	0.221 , 0.276 0.206 , 0.244	Depositor DCC
R_{free} test set	1998 reflections (4.01%)	DCC
Wilson B-factor (Å ²)	66.4	Xtriage
Anisotropy	0.703	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 40.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.43$, $\langle L^2 \rangle = 0.25$	Xtriage
Estimated twinning fraction	0.129 for h,-h-k,-l	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	13194	wwPDB-VP
Average B, all atoms (Å ²)	73.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.03% 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

5.1 Standard geometry

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	0.35	0/1703	0.75	7/2300 (0.3%)
1	B	0.32	0/1652	0.69	1/2229 (0.0%)
1	C	0.31	0/1726	0.71	0/2329
1	D	0.33	0/1597	0.80	5/2156 (0.2%)
1	E	0.38	0/1641	0.75	6/2216 (0.3%)
1	F	0.35	0/1646	0.72	1/2223 (0.0%)
1	G	0.37	0/1705	0.82	5/2301 (0.2%)
1	H	0.34	0/1638	0.71	2/2212 (0.1%)
All	All	0.34	0/13308	0.74	27/17966 (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	4
1	C	0	1
1	D	0	3
1	F	0	2
1	H	0	2
All	All	0	12

There are no bond length outliers.

The worst 5 of 27 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	62	LEU	CA-CB-CG	11.19	141.02	115.30
1	A	149	TYR	C-N-CA	9.09	144.44	121.70
1	E	168	HIS	C-N-CD	-8.39	102.15	120.60

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	G	155	LEU	CA-CB-CG	8.30	134.38	115.30
1	H	155	LEU	CA-CB-CG	7.54	132.65	115.30

There are no chirality outliers.

5 of 12 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	148	LYS	Peptide
1	A	163	LEU	Peptide
1	A	60	PHE	Peptide
1	A	61	LYS	Peptide
1	C	167	LYS	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	1661	0	1687	64	0
1	B	1611	0	1641	38	1
1	C	1684	0	1712	49	0
1	D	1558	0	1582	53	0
1	E	1600	0	1619	71	1
1	F	1605	0	1631	61	0
1	G	1662	0	1685	59	0
1	H	1597	0	1620	60	0
2	A	27	0	22	7	0
2	B	27	0	22	2	0
2	C	27	0	22	6	0
2	D	27	0	22	2	0
2	E	27	0	22	8	0
2	F	27	0	22	4	0
2	G	27	0	22	6	0
2	H	27	0	22	6	0
All	All	13194	0	13353	412	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 412 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:155:LEU:HD23	1:A:157:SER:H	1.25	1.02
1:G:151:ARG:HH22	1:G:154:ILE:HG22	1.34	0.91
1:G:153:LEU:HG	1:G:155:LEU:H	1.38	0.88
1:F:13:ILE:HG13	1:F:17:TYR:HE2	1.42	0.84
1:B:108:TRP:HE1	1:B:155:LEU:HB3	1.43	0.82

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)
1:B:20:ASN:OD1	1:E:20:ASN:ND2[6_564]	2.16	0.04

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	195/232 (84%)	178 (91%)	12 (6%)	5 (3%)	7	32
1	B	188/232 (81%)	176 (94%)	10 (5%)	2 (1%)	17	55
1	C	198/232 (85%)	181 (91%)	15 (8%)	2 (1%)	19	58
1	D	182/232 (78%)	168 (92%)	12 (7%)	2 (1%)	17	55
1	E	187/232 (81%)	171 (91%)	12 (6%)	4 (2%)	9	37
1	F	188/232 (81%)	171 (91%)	16 (8%)	1 (0%)	34	72
1	G	194/232 (84%)	175 (90%)	16 (8%)	3 (2%)	13	46
1	H	187/232 (81%)	169 (90%)	15 (8%)	3 (2%)	12	44
All	All	1519/1856 (82%)	1389 (91%)	108 (7%)	22 (1%)	14	48

5 of 22 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	H	118	ILE
1	E	171	GLN
1	G	155	LEU
1	A	148	LYS
1	C	150	LEU

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	182/210 (87%)	179 (98%)	3 (2%)	70	89
1	B	176/210 (84%)	173 (98%)	3 (2%)	68	89
1	C	185/210 (88%)	184 (100%)	1 (0%)	92	96
1	D	170/210 (81%)	168 (99%)	2 (1%)	78	92
1	E	175/210 (83%)	168 (96%)	7 (4%)	38	75
1	F	176/210 (84%)	172 (98%)	4 (2%)	58	84
1	G	182/210 (87%)	176 (97%)	6 (3%)	45	79
1	H	175/210 (83%)	172 (98%)	3 (2%)	68	89
All	All	1421/1680 (85%)	1392 (98%)	29 (2%)	63	86

5 of 29 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	145	LYS
1	F	14	LYS
1	H	18	GLN
1	E	170	THR
1	F	18	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	112	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	E	168	HIS
1	G	168	HIS
1	E	59	ASN
1	F	111	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](#)

8 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	$\# Z > 2$
2	SAM	A	301	-	23,29,29	1.03	2 (8%)	15,42,42	3.27	2 (13%)
2	SAM	B	301	-	23,29,29	1.04	2 (8%)	15,42,42	3.05	1 (6%)
2	SAM	C	301	-	23,29,29	1.10	2 (8%)	15,42,42	3.37	4 (26%)
2	SAM	D	301	-	23,29,29	1.06	2 (8%)	15,42,42	3.00	2 (13%)
2	SAM	E	301	-	23,29,29	1.06	2 (8%)	15,42,42	3.10	3 (20%)
2	SAM	F	301	-	23,29,29	1.09	2 (8%)	15,42,42	3.06	2 (13%)
2	SAM	G	301	-	23,29,29	1.07	2 (8%)	15,42,42	3.07	2 (13%)
2	SAM	H	301	-	23,29,29	1.06	2 (8%)	15,42,42	3.33	3 (20%)

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	SAM	A	301	-	-	0/8/33/33	0/3/3/3
2	SAM	B	301	-	-	0/8/33/33	0/3/3/3
2	SAM	C	301	-	-	0/8/33/33	0/3/3/3
2	SAM	D	301	-	-	0/8/33/33	0/3/3/3
2	SAM	E	301	-	-	0/8/33/33	0/3/3/3
2	SAM	F	301	-	-	0/8/33/33	0/3/3/3
2	SAM	G	301	-	-	0/8/33/33	0/3/3/3
2	SAM	H	301	-	-	0/8/33/33	0/3/3/3

The worst 5 of 16 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	301	SAM	C2-N1	2.32	1.38	1.33
2	E	301	SAM	C2-N1	2.35	1.38	1.33
2	G	301	SAM	C2-N1	2.35	1.38	1.33
2	H	301	SAM	C2-N1	2.37	1.38	1.33
2	B	301	SAM	C2-N1	2.38	1.38	1.33

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	301	SAM	N3-C2-N1	-11.77	119.62	128.87
2	F	301	SAM	N3-C2-N1	-11.39	119.93	128.87
2	B	301	SAM	N3-C2-N1	-11.29	120.00	128.87
2	G	301	SAM	N3-C2-N1	-11.22	120.06	128.87
2	C	301	SAM	N3-C2-N1	-11.09	120.16	128.87

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

8 monomers are involved in 41 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	301	SAM	7	0
2	B	301	SAM	2	0
2	C	301	SAM	6	0
2	D	301	SAM	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	E	301	SAM	8	0
2	F	301	SAM	4	0
2	G	301	SAM	6	0
2	H	301	SAM	6	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 ⓘ

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	201/232 (86%)	0.03	2 (0%) 84 69	43, 66, 95, 104	0
1	B	194/232 (83%)	-0.07	1 (0%) 91 83	42, 70, 101, 113	0
1	C	204/232 (87%)	0.04	0 100 100	47, 70, 92, 108	0
1	D	188/232 (81%)	0.06	1 (0%) 91 83	49, 72, 103, 123	0
1	E	193/232 (83%)	0.10	3 (1%) 74 55	47, 70, 108, 123	0
1	F	194/232 (83%)	0.14	0 100 100	43, 72, 102, 121	0
1	G	200/232 (86%)	-0.04	3 (1%) 76 58	46, 72, 109, 131	0
1	H	193/232 (83%)	0.03	3 (1%) 74 55	49, 74, 107, 133	0
All	All	1567/1856 (84%)	0.04	13 (0%) 87 75	42, 71, 104, 133	0

The worst 5 of 13 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	149	TYR	3.1
1	G	116	ARG	2.9
1	H	149	TYR	2.8
1	D	149	TYR	2.8
1	E	224	THR	2.7

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 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å ²)	Q<0.9
2	SAM	D	301	27/27	0.85	0.23	0.55	73,90,99,122	0
2	SAM	H	301	27/27	0.91	0.23	0.18	67,74,88,141	0
2	SAM	B	301	27/27	0.92	0.22	0.16	52,70,74,79	0
2	SAM	C	301	27/27	0.92	0.26	0.13	49,63,80,132	0
2	SAM	A	301	27/27	0.95	0.22	-0.06	53,66,80,86	0
2	SAM	G	301	27/27	0.93	0.21	-0.07	57,63,70,70	0
2	SAM	F	301	27/27	0.93	0.24	-0.10	37,68,74,114	0
2	SAM	E	301	27/27	0.89	0.20	-0.63	62,74,100,101	0

6.5 Other polymers [i](#)

There are no such residues in this entry.