wwPDB X-ray Structure Validation Summary Report (i)

May 14, 2020 – 12:16 am BST

PDB ID	:	5HYN
Title	:	Structure of Human Polycomb Repressive Complex 2 (PRC2) with oncogenic histone H3K27M peptide
Authors	:	Zhang, Y.; Justin, N.; Wilson, J.R.; Gamblin, S.J.
Deposited on	:	2016-02-01
Resolution	:	2.95 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 following versions of software and data (see references (1)) were used in the production of this report:

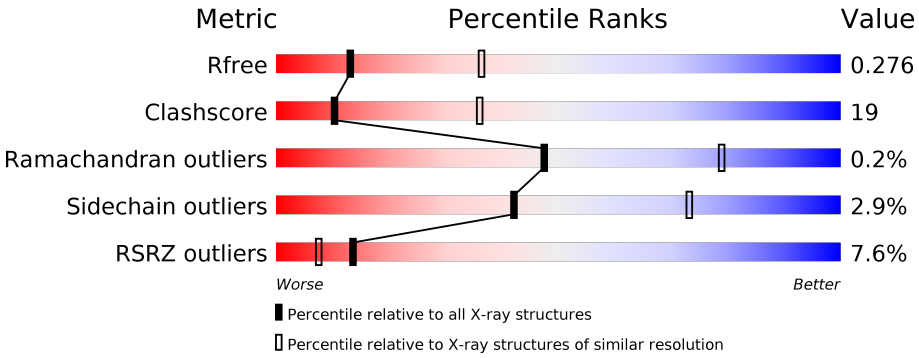
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.95 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.



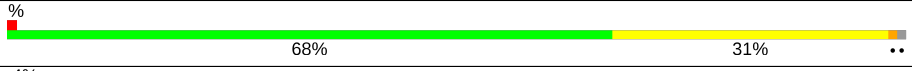
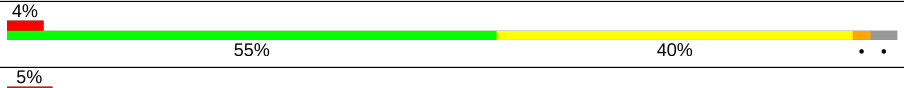
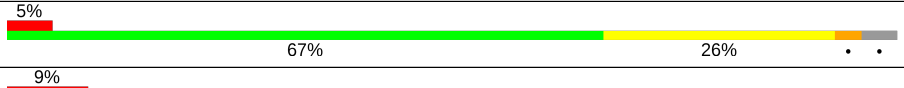
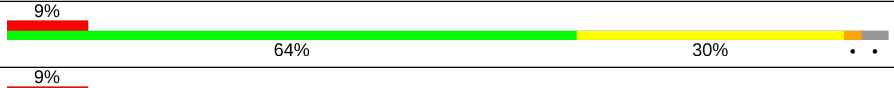
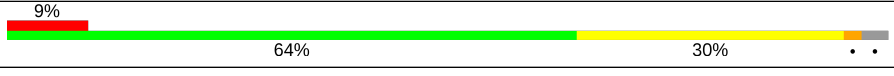
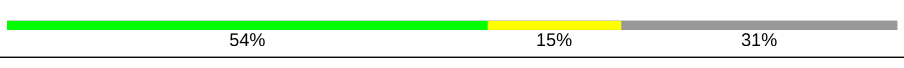
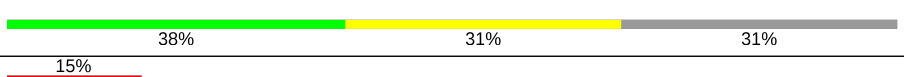
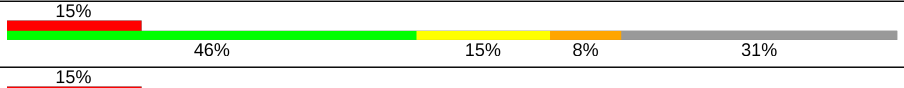
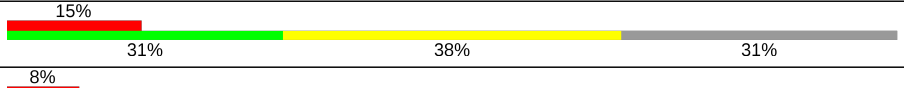
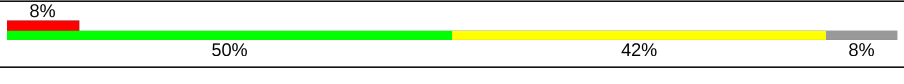
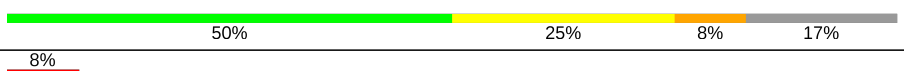
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3104 (3.00-2.92)
Clashscore	141614	3462 (3.00-2.92)
Ramachandran outliers	138981	3340 (3.00-2.92)
Sidechain outliers	138945	3343 (3.00-2.92)
RSRZ outliers	127900	2986 (3.00-2.92)

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 <=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	746	7% 47% 28% • 22%
1	F	746	8% 48% 25% • 24%
1	K	746	10% 50% 23% • 25%
1	Q	746	9% 45% 27% • 24%
2	B	367	3% 62% 37% ••
2	G	367	4% 67% 31% ••

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	L	367	
2	R	367	
3	C	129	
3	H	129	
3	M	129	
3	S	129	
4	D	13	
4	I	13	
4	O	13	
4	T	13	
5	E	12	
5	J	12	
5	P	12	
5	U	12	

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 35028 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 Histone-lysine N-methyltransferase EZH2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	581	Total	C	N	O	S	0	0	0
			4673	2932	825	874	42			
1	F	568	Total	C	N	O	S	0	0	0
			4567	2869	804	852	42			
1	K	562	Total	C	N	O	S	0	0	0
			4521	2836	799	844	42			
1	Q	565	Total	C	N	O	S	0	0	0
			4542	2850	801	849	42			

- Molecule 2 is a protein called Polycomb protein EED.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	365	Total	C	N	O	S	0	0	0
			2959	1873	521	543	22			
2	G	365	Total	C	N	O	S	0	0	0
			2959	1873	521	543	22			
2	L	365	Total	C	N	O	S	0	0	0
			2959	1873	521	543	22			
2	R	365	Total	C	N	O	S	0	0	0
			2959	1873	521	543	22			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	75	GLY	-	expression tag	UNP O75530
B	76	SER	-	expression tag	UNP O75530
G	75	GLY	-	expression tag	UNP O75530
G	76	SER	-	expression tag	UNP O75530
L	75	GLY	-	expression tag	UNP O75530
L	76	SER	-	expression tag	UNP O75530
R	75	GLY	-	expression tag	UNP O75530
R	76	SER	-	expression tag	UNP O75530

- Molecule 3 is a protein called Polycomb protein SUZ12.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	125	Total	C	N	O	S	0	0	0
			1042	657	180	193	12			
3	H	124	Total	C	N	O	S	0	0	0
			1032	651	177	192	12			
3	M	125	Total	C	N	O	S	0	0	0
			1042	657	180	193	12			
3	S	125	Total	C	N	O	S	0	0	0
			1042	657	180	193	12			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	557	GLY	-	expression tag	UNP Q15022
H	557	GLY	-	expression tag	UNP Q15022
M	557	GLY	-	expression tag	UNP Q15022
S	557	GLY	-	expression tag	UNP Q15022

- Molecule 4 is a protein called H3K27M.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	9	Total	C	N	O	S	0	0	0
			63	38	13	11	1			
4	I	9	Total	C	N	O	S	0	0	0
			63	38	13	11	1			
4	O	9	Total	C	N	O	S	0	0	0
			63	38	13	11	1			
4	T	9	Total	C	N	O	S	0	0	0
			63	38	13	11	1			

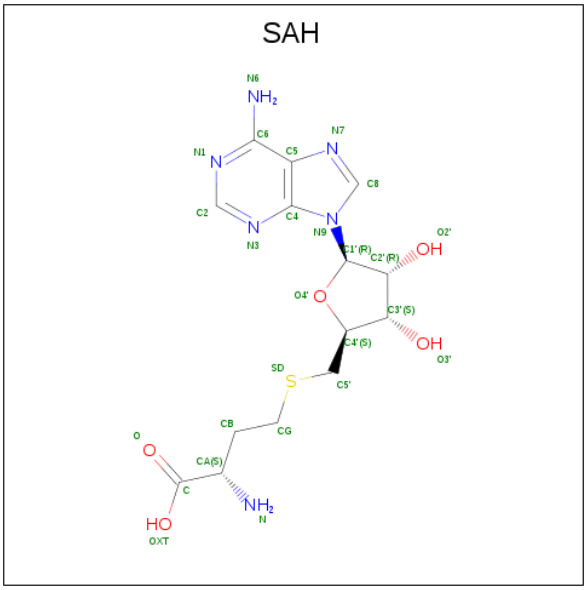
- Molecule 5 is a protein called JARID2 K116me3.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
5	E	11	Total	C	N	O	0	0	0
			96	60	21	15			
5	J	10	Total	C	N	O	0	0	0
			85	54	17	14			
5	P	9	Total	C	N	O	0	0	0
			77	48	16	13			
5	U	10	Total	C	N	O	0	0	0
			85	54	17	14			

- Molecule 6 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	Q	8	Total Zn 8 8	0	0
6	A	8	Total Zn 8 8	0	0
6	K	8	Total Zn 8 8	0	0
6	F	8	Total Zn 8 8	0	0

- Molecule 7 is S-ADENOSYL-L-HOMOCYSTEINE (three-letter code: SAH) (formula: $C_{14}H_{20}N_6O_5S$).

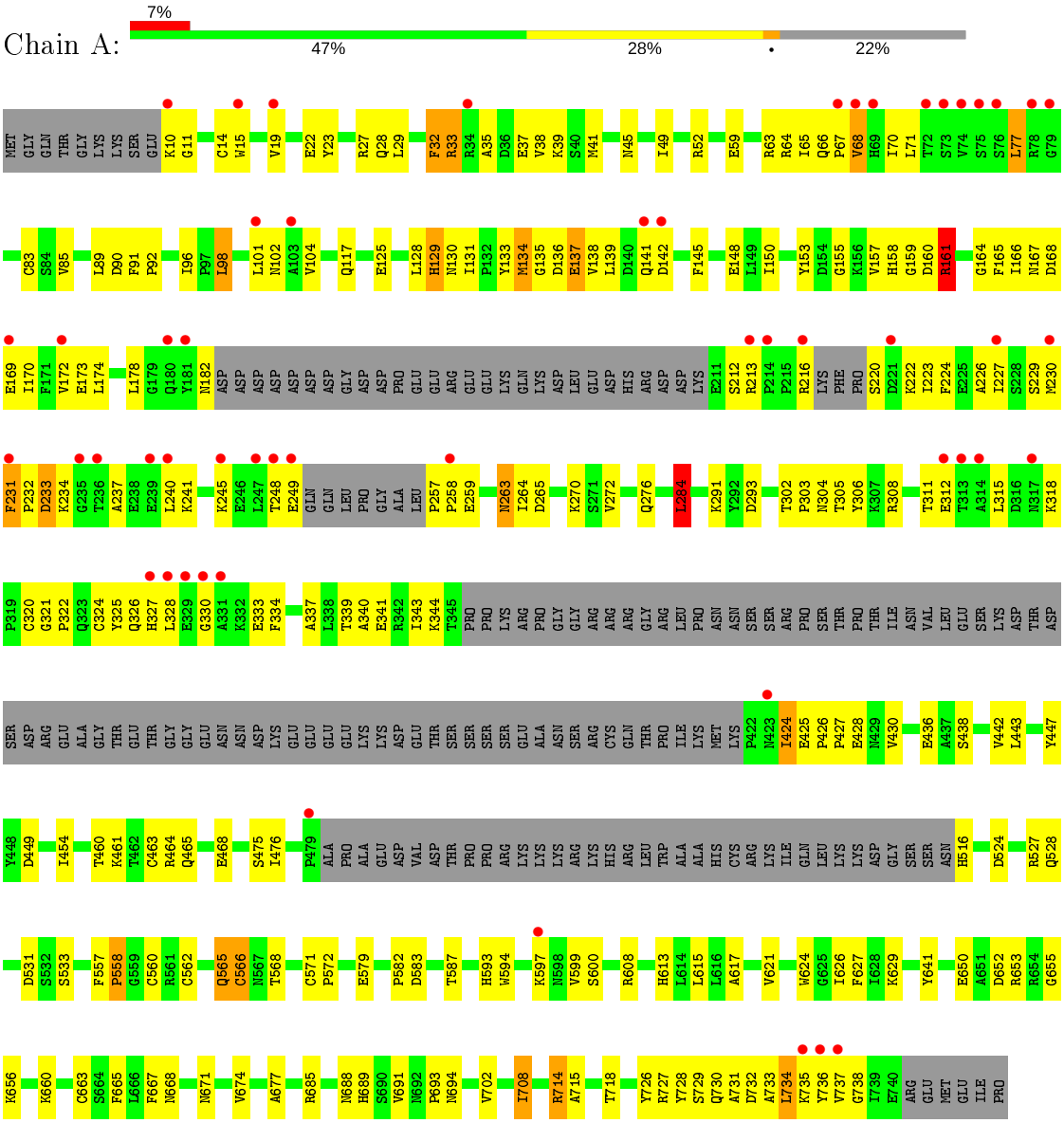


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	A	1	Total C N O S 26 14 6 5 1	0	0
7	F	1	Total C N O S 26 14 6 5 1	0	0
7	K	1	Total C N O S 26 14 6 5 1	0	0
7	Q	1	Total C N O S 26 14 6 5 1	0	0

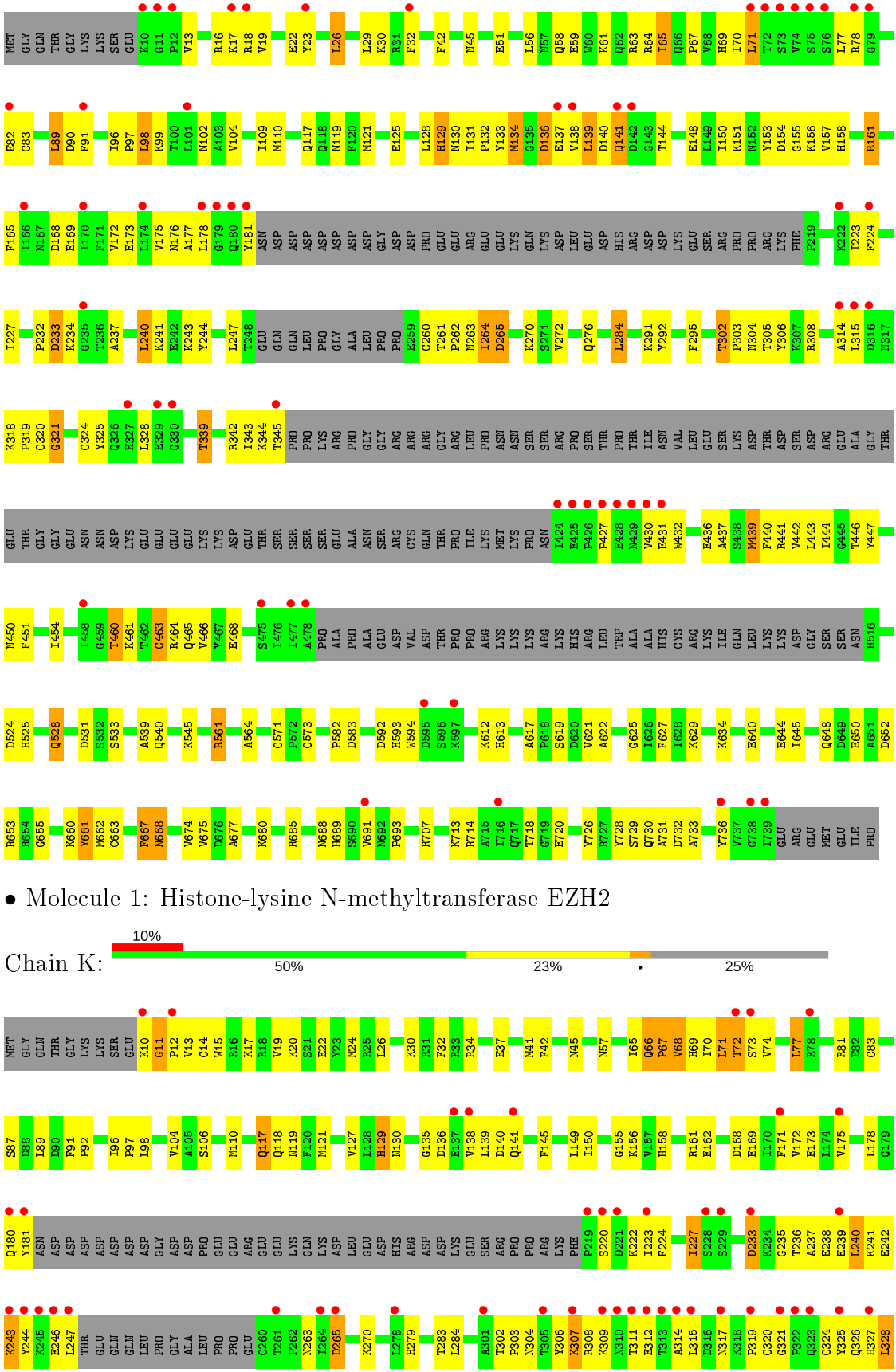
3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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: Histone-lysine N-methyltransferase EZH2

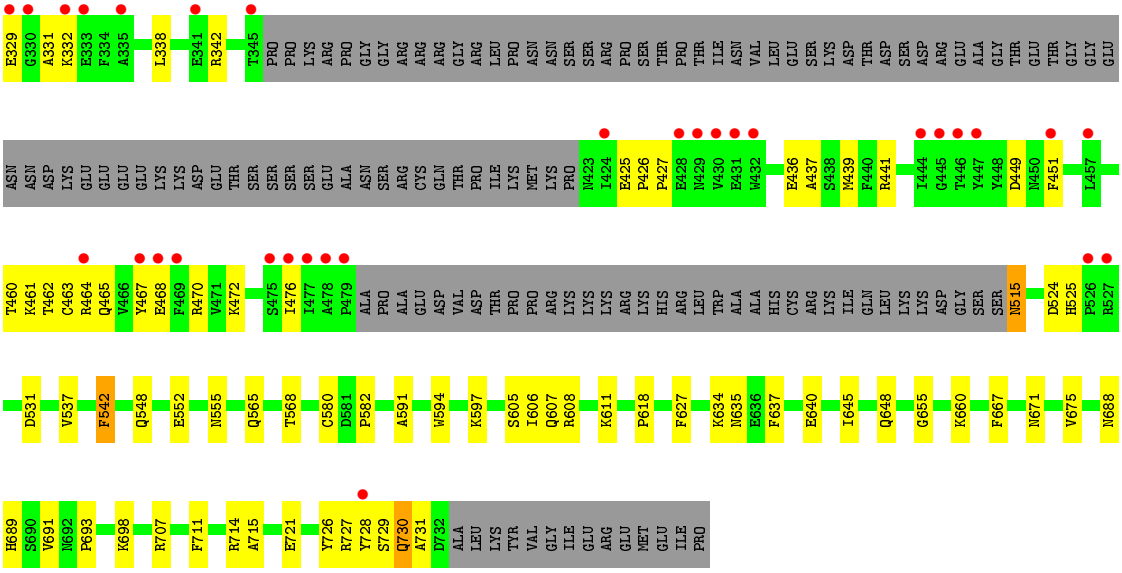


- Molecule 1: Histone-lysine N-methyltransferase EZH2



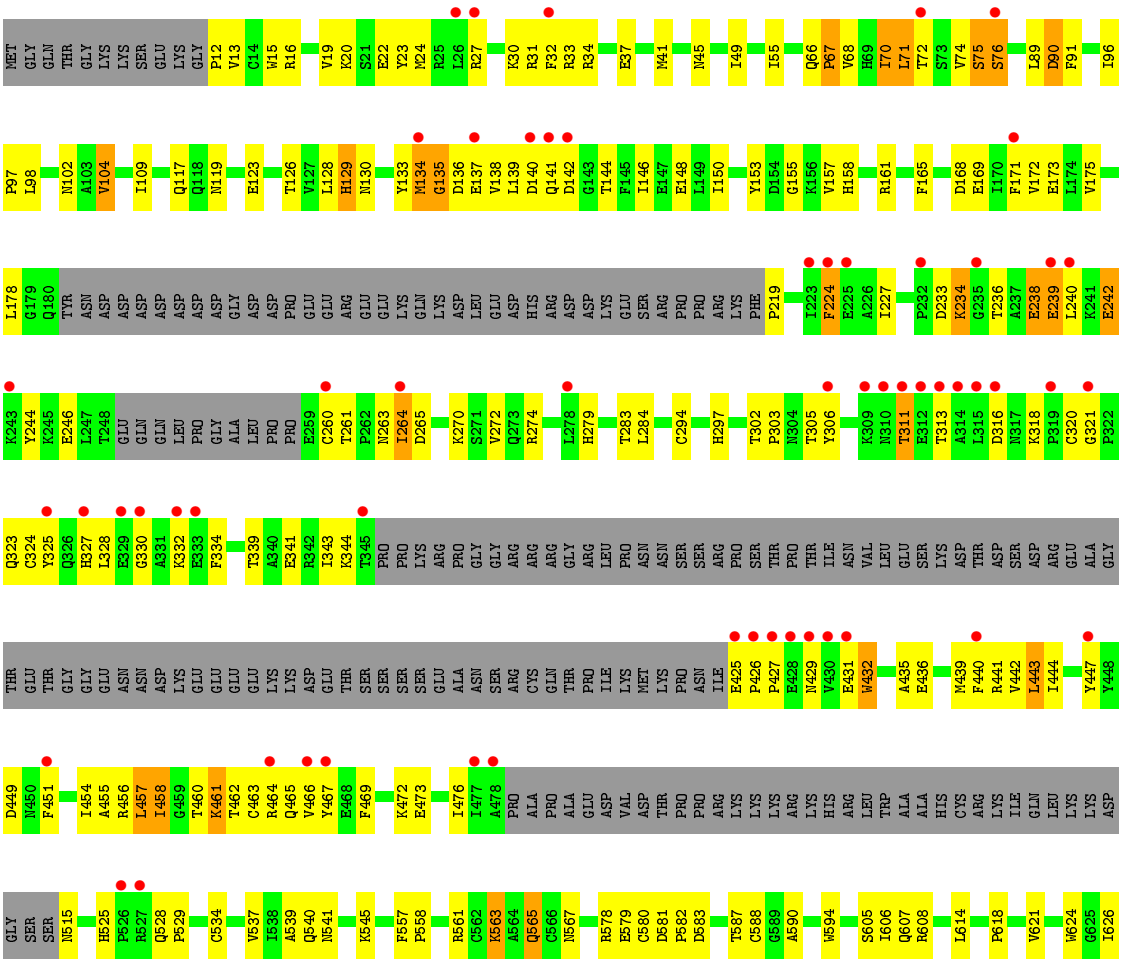
• Molecule 1: Histone-lysine N-methyltransferase EZH2

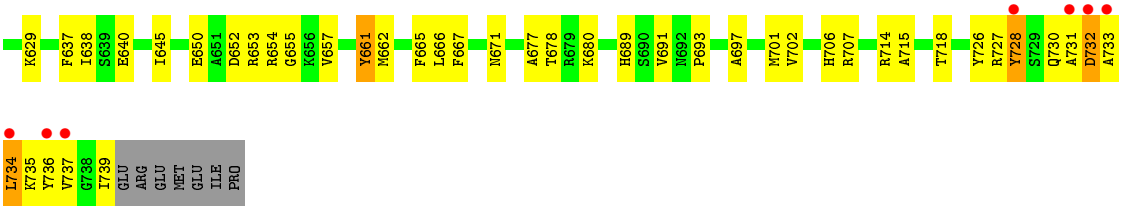
Chain K:



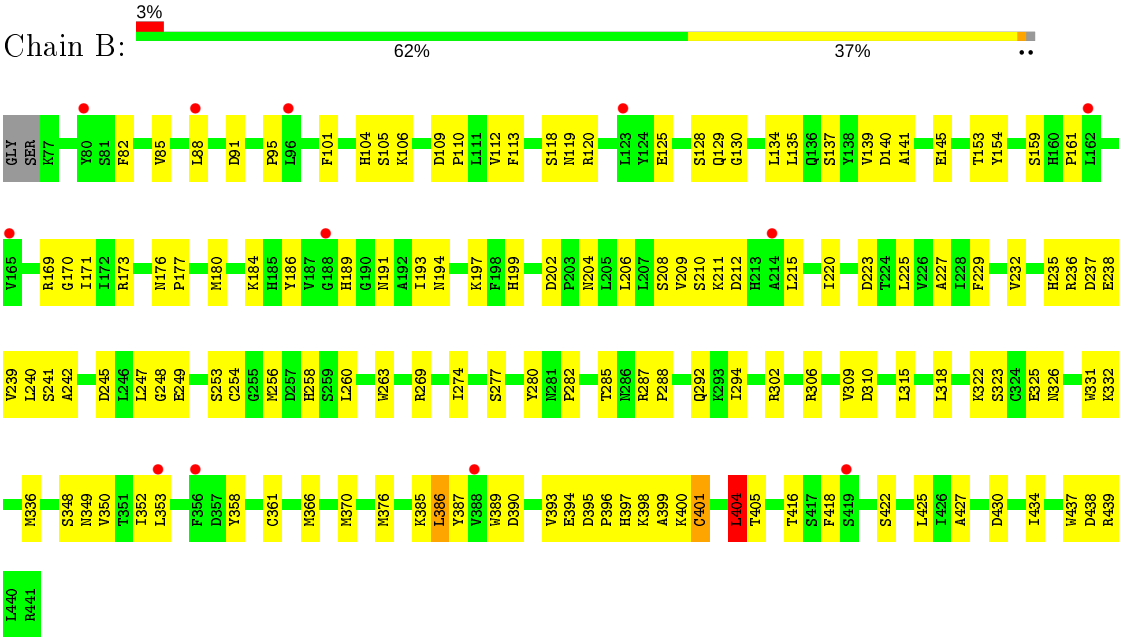
• Molecule 1: Histone-lysine N-methyltransferase EZH2

Chain Q:

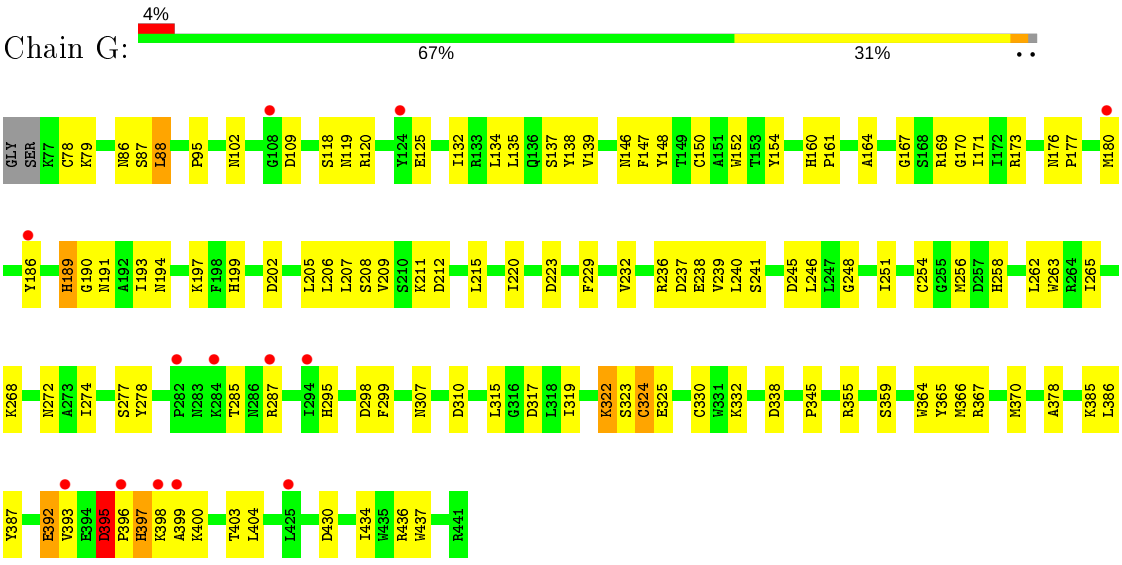




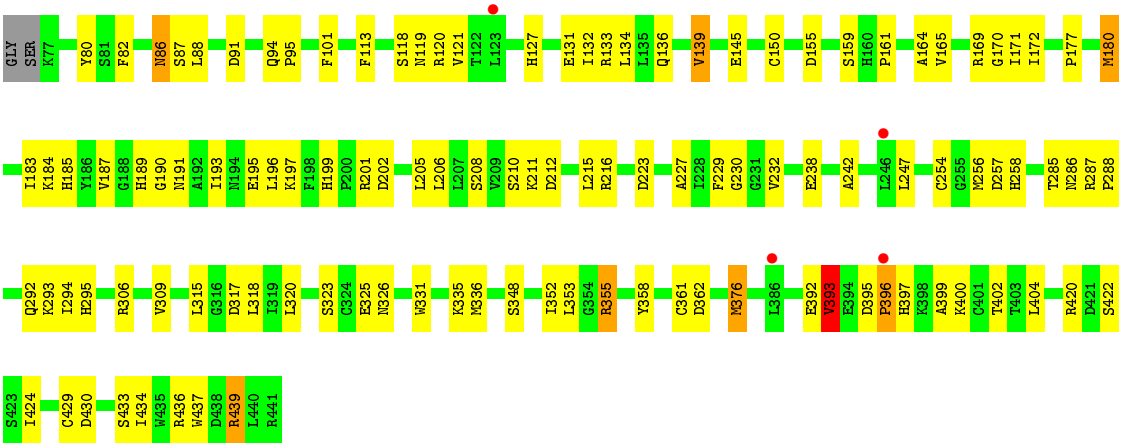
• Molecule 2: Polycomb protein EED



• Molecule 2: Polycomb protein EED



• Molecule 2: Polycomb protein EED

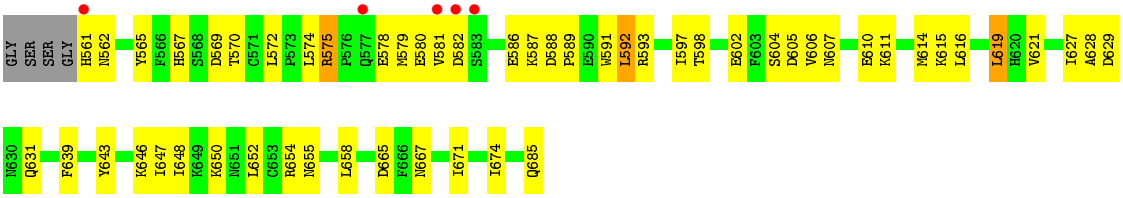


• Molecule 2: Polycomb protein EED

Chain R:

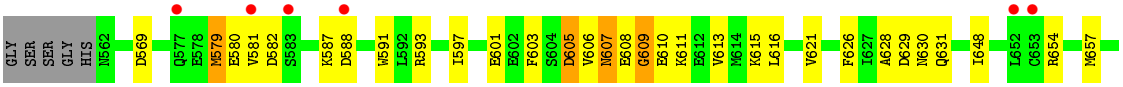
• Molecule 3: Polycomb protein SUZ12

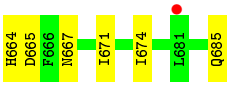
Chain C:



• Molecule 3: Polycomb protein SUZ12

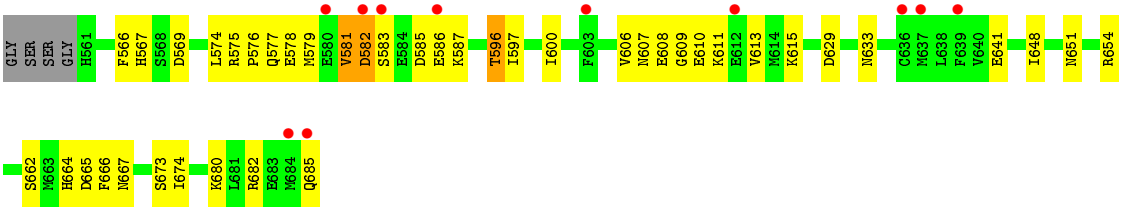
Chain H:





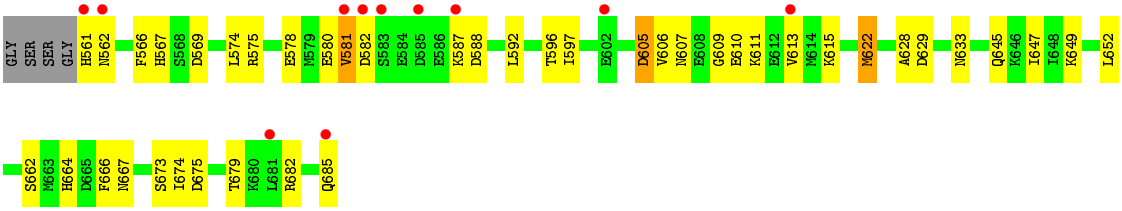
• Molecule 3: Polycomb protein SUZ12

Chain M:



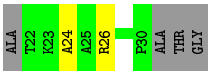
• Molecule 3: Polycomb protein SUZ12

Chain S:



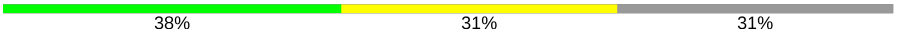
• Molecule 4: H3K27M

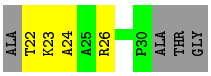
Chain D:



• Molecule 4: H3K27M

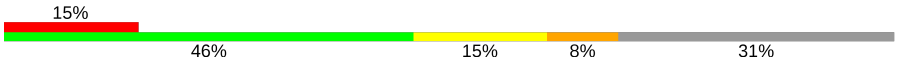
Chain I:

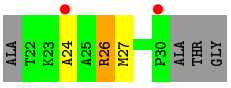




• Molecule 4: H3K27M

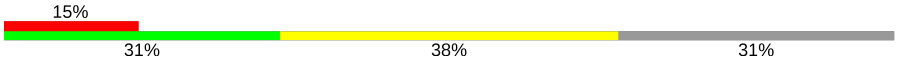
Chain O:

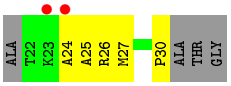




• Molecule 4: H3K27M

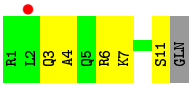
Chain T:





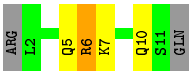
- Molecule 5: JARID2 K116me3

Chain E:



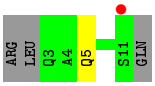
- Molecule 5: JARID2 K116me3

Chain J:



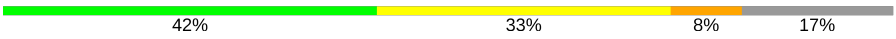
- Molecule 5: JARID2 K116me3

Chain P:



- Molecule 5: JARID2 K116me3

Chain U:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	131.64Å 171.51Å 274.55Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	95.01 – 2.95 104.43 – 2.94	Depositor EDS
% Data completeness (in resolution range)	99.8 (95.01-2.95) 99.9 (104.43-2.94)	Depositor EDS
R_{merge}	0.17	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.03 (at 2.96Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.219 , 0.273 0.224 , 0.276	Depositor DCC
R_{free} test set	6527 reflections (4.93%)	wwPDB-VP
Wilson B-factor (Å^2)	63.5	Xtriage
Anisotropy	0.627	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 72.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	35028	wwPDB-VP
Average B, all atoms (Å^2)	88.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 33.92 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 7.4013e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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: ZN, SAH, M3L

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	0.66	6/4777 (0.1%)	0.80	8/6443 (0.1%)
1	F	0.67	6/4667 (0.1%)	0.86	14/6293 (0.2%)
1	K	0.61	1/4621 (0.0%)	0.84	8/6232 (0.1%)
1	Q	0.72	6/4641 (0.1%)	0.93	13/6258 (0.2%)
2	B	0.50	0/3034	0.75	1/4107 (0.0%)
2	G	0.57	1/3034 (0.0%)	0.84	5/4107 (0.1%)
2	L	0.58	2/3034 (0.1%)	0.81	3/4107 (0.1%)
2	R	0.59	0/3034	0.82	1/4107 (0.0%)
3	C	0.54	0/1063	0.85	2/1427 (0.1%)
3	H	0.58	1/1052 (0.1%)	0.78	1/1412 (0.1%)
3	M	0.68	3/1063 (0.3%)	0.83	2/1427 (0.1%)
3	S	0.51	0/1063	0.74	2/1427 (0.1%)
4	D	0.50	0/63	0.86	0/83
4	I	0.50	0/63	0.72	0/83
4	O	0.57	0/63	1.07	1/83 (1.2%)
4	T	0.50	0/63	0.88	0/83
5	E	0.60	0/84	0.82	0/110
5	J	0.80	0/73	1.43	1/96 (1.0%)
5	P	0.56	0/65	0.62	0/85
5	U	0.95	1/73 (1.4%)	1.05	0/96
All	All	0.62	27/35630 (0.1%)	0.84	62/48066 (0.1%)

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	F	0	1
1	K	0	1

Continued on next page...

Continued from previous page...

Mol	Chain	#Chirality outliers	#Planarity outliers
1	Q	0	4
2	G	0	2
2	L	0	3
2	R	0	1
3	H	0	1
All	All	0	15

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	33	ARG	NE-CZ	-12.29	1.17	1.33
1	F	181	TYR	CB-CG	-10.42	1.36	1.51
1	A	32	PHE	CE1-CZ	8.05	1.52	1.37
1	A	249	GLU	CD-OE1	7.87	1.34	1.25
2	G	396	PRO	N-CD	7.52	1.58	1.47

The worst 5 of 62 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	Q	443	LEU	CB-CG-CD1	-13.41	88.20	111.00
1	Q	33	ARG	NE-CZ-NH2	-11.52	114.54	120.30
1	Q	70	ILE	CG1-CB-CG2	-10.19	88.99	111.40
5	J	6	ARG	NE-CZ-NH2	-9.21	115.69	120.30
2	G	324	CYS	CA-CB-SG	-8.72	98.31	114.00

There are no chirality outliers.

5 of 15 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	102	ASN	Peptide
1	A	134	MET	Peptide
1	F	564	ALA	Peptide
2	G	392	GLU	Peptide
2	G	395	ASP	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	4673	0	4524	247	1
1	F	4567	0	4427	192	1
1	K	4521	0	4373	180	1
1	Q	4542	0	4400	233	1
2	B	2959	0	2881	148	0
2	G	2959	0	2881	98	0
2	L	2959	0	2881	94	0
2	R	2959	0	2881	98	0
3	C	1042	0	1021	56	0
3	H	1032	0	1014	37	0
3	M	1042	0	1021	49	0
3	S	1042	0	1021	35	0
4	D	63	0	68	3	0
4	I	63	0	68	4	0
4	O	63	0	68	4	0
4	T	63	0	68	12	0
5	E	96	0	106	3	0
5	J	85	0	90	5	0
5	P	77	0	79	1	0
5	U	85	0	90	8	0
6	A	8	0	0	0	0
6	F	8	0	0	0	0
6	K	8	0	0	0	0
6	Q	8	0	0	0	0
7	A	26	0	19	6	0
7	F	26	0	19	2	0
7	K	26	0	19	1	0
7	Q	26	0	19	1	0
All	All	35028	0	34038	1288	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 19.

The worst 5 of 1288 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:Q:432:TRP:CH2	1:Q:461:LYS:HG3	1.65	1.32
1:Q:318:LYS:NZ	1:Q:324:CYS:SG	2.26	1.08
1:Q:432:TRP:HD1	1:Q:469:PHE:CD1	1.75	1.05
1:Q:432:TRP:HH2	1:Q:461:LYS:CG	1.70	1.04
1:A:430:VAL:HG21	1:A:465:GLN:HE21	1.19	1.03

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:568:THR:OG1	1:Q:561:ARG:NH1[3_644]	2.11	0.09
1:F:561:ARG:NH1	1:K:568:THR:OG1[2_545]	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	569/746 (76%)	534 (94%)	34 (6%)	1 (0%)	47	79
1	F	558/746 (75%)	526 (94%)	31 (6%)	1 (0%)	47	79
1	K	552/746 (74%)	527 (96%)	23 (4%)	2 (0%)	34	69
1	Q	555/746 (74%)	527 (95%)	27 (5%)	1 (0%)	47	79
2	B	363/367 (99%)	348 (96%)	15 (4%)	0	100	100
2	G	363/367 (99%)	355 (98%)	8 (2%)	0	100	100
2	L	363/367 (99%)	354 (98%)	9 (2%)	0	100	100
2	R	363/367 (99%)	352 (97%)	11 (3%)	0	100	100
3	C	123/129 (95%)	120 (98%)	3 (2%)	0	100	100
3	H	122/129 (95%)	119 (98%)	2 (2%)	1 (1%)	19	53
3	M	123/129 (95%)	118 (96%)	4 (3%)	1 (1%)	19	53
3	S	123/129 (95%)	121 (98%)	2 (2%)	0	100	100
4	D	7/13 (54%)	7 (100%)	0	0	100	100
4	I	7/13 (54%)	7 (100%)	0	0	100	100
4	O	7/13 (54%)	7 (100%)	0	0	100	100
4	T	7/13 (54%)	7 (100%)	0	0	100	100
5	E	8/12 (67%)	8 (100%)	0	0	100	100
5	J	7/12 (58%)	7 (100%)	0	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
5	P	6/12 (50%)	6 (100%)	0	0	100 100
5	U	7/12 (58%)	6 (86%)	1 (14%)	0	100 100
All	All	4233/5068 (84%)	4056 (96%)	170 (4%)	7 (0%)	47 79

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	68	VAL
1	K	66	GLN
1	F	139	LEU
3	M	582	ASP
1	K	265	ASP

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	520/667 (78%)	496 (95%)	24 (5%)	27 60
1	F	507/667 (76%)	486 (96%)	21 (4%)	30 64
1	K	503/667 (75%)	481 (96%)	22 (4%)	28 62
1	Q	505/667 (76%)	485 (96%)	20 (4%)	31 64
2	B	328/329 (100%)	322 (98%)	6 (2%)	59 82
2	G	328/329 (100%)	319 (97%)	9 (3%)	44 74
2	L	328/329 (100%)	326 (99%)	2 (1%)	86 94
2	R	328/329 (100%)	327 (100%)	1 (0%)	92 97
3	C	119/121 (98%)	118 (99%)	1 (1%)	81 92
3	H	118/121 (98%)	115 (98%)	3 (2%)	47 76
3	M	119/121 (98%)	117 (98%)	2 (2%)	60 83
3	S	119/121 (98%)	118 (99%)	1 (1%)	81 92
4	D	6/7 (86%)	6 (100%)	0	100 100
4	I	6/7 (86%)	6 (100%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
4	O	6/7 (86%)	6 (100%)	0	100 100
4	T	6/7 (86%)	6 (100%)	0	100 100
5	E	8/9 (89%)	6 (75%)	2 (25%)	0 2
5	J	7/9 (78%)	7 (100%)	0	100 100
5	P	6/9 (67%)	6 (100%)	0	100 100
5	U	7/9 (78%)	7 (100%)	0	100 100
All	All	3874/4532 (86%)	3760 (97%)	114 (3%)	42 73

5 of 114 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	663	CYS
3	H	579	MET
1	Q	525	HIS
1	F	668	ASN
2	G	223	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 102 such sidechains are listed below:

Mol	Chain	Res	Type
2	G	286	ASN
1	K	326	GLN
2	R	286	ASN
2	G	349	ASN
5	J	10	GLN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

4 non-standard protein/DNA/RNA residues 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	M3L	P	7	5	10,11,12	0.49	0	9,14,16	0.94	0
5	M3L	E	7	5	10,11,12	0.52	0	9,14,16	1.50	2 (22%)
5	M3L	U	7	5	10,11,12	0.58	0	9,14,16	1.69	1 (11%)
5	M3L	J	7	5	10,11,12	0.57	0	9,14,16	1.36	1 (11%)

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	M3L	P	7	5	-	3/9/10/12	-
5	M3L	E	7	5	-	4/9/10/12	-
5	M3L	U	7	5	-	4/9/10/12	-
5	M3L	J	7	5	-	2/9/10/12	-

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	U	7	M3L	CM3-NZ-CM1	-3.38	100.27	108.97
5	E	7	M3L	CM3-NZ-CM1	-3.30	100.49	108.97
5	J	7	M3L	CM2-NZ-CM1	-3.27	100.56	108.97
5	E	7	M3L	CM2-NZ-CM1	2.01	114.14	108.97

There are no chirality outliers.

5 of 13 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	E	7	M3L	N-CA-CB-CG
5	E	7	M3L	CD-CE-NZ-CM2
5	U	7	M3L	CD-CE-NZ-CM2
5	E	7	M3L	CD-CE-NZ-CM3
5	P	7	M3L	CD-CE-NZ-CM2

There are no ring outliers.

3 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	E	7	M3L	1	0
5	U	7	M3L	1	0
5	J	7	M3L	2	0

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 36 ligands modelled in this entry, 32 are monoatomic - leaving 4 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	$\#\|Z\| > 2$
7	SAH	F	1009	-	21,28,28	1.24	2 (9%)	20,40,40	1.77	4 (20%)
7	SAH	A	1009	-	21,28,28	1.15	2 (9%)	20,40,40	1.64	3 (15%)
7	SAH	K	1009	-	21,28,28	1.18	2 (9%)	20,40,40	1.87	3 (15%)
7	SAH	Q	1009	-	21,28,28	1.18	2 (9%)	20,40,40	1.70	2 (10%)

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
7	SAH	F	1009	-	-	2/7/31/31	0/3/3/3
7	SAH	A	1009	-	-	2/7/31/31	0/3/3/3
7	SAH	K	1009	-	-	3/7/31/31	0/3/3/3
7	SAH	Q	1009	-	-	4/7/31/31	0/3/3/3

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	K	1009	SAH	C2-N3	4.22	1.38	1.32
7	Q	1009	SAH	C2-N3	3.68	1.38	1.32
7	F	1009	SAH	C2-N3	3.58	1.37	1.32
7	A	1009	SAH	C2-N3	3.57	1.37	1.32
7	F	1009	SAH	C2-N1	2.76	1.39	1.33

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
7	K	1009	SAH	C5'-SD-CG	-5.62	85.42	102.27
7	A	1009	SAH	N3-C2-N1	-5.19	120.56	128.68
7	K	1009	SAH	N3-C2-N1	-5.19	120.57	128.68
7	Q	1009	SAH	C5'-SD-CG	-5.09	86.98	102.27
7	F	1009	SAH	C5'-SD-CG	-4.96	87.40	102.27

There are no chirality outliers.

5 of 11 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
7	K	1009	SAH	O4'-C4'-C5'-SD
7	K	1009	SAH	C3'-C4'-C5'-SD
7	Q	1009	SAH	C-CA-CB-CG
7	Q	1009	SAH	CA-CB-CG-SD
7	F	1009	SAH	CA-CB-CG-SD

There are no ring outliers.

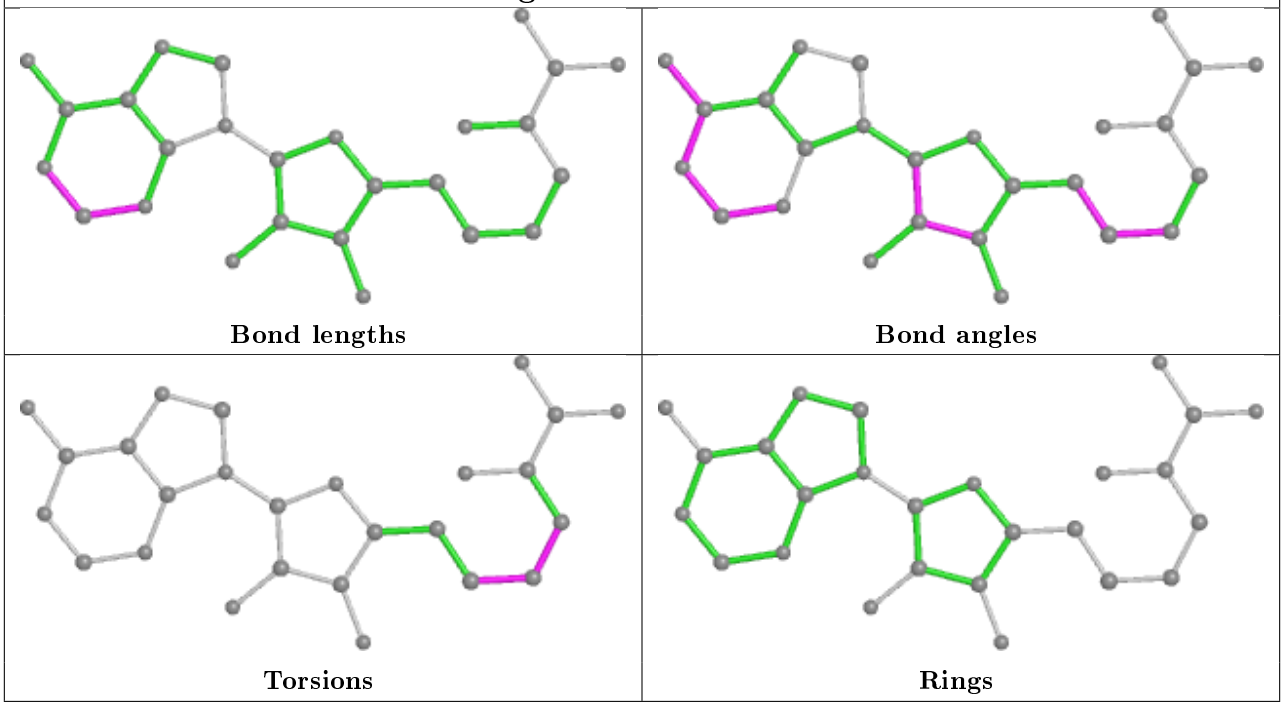
4 monomers are involved in 10 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
7	F	1009	SAH	2	0
7	A	1009	SAH	6	0
7	K	1009	SAH	1	0
7	Q	1009	SAH	1	0

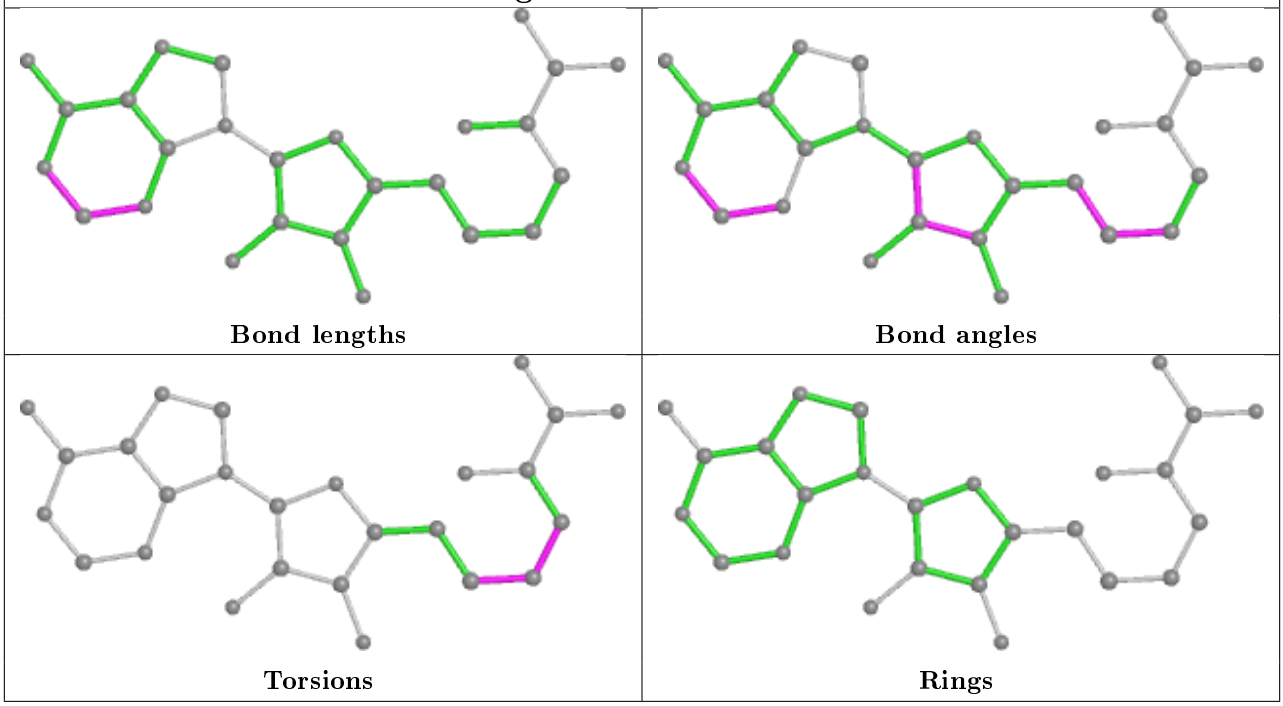
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less 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.

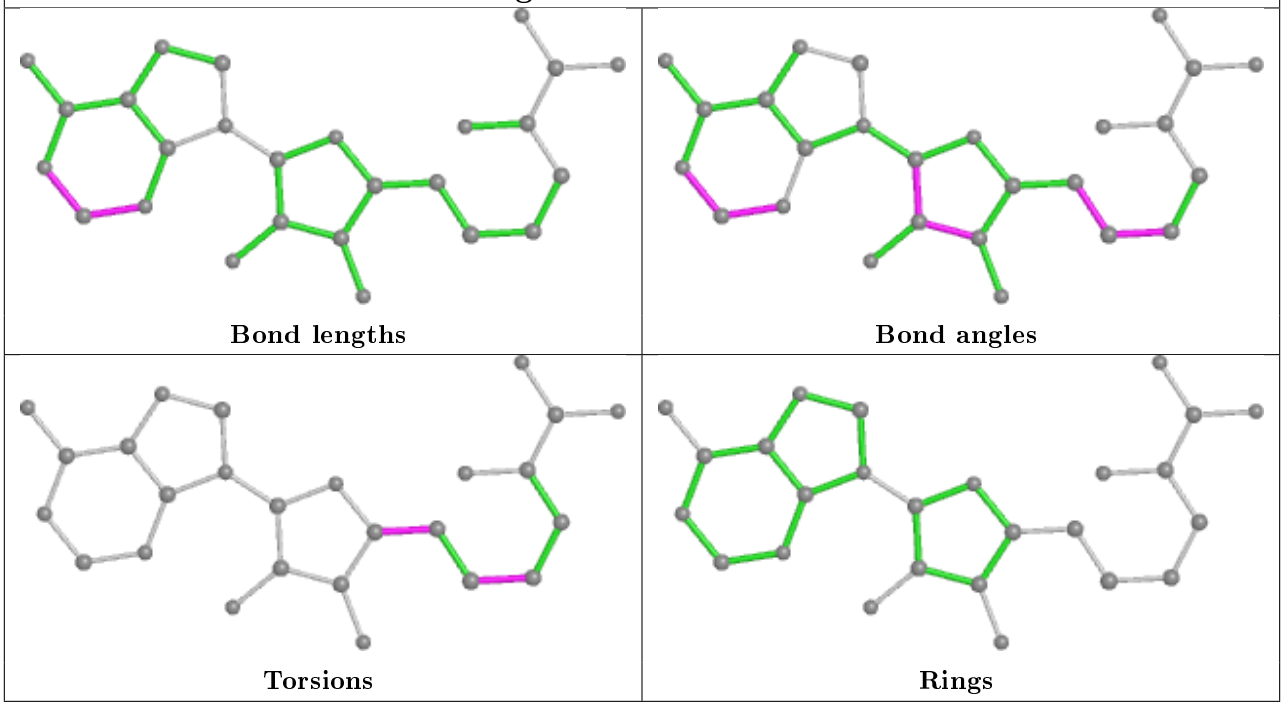
Ligand SAH F 1009



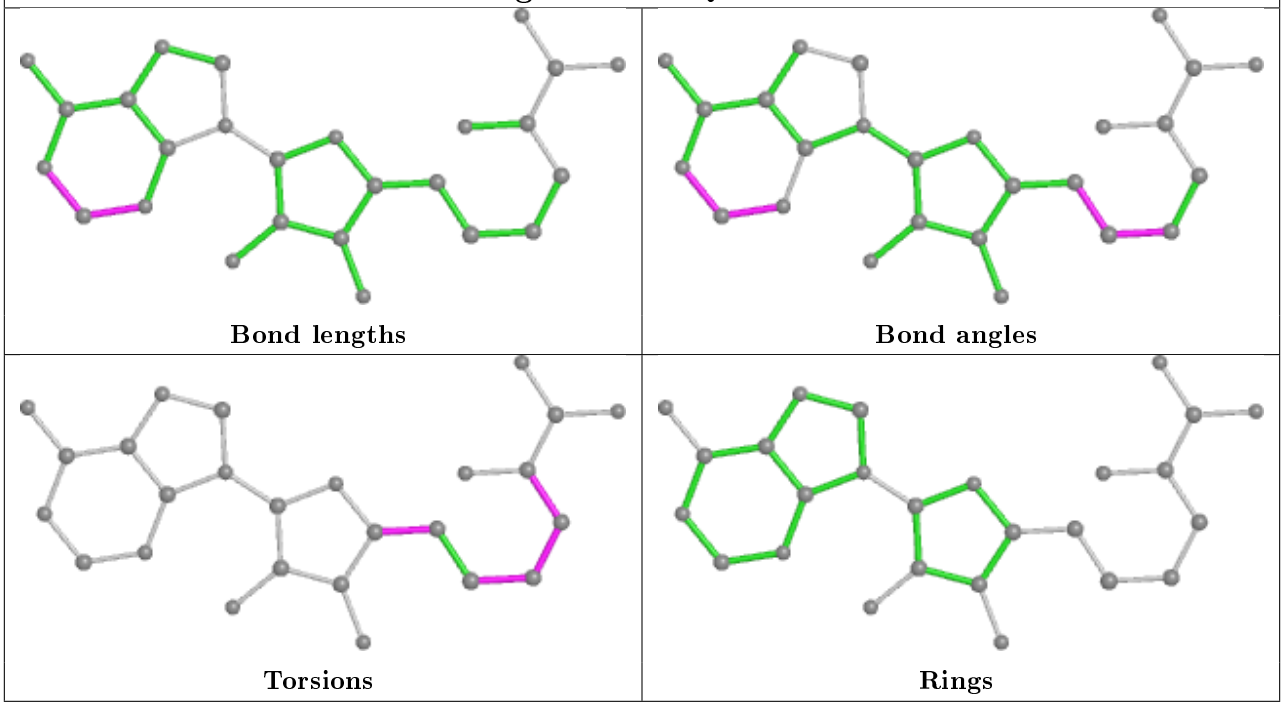
Ligand SAH A 1009



Ligand SAH K 1009



Ligand SAH Q 1009



5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues

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^{th} 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	581/746 (77%)	0.60	53 (9%) 9 5	40, 84, 159, 196	0
1	F	568/746 (76%)	0.67	58 (10%) 6 4	40, 92, 162, 224	0
1	K	562/746 (75%)	0.80	77 (13%) 3 1	43, 92, 165, 189	0
1	Q	565/746 (75%)	0.71	64 (11%) 5 3	45, 86, 165, 202	0
2	B	365/367 (99%)	0.40	12 (3%) 46 30	51, 79, 121, 171	0
2	G	365/367 (99%)	0.38	13 (3%) 42 28	49, 76, 115, 177	0
2	L	365/367 (99%)	0.29	4 (1%) 80 65	41, 65, 103, 140	0
2	R	365/367 (99%)	0.28	5 (1%) 75 59	35, 61, 92, 147	0
3	C	125/129 (96%)	0.47	5 (4%) 38 25	48, 72, 148, 172	0
3	H	124/129 (96%)	0.51	7 (5%) 24 15	58, 89, 143, 187	0
3	M	125/129 (96%)	0.74	11 (8%) 10 5	56, 92, 135, 158	0
3	S	125/129 (96%)	0.66	11 (8%) 10 5	54, 91, 139, 159	0
4	D	9/13 (69%)	0.15	0 100 100	67, 86, 111, 118	0
4	I	9/13 (69%)	0.18	0 100 100	62, 81, 103, 113	0
4	O	9/13 (69%)	1.70	2 (22%) 0 0	68, 76, 88, 93	9 (100%)
4	T	9/13 (69%)	1.74	2 (22%) 0 0	70, 73, 81, 94	9 (100%)
5	E	10/12 (83%)	0.34	1 (10%) 7 4	67, 89, 139, 140	0
5	J	9/12 (75%)	0.04	0 100 100	63, 82, 127, 133	0
5	P	8/12 (66%)	0.52	1 (12%) 3 2	69, 84, 117, 140	0
5	U	9/12 (75%)	0.89	0 100 100	71, 95, 133, 134	0
All	All	4307/5068 (84%)	0.56	326 (7%) 13 7	35, 79, 155, 224	18 (0%)

The worst 5 of 326 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	K	311	THR	13.7
1	Q	314	ALA	12.6
1	Q	235	GLY	12.2
1	Q	430	VAL	10.5
1	F	75	SER	10.4

6.2 Non-standard residues in protein, DNA, RNA chains (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^{th} 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	M3L	P	7	12/13	0.95	0.26	39,57,65,71	0
5	M3L	E	7	12/13	0.97	0.26	40,58,78,90	0
5	M3L	U	7	12/13	0.97	0.24	34,59,68,68	0
5	M3L	J	7	12/13	0.98	0.22	41,52,63,65	0

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, 95^{th} 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
7	SAH	Q	1009	26/26	0.88	0.21	55,77,177,321	0
7	SAH	K	1009	26/26	0.89	0.21	60,74,184,324	0
7	SAH	F	1009	26/26	0.92	0.20	40,75,86,146	0
6	ZN	K	1002	1/1	0.92	0.12	122,122,122,122	0
7	SAH	A	1009	26/26	0.94	0.21	58,77,90,93	0
6	ZN	K	1001	1/1	0.97	0.23	69,69,69,69	0
6	ZN	A	1002	1/1	0.97	0.16	64,64,64,64	0
6	ZN	F	1002	1/1	0.98	0.16	110,110,110,110	0
6	ZN	K	1004	1/1	0.98	0.19	70,70,70,70	0
6	ZN	K	1003	1/1	0.98	0.17	66,66,66,66	0

Continued on next page...

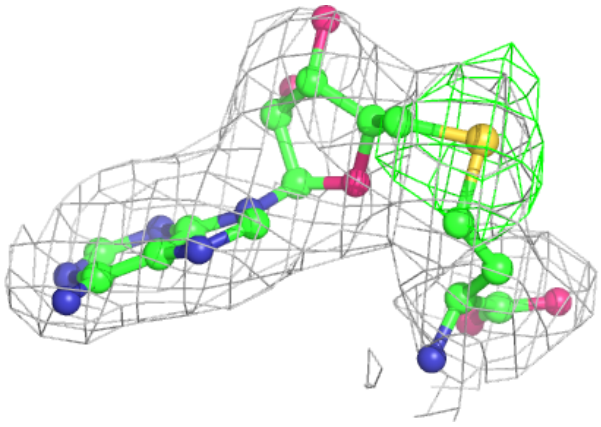
Continued from previous page...

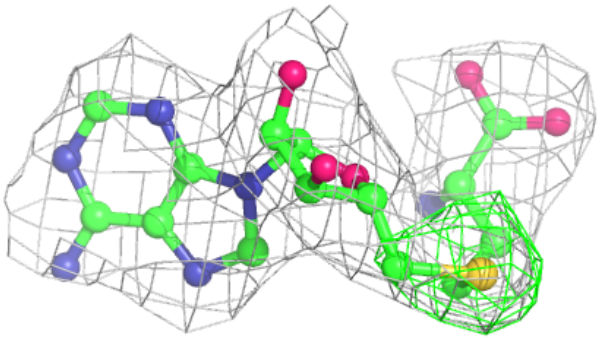
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
6	ZN	K	1007	1/1	0.98	0.21	54,54,54,54	0
6	ZN	Q	1003	1/1	0.98	0.18	69,69,69,69	0
6	ZN	Q	1006	1/1	0.98	0.22	66,66,66,66	0
6	ZN	F	1008	1/1	0.98	0.22	54,54,54,54	0
6	ZN	K	1006	1/1	0.98	0.21	63,63,63,63	0
6	ZN	K	1008	1/1	0.98	0.22	59,59,59,59	0
6	ZN	A	1007	1/1	0.98	0.21	52,52,52,52	0
6	ZN	A	1004	1/1	0.98	0.24	63,63,63,63	0
6	ZN	Q	1002	1/1	0.99	0.15	128,128,128,128	0
6	ZN	F	1006	1/1	0.99	0.20	60,60,60,60	0
6	ZN	Q	1008	1/1	0.99	0.22	50,50,50,50	0
6	ZN	A	1003	1/1	0.99	0.22	51,51,51,51	0
6	ZN	F	1003	1/1	0.99	0.21	65,65,65,65	0
6	ZN	Q	1001	1/1	0.99	0.21	60,60,60,60	0
6	ZN	A	1008	1/1	0.99	0.21	58,58,58,58	0
6	ZN	A	1006	1/1	0.99	0.20	58,58,58,58	0
6	ZN	F	1004	1/1	0.99	0.21	65,65,65,65	0
6	ZN	Q	1004	1/1	0.99	0.17	77,77,77,77	0
6	ZN	A	1005	1/1	0.99	0.21	56,56,56,56	0
6	ZN	K	1005	1/1	0.99	0.17	74,74,74,74	0
6	ZN	F	1007	1/1	0.99	0.21	54,54,54,54	0
6	ZN	Q	1005	1/1	0.99	0.19	72,72,72,72	0
6	ZN	Q	1007	1/1	0.99	0.22	54,54,54,54	0
6	ZN	A	1001	1/1	0.99	0.20	54,54,54,54	0
6	ZN	F	1005	1/1	0.99	0.20	67,67,67,67	0
6	ZN	F	1001	1/1	0.99	0.20	63,63,63,63	0

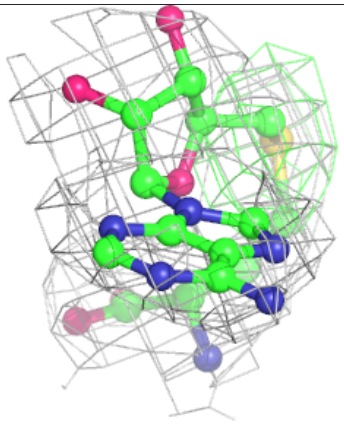
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around SAH Q 1009:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

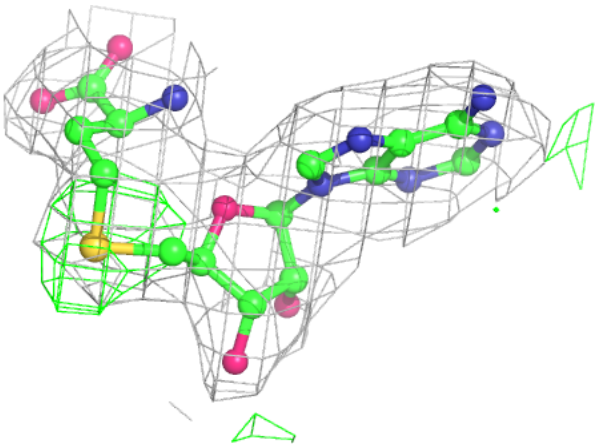


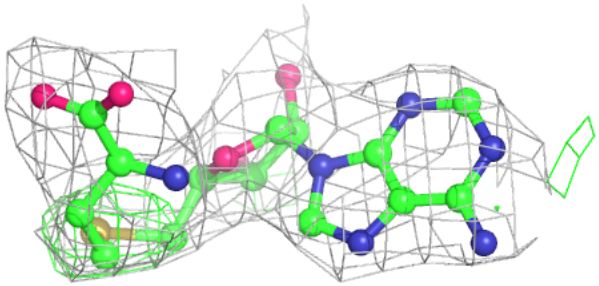


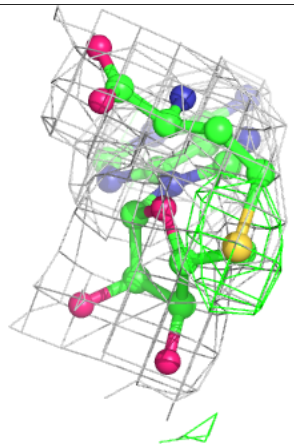


Electron density around SAH K 1009:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

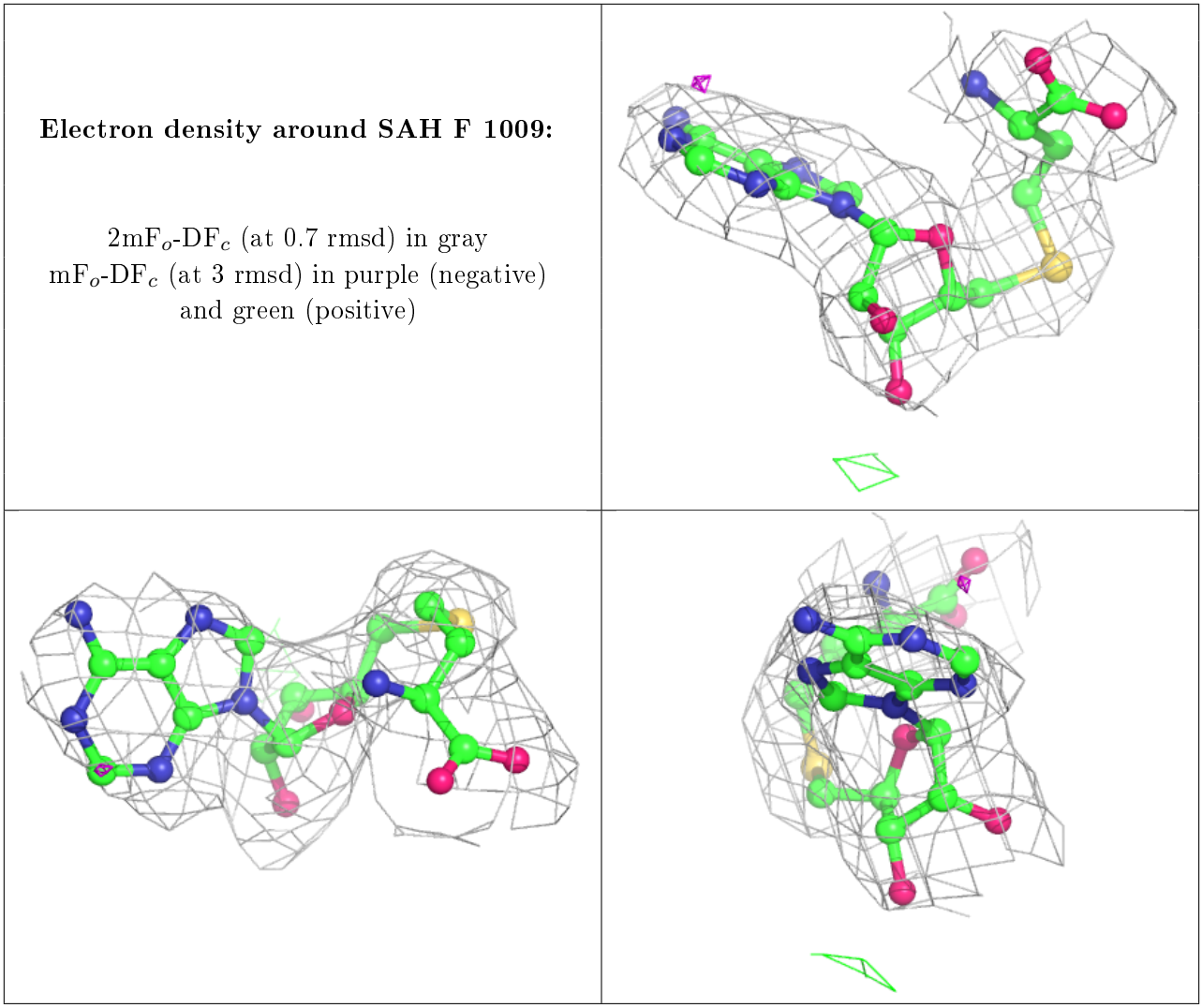






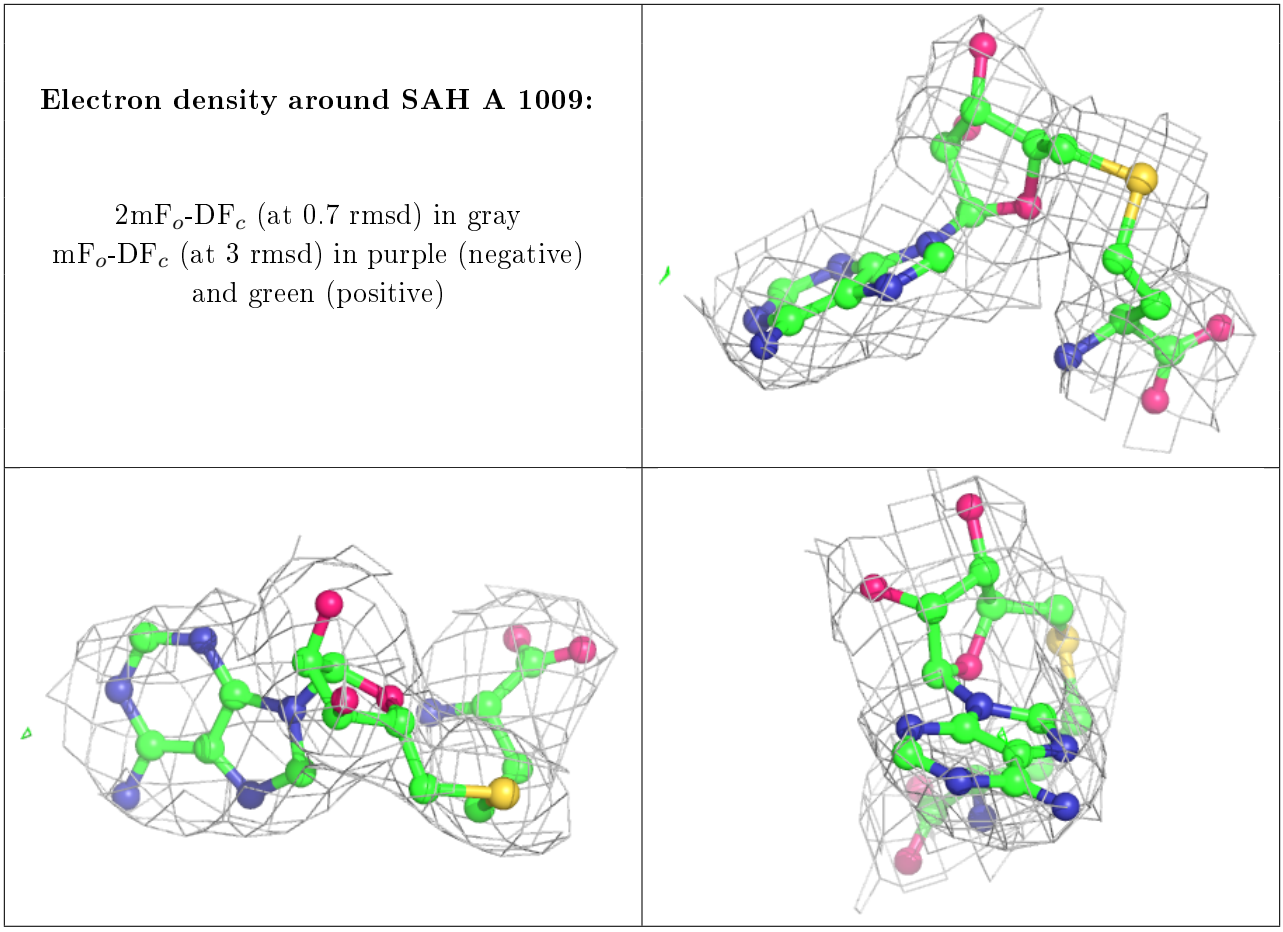
Electron density around SAH F 1009:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)



Electron density around SAH A 1009:

$2mF_o$-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.