wwPDB X-ray Structure Validation Summary Report (i)

Mar 4, 2024 – 10:34 PM EST

PDB ID	:	1U2Z
Title	:	Crystal structure of histone K79 methyltransferase Dot1p from yeast
Authors	:	Sawada, K.; Yang, Z.; Horton, J.R.; Collins, R.E.; Zhang, X.; Cheng, X.
Deposited on	:	2004-07-20
Resolution	:	2.20 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

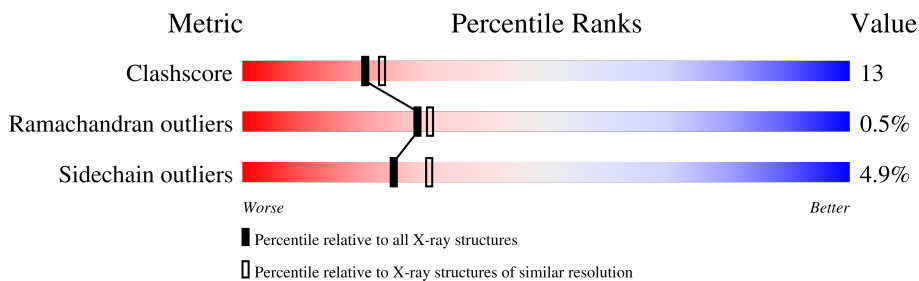|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.20 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	Percentile Ranks	Value
Clashscore		13
Ramachandran outliers		0.5%
Sidechain outliers		4.9%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)

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 of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	433	65% 20% • 12%
1	B	433	69% 15% • 12%
1	C	433	67% 19% • 11%

2   Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 9909 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, H3 lysine-79 specific.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	379	Total	C	N	O	S	0	0	0
			3082	1976	508	581	17			
1	B	379	Total	C	N	O	S	0	0	0
			3085	1979	508	581	17			
1	C	385	Total	C	N	O	S	0	0	0
			3131	2004	520	590	17			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	150	MET	-	expression tag	UNP Q04089
A	151	GLY	-	expression tag	UNP Q04089
A	152	HIS	-	expression tag	UNP Q04089
A	153	HIS	-	expression tag	UNP Q04089
A	154	HIS	-	expression tag	UNP Q04089
A	155	HIS	-	expression tag	UNP Q04089
A	156	HIS	-	expression tag	UNP Q04089
A	157	HIS	-	expression tag	UNP Q04089
B	150	MET	-	expression tag	UNP Q04089
B	151	GLY	-	expression tag	UNP Q04089
B	152	HIS	-	expression tag	UNP Q04089
B	153	HIS	-	expression tag	UNP Q04089
B	154	HIS	-	expression tag	UNP Q04089
B	155	HIS	-	expression tag	UNP Q04089
B	156	HIS	-	expression tag	UNP Q04089
B	157	HIS	-	expression tag	UNP Q04089
C	150	MET	-	expression tag	UNP Q04089
C	151	GLY	-	expression tag	UNP Q04089
C	152	HIS	-	expression tag	UNP Q04089
C	153	HIS	-	expression tag	UNP Q04089
C	154	HIS	-	expression tag	UNP Q04089
C	155	HIS	-	expression tag	UNP Q04089
C	156	HIS	-	expression tag	UNP Q04089

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	157	HIS	-	expression tag	UNP Q04089

- Molecule 2 is S-ADENOSYL-L-HOMOCYSTEINE (three-letter code: SAH) (formula: $C_{14}H_{20}N_6O_5S$).
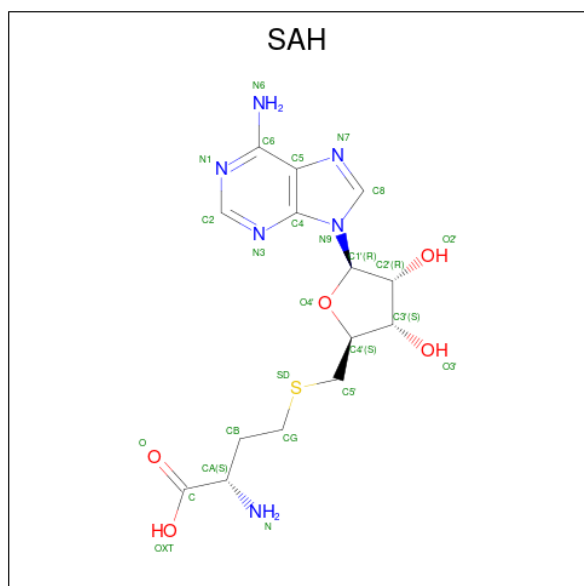
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	S	0	0
			26	14	6	5	1		
2	B	1	Total	C	N	O	S	0	0
			26	14	6	5	1		
2	C	1	Total	C	N	O	S	0	0
			26	14	6	5	1		

- Molecule 3 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	142	Total	O	0	0
			142	142		
3	B	203	Total	O	0	0
			203	203		
3	C	188	Total	O	0	0
			188	188		

3 Residue-property plots ⓘ

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. 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. 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.

Note EDS was not executed.

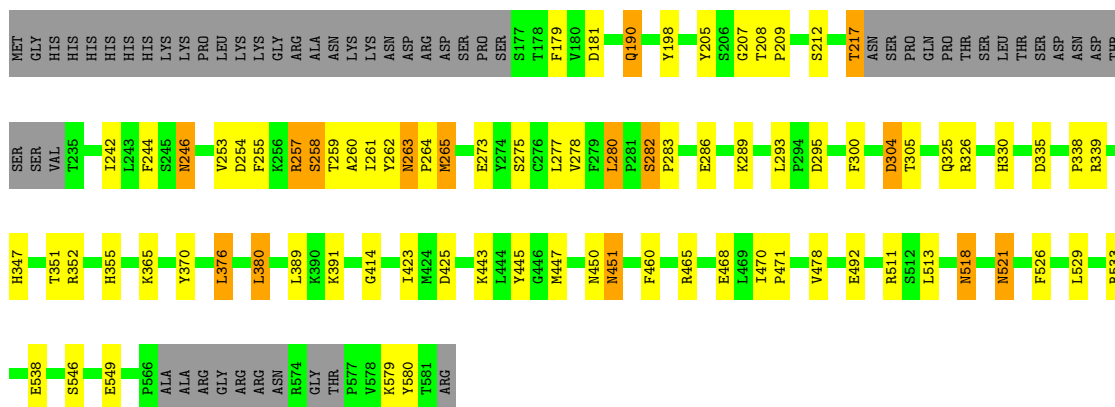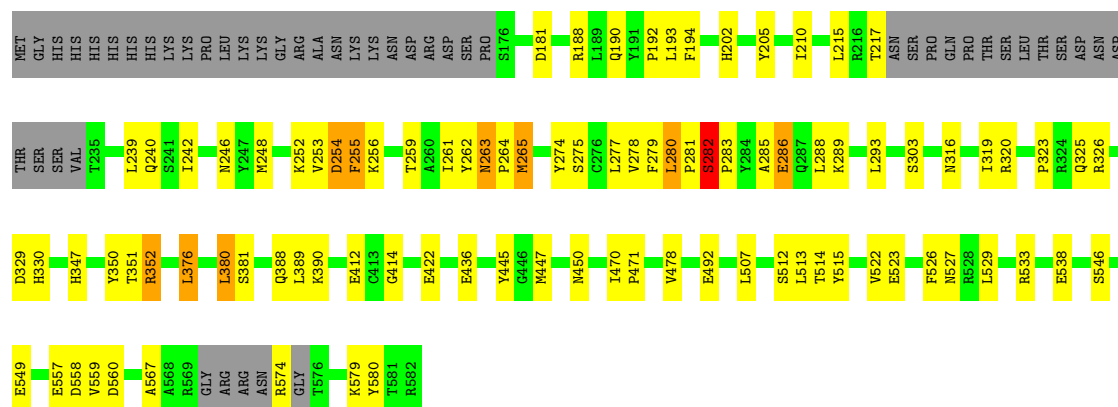• Molecule 1: Histone-lysine N-methyltransferase, H3 lysine-79 specific

Chain A:
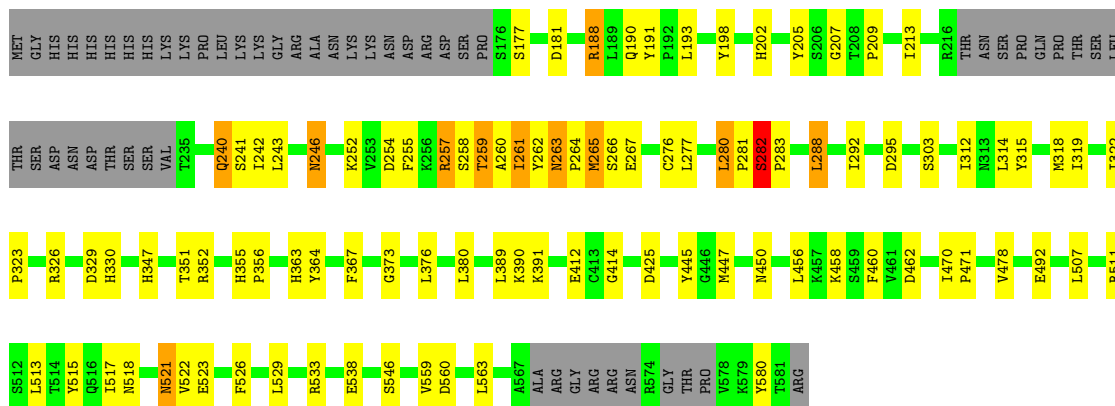
• Molecule 1: Histone-lysine N-methyltransferase, H3 lysine-79 specific

Chain B:

• Molecule 1: Histone-lysine N-methyltransferase, H3 lysine-79 specific

Chain C:

MET GLY HIS HIS HIS HIS HIS HIS LYS LYS LYS PRO LEU LYS LYS GLY ARG ALA ASN LYS LYS ASN ASP ARG ASP SER PRO **S176** D181 **R188 L189 Q190 Y191** P192 L193 **F194** H202 Y205 I210 L215 **R216 T217** ASN SER PRO GLN PRO THR SER LEU THR SER ASP ASN ASP

THR SER SER VAL **T235** L239 Q240 **S241** I242 **N246 Y247** M248 K252 V253 **D254 F255** K256 T259 **A260** I261 Y262 **N263** P264 **M265** Y274 **S275 C276** L277 V278 F279 **L280** P281 **S282 P283 Y284** A285 **E286 Q287** L288 **K289** L293 **S303** N316 I319 R320 P323 **R324** Q325 **R326**

D329 H330 **H347** Y350 T351 **R352** L376 **L380** S381 **Q388** L389 K390 E412 **C413 G414** E422 E436 Y445 **G446** M447 N450 I470 **P471** V478 E492 L507 **S512** L513 **T514 Y515** V522 **E523** F526 N527 **R528** L529 R533 **E538** S546

E549 **E557** D558 V559 **D560** A567 **A568 R569** GLY ARG ARG ASN **R574** GLY **T576** K579 **Y580 T581 R582**

4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	73.07Å 146.28Å 75.41Å 90.00° 97.68° 90.00°	Depositor
Resolution (Å)	19.90 – 2.20	Depositor
% Data completeness (in resolution range)	98.9 (19.90-2.20)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	0.05	Depositor
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.213 , 0.246	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	9909	wwPDB-VP
Average B, all atoms (Å^2)	46.0	wwPDB-VP

5   Model quality ⓘ

5.1   Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: SAH

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	0.40	1/3144 (0.0%)	0.67	3/4245 (0.1%)
1	B	0.46	1/3148 (0.0%)	0.70	4/4251 (0.1%)
1	C	0.45	1/3194 (0.0%)	0.73	6/4312 (0.1%)
All	All	0.44	3/9486 (0.0%)	0.70	13/12808 (0.1%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	283	PRO	CA-C	-6.02	1.40	1.52
1	A	283	PRO	CA-C	-5.06	1.42	1.52
1	C	283	PRO	CA-C	-5.01	1.42	1.52

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	282	SER	C-N-CD	-13.69	90.49	120.60
1	A	282	SER	C-N-CD	-13.52	90.85	120.60
1	B	282	SER	C-N-CD	-8.76	101.34	120.60
1	C	255	PHE	CB-CG-CD2	-7.37	115.64	120.80
1	A	282	SER	C-N-CA	6.31	148.49	122.00

There are no chirality outliers.

There are no planarity outliers.

5.2   Too-close contacts ⓘ

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within

the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	3082	0	3064	87	0
1	B	3085	0	3069	68	0
1	C	3131	0	3116	84	0
2	A	26	0	19	2	0
2	B	26	0	19	0	0
2	C	26	0	19	1	0
3	A	142	0	0	0	0
3	B	203	0	0	0	0
3	C	188	0	0	0	0
All	All	9909	0	9306	234	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 13.

The worst 5 of 234 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:188:ARG:HG2	1:A:188:ARG:HH11	1.17	1.04
1:C:286:GLU:N	1:C:286:GLU:OE1	1.90	1.04
1:C:253:VAL:HA	1:C:265:MET:HE2	1.38	1.02
1:B:217:THR:HG22	1:B:261:ILE:HD12	1.48	0.94
1:B:253:VAL:HA	1:B:265:MET:HE2	1.51	0.93

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	372/433 (86%)	358 (96%)	11 (3%)	3 (1%)	19	19
1	B	372/433 (86%)	358 (96%)	11 (3%)	3 (1%)	19	19

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	378/433 (87%)	368 (97%)	10 (3%)	0	100	100
All	All	1122/1299 (86%)	1084 (97%)	32 (3%)	6 (0%)	29	31

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	257	ARG
1	A	259	THR
1	A	261	ILE
1	B	260	ALA
1	B	258	SER

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	350/398 (88%)	331 (95%)	19 (5%)	22	26
1	B	351/398 (88%)	332 (95%)	19 (5%)	22	26
1	C	355/398 (89%)	341 (96%)	14 (4%)	32	41
All	All	1056/1194 (88%)	1004 (95%)	52 (5%)	25	31

5 of 52 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	335	ASP
1	B	513	LEU
1	C	436	GLU
1	B	339	ARG
1	B	389	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 37 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	240	GLN
1	C	450	ASN
1	C	246	ASN
1	C	310	ASN
1	B	202	HIS

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

3 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	SAH	A	801	-	24,28,28	1.67	2 (8%)	25,40,40	1.97	5 (20%)
2	SAH	B	802	-	24,28,28	0.84	1 (4%)	25,40,40	1.04	2 (8%)
2	SAH	C	803	-	24,28,28	1.55	3 (12%)	25,40,40	1.46	3 (12%)

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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	SAH	A	801	-	-	1/11/31/31	0/3/3/3
2	SAH	B	802	-	-	2/11/31/31	0/3/3/3
2	SAH	C	803	-	-	1/11/31/31	0/3/3/3

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	801	SAH	CB-CA	-6.60	1.39	1.53
2	C	803	SAH	CB-CA	-5.32	1.41	1.53
2	C	803	SAH	O-C	3.34	1.32	1.22
2	A	801	SAH	O-C	3.02	1.31	1.22
2	B	802	SAH	O-C	2.97	1.31	1.22

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	801	SAH	CB-CA-N	-5.97	94.51	110.17
2	A	801	SAH	CB-CA-C	4.87	121.89	110.30
2	C	803	SAH	CB-CA-N	-4.02	99.63	110.17
2	C	803	SAH	CB-CA-C	3.43	118.46	110.30
2	A	801	SAH	C5'-SD-CG	-2.97	93.37	102.27

There are no chirality outliers.

All (4) torsion outliers are listed below:

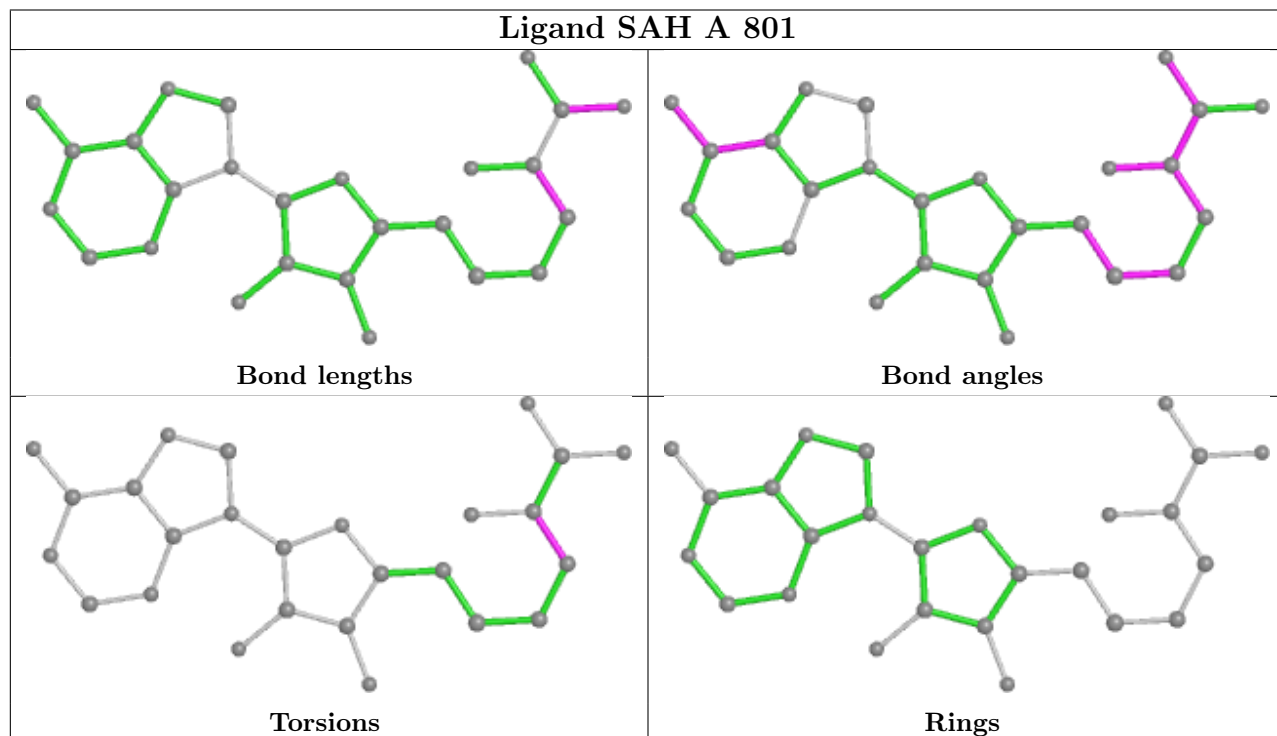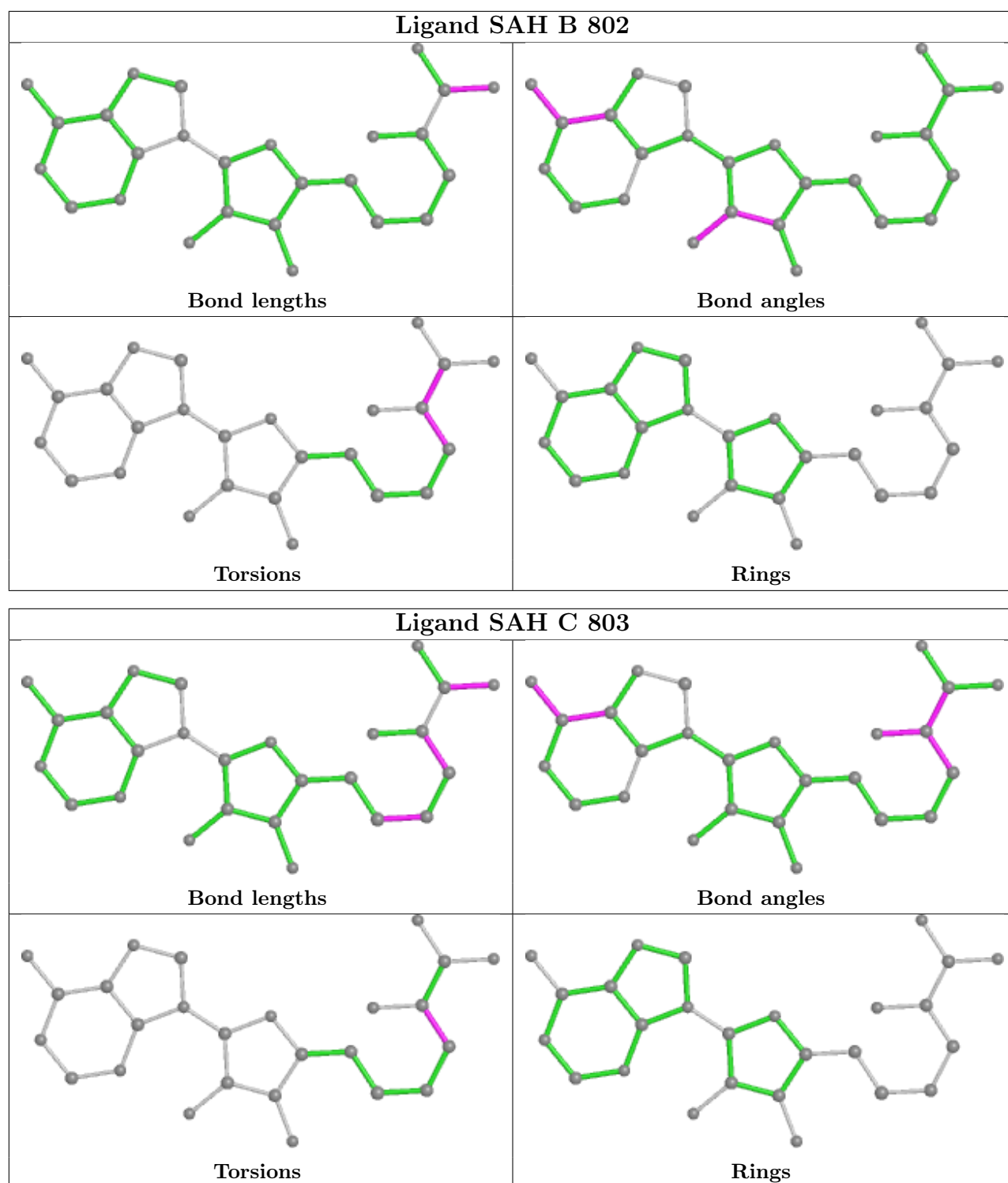| Mol	Chain	Res	Type	Atoms
2	B	802	SAH	OXT-C-CA-N
2	A	801	SAH	C-CA-CB-CG
2	C	803	SAH	C-CA-CB-CG
2	B	802	SAH	C-CA-CB-CG

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	801	SAH	2	0
2	C	803	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 A 801

Bond lengths

Bond angles

Torsions

Rings

Ligand SAH B 802

Bond lengths

Bond angles

Torsions

Rings

Ligand SAH C 803

Bond lengths

Bond angles

Torsions

Rings

5.7   Other polymers ⓘ

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 ⓘ

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates ⓘ

EDS was not executed - this section is therefore empty.

6.4 Ligands ⓘ

EDS was not executed - this section is therefore empty.

6.5 Other polymers ⓘ

EDS was not executed - this section is therefore empty.