



# wwPDB X-ray Structure Validation Summary Report

Feb 28, 2014 – 10:58 PM GMT

PDB ID : 3HTX  
Title : Crystal structure of small RNA methyltransferase HEN1  
Authors : Huang, Y.; Ji, L.-J.; Huang, Q.-C.; Vassylyev, D.G.; Chen, X.-M.; Ma, J.-B.  
Deposited on : 2009-06-12  
Resolution : 3.10 Å(reported)

This is a wwPDB validation summary report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

---

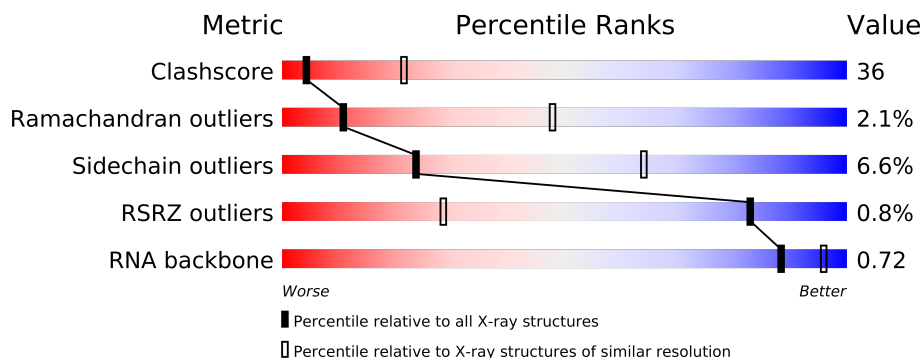
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22683

# 1 Overall quality at a glance

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))
Clashscore	79885	1078 (3.16-3.04)
Ramachandran outliers	78287	1044 (3.16-3.04)
Sidechain outliers	78261	1044 (3.16-3.04)
RSRZ outliers	66119	1008 (3.18-3.02)
RNA backbone	1838	1047 (3.60-2.60)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	950	
1	D	950	
2	B	22	
2	E	22	
3	C	22	
3	F	22	

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 14269 atoms, of which 0 are hydrogen and 0 are deuterium.

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 HEN1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	779	Total	C	N	O	S	0	0	0
			6093	3847	1046	1170	30			
1	D	795	Total	C	N	O	S	0	0	0
			6217	3928	1065	1194	30			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-7	MET	-	EXPRESSION TAG	UNP Q945R3
A	-6	GLY	-	EXPRESSION TAG	UNP Q945R3
A	-5	HIS	-	EXPRESSION TAG	UNP Q945R3
A	-4	HIS	-	EXPRESSION TAG	UNP Q945R3
A	-3	HIS	-	EXPRESSION TAG	UNP Q945R3
A	-2	HIS	-	EXPRESSION TAG	UNP Q945R3
A	-1	HIS	-	EXPRESSION TAG	UNP Q945R3
A	0	HIS	-	EXPRESSION TAG	UNP Q945R3
A	604	PRO	LEU	ENGINEERED	UNP Q945R3
A	640	LYS	ARG	ENGINEERED	UNP Q945R3
D	-7	MET	-	EXPRESSION TAG	UNP Q945R3
D	-6	GLY	-	EXPRESSION TAG	UNP Q945R3
D	-5	HIS	-	EXPRESSION TAG	UNP Q945R3
D	-4	HIS	-	EXPRESSION TAG	UNP Q945R3
D	-3	HIS	-	EXPRESSION TAG	UNP Q945R3
D	-2	HIS	-	EXPRESSION TAG	UNP Q945R3
D	-1	HIS	-	EXPRESSION TAG	UNP Q945R3
D	0	HIS	-	EXPRESSION TAG	UNP Q945R3
D	604	PRO	LEU	ENGINEERED	UNP Q945R3
D	640	LYS	ARG	ENGINEERED	UNP Q945R3

- Molecule 2 is a RNA chain called 5'-R(\*GP\*AP\*UP\*UP\*UP\*CP\*UP\*CP\*UP\*CP\*UP\*GP\*CP\*AP\*AP\*GP\*CP\*GP\*AP\*AP\*AP\*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	22	Total	C	N	O	P	0	0	0
			464	209	82	152	21			
2	E	22	Total	C	N	O	P	0	0	0
			464	209	82	152	21			

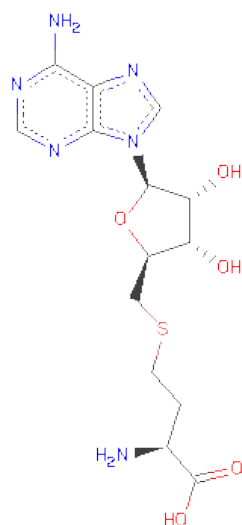
- Molecule 3 is a RNA chain called 5'-R(P\*UP\*UP\*CP\*GP\*CP\*UP\*UP\*GP\*CP\*AP\*GP\*AP\*GP\*AP\*GP\*AP\*AP\*UP\*CP\*AP\*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	22	Total	C	N	O	P	0	0	0
			470	210	85	153	22			
3	F	22	Total	C	N	O	P	0	0	0
			470	210	85	153	22			

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Mg	0	0
			1	1		
4	D	1	Total	Mg	0	0
			1	1		

- Molecule 5 is S-ADENOSYL-L-HOMOCYSTEINE (three-letter code: SAH) (formula: C<sub>14</sub>H<sub>20</sub>N<sub>6</sub>O<sub>5</sub>S).

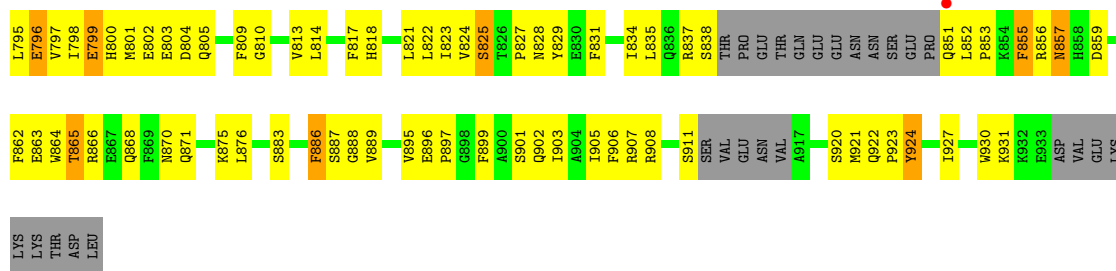


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	A	1	Total	C	N	O	S	0	0
			26	14	6	5	1		
5	D	1	Total	C	N	O	S	0	0
			26	14	6	5	1		

- Molecule 6 is water.

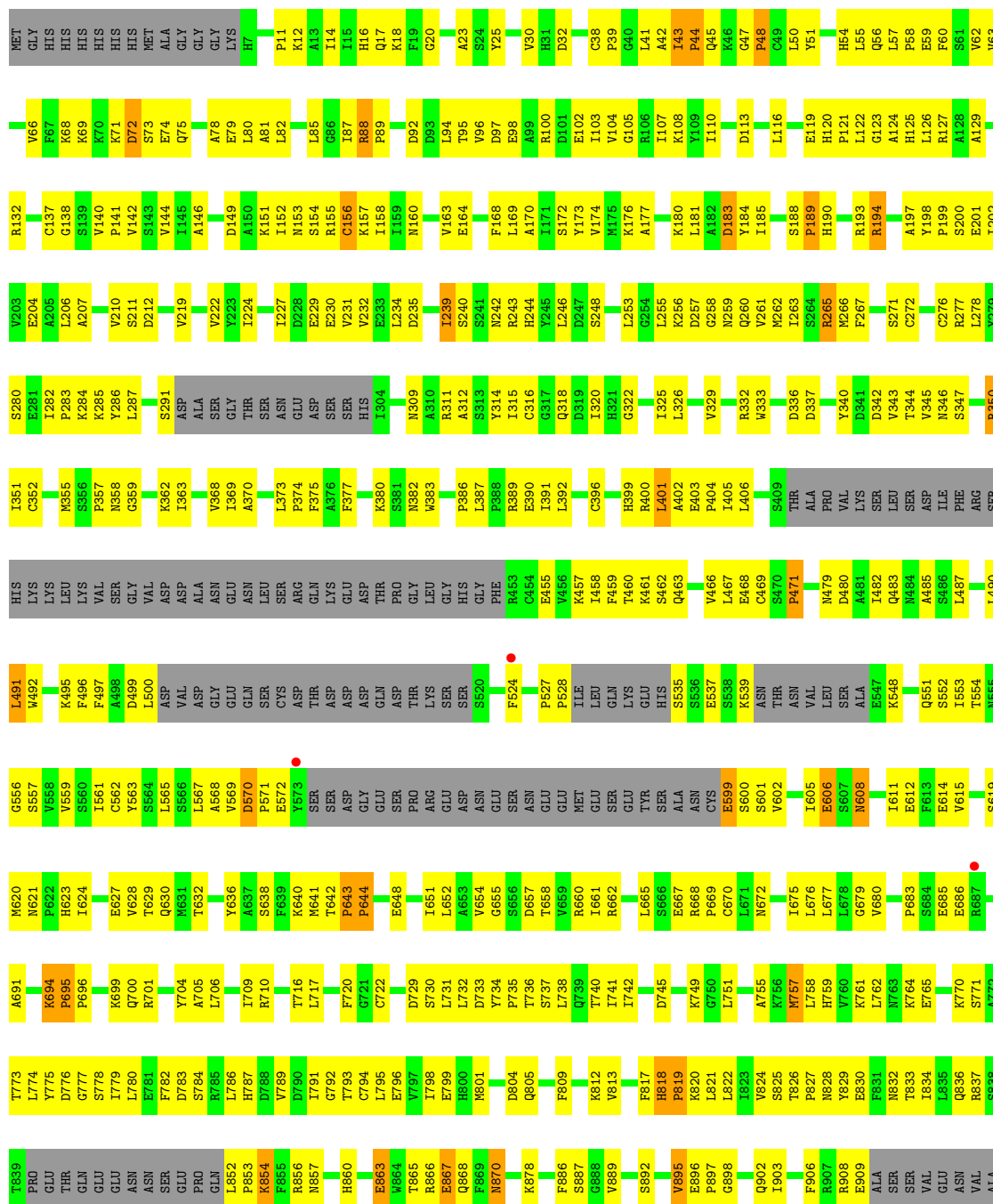
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	14	Total	O	0	0
			14	14		
6	B	2	Total	O	0	0
			2	2		
6	C	2	Total	O	0	0
			2	2		
6	D	16	Total	O	0	0
			16	16		
6	E	1	Total	O	0	0
			1	1		
6	F	2	Total	O	0	0
			2	2		





• Molecule 1: HEN1

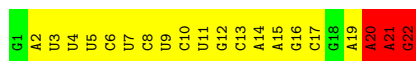
Chain D:





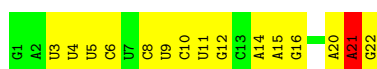
- Molecule 2: 5'-R(\*GP\*AP\*UP\*UP\*UP\*CP\*UP\*CP\*UP\*CP\*UP\*GP\*CP\*AP\*AP\*GP\*CP\*GP\*AP\*AP\*AP\*G)-3'

Chain B:



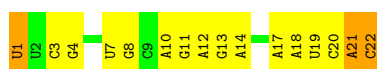
- Molecule 2: 5'-R(\*GP\*AP\*UP\*UP\*UP\*CP\*UP\*CP\*UP\*CP\*UP\*GP\*CP\*AP\*AP\*GP\*CP\*GP\*AP\*AP\*AP\*G)-3'

Chain E:



- Molecule 3: 5'-R(P\*UP\*UP\*CP\*GP\*CP\*UP\*UP\*GP\*CP\*AP\*GP\*AP\*GP\*AP\*GP\*AP\*AP\*AP\*UP\*CP\*AP\*C)-3'

Chain C:



- Molecule 3: 5'-R(P\*UP\*UP\*CP\*GP\*CP\*UP\*UP\*GP\*CP\*AP\*GP\*AP\*GP\*AP\*GP\*AP\*AP\*AP\*UP\*CP\*AP\*C)-3'

Chain F:





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	97.00Å 124.35Å 101.38Å 90.00° 93.47° 90.00°	Depositor
Resolution (Å)	20.00 – 3.10 19.99 – 3.10	Depositor EDS
% Data completeness (in resolution range)	92.8 (20.00-3.10) 78.0 (19.99-3.10)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.59 (at 3.09Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.260 , 0.288 0.248 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	DCC
Wilson B-factor (Å <sup>2</sup> )	72.0	Xtriage
Anisotropy	0.237	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.24 , 8.9	EDS
Estimated twinning fraction	0.054 for l,-k,h	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.42$ , $\langle L^2 \rangle = 0.24$	Xtriage
Outliers	0 of 40951 reflections	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	14269	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	106.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.41% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

## 5 Model quality ⓘ

### 5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: MG, 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	0.51	0/6214	0.75	4/8395 (0.0%)
1	D	0.52	0/6343	0.76	5/8573 (0.1%)
2	B	0.46	0/518	0.92	3/805 (0.4%)
2	E	0.46	0/518	0.78	1/805 (0.1%)
3	C	0.55	1/525 (0.2%)	0.69	0/814
3	F	0.58	1/525 (0.2%)	0.84	1/814 (0.1%)
All	All	0.51	2/14643 (0.0%)	0.77	14/20206 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	1	U	OP3-P	-6.77	1.53	1.61
3	F	1	U	OP3-P	-6.57	1.53	1.61

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	20	A	C2'-C3'-O3'	9.92	131.33	109.50
3	F	19	U	C2'-C3'-O3'	8.29	127.75	109.50
1	A	789	VAL	CB-CA-C	-7.54	97.08	111.40
1	D	20	GLY	N-CA-C	-7.19	95.12	113.10
1	D	818	HIS	C-N-CD	-7.07	105.05	120.60

There are no chirality outliers.

There are no planarity outliers.

## 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	6093	0	6031	434	0
1	D	6217	0	6139	493	0
2	B	464	0	236	35	0
2	E	464	0	236	21	0
3	C	470	0	238	20	0
3	F	470	0	238	17	0
4	A	1	0	0	0	0
4	D	1	0	0	0	0
5	A	26	0	19	2	0
5	D	26	0	19	1	0
6	A	14	0	0	5	0
6	B	2	0	0	2	0
6	C	2	0	0	0	0
6	D	16	0	0	2	0
6	E	1	0	0	0	0
6	F	2	0	0	1	0
All	All	14269	0	13156	987	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 36.

The worst 5 of 987 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:802:GLU:HB2	1:A:805:GLN:HG3	1.21	1.20
3:F:21:A:H2'	6:F:226:HOH:O	1.45	1.12
1:A:262:MET:HE3	1:A:263:ILE:H	1.16	1.07
1:D:776:ASP:HB2	1:D:927:ILE:HD11	1.34	1.07
1:A:45:GLN:HE21	1:A:752:ALA:HB1	1.18	1.07

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	761/950 (80%)	671 (88%)	72 (10%)	18 (2%)	9	43
1	D	777/950 (82%)	692 (89%)	70 (9%)	15 (2%)	12	51
All	All	1538/1900 (81%)	1363 (89%)	142 (9%)	33 (2%)	11	48

5 of 33 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	44	PRO
1	A	59	GLU
1	A	284	LYS
1	D	48	PRO
1	D	854	LYS

### 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	676/828 (82%)	621 (92%)	55 (8%)	17	53
1	D	689/828 (83%)	654 (95%)	35 (5%)	33	74
All	All	1365/1656 (82%)	1275 (93%)	90 (7%)	24	64

5 of 90 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	710	ARG
1	A	825	SER
1	D	804	ASP
1	A	715	SER

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	776	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 28 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	739	GLN
1	A	870	ASN
1	D	621	ASN
1	A	857	ASN
1	A	868	GLN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	B	21/22 (95%)	3 (14%)	2 (9%)
2	E	21/22 (95%)	2 (9%)	0
3	C	21/22 (95%)	2 (9%)	0
3	F	21/22 (95%)	4 (19%)	1 (4%)
All	All	84/88 (95%)	11 (13%)	3 (3%)

5 of 11 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	B	20	A
2	B	21	A
2	B	22	G
3	C	21	A
3	C	22	C

All (3) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	B	20	A
2	B	21	A
3	F	19	U

## 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 carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 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	$\# Z  > 2$
5	SAH	A	951	-	28,28,28	0.95	2 (7%)	40,40,40	1.09	4 (10%)
5	SAH	D	951	-	28,28,28	0.90	1 (3%)	40,40,40	0.91	2 (5%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	SAH	A	951	-	-	0/15/31/31	0/1/3/3
5	SAH	D	951	-	-	0/15/31/31	0/1/3/3

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	D	951	SAH	C8-N9	2.17	1.39	1.36
5	A	951	SAH	O-C	2.07	1.29	1.22
5	A	951	SAH	C8-N9	2.04	1.39	1.36

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	951	SAH	O4'-C1'-C2'	-3.62	101.23	106.77
5	A	951	SAH	C8-N9-C4	-2.64	104.89	106.90
5	A	951	SAH	O4'-C1'-N9	-2.58	106.04	108.44
5	D	951	SAH	O4'-C1'-N9	-2.15	106.44	108.44

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	D	951	SAH	O4'-C1'-C2'	-2.05	103.63	106.77

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 5.7 Other polymers ⓘ

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<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	A	779/950 (82%)	-0.10	11 (1%) 72 17	60, 96, 160, 193	0
1	D	795/950 (83%)	-0.12	3 (0%) 90 45	55, 98, 154, 188	0
2	B	22/22 (100%)	-0.72	0 100 100	91, 120, 146, 151	0
2	E	22/22 (100%)	-0.71	0 100 100	99, 114, 123, 130	0
3	C	22/22 (100%)	-0.65	0 100 100	89, 137, 168, 199	0
3	F	22/22 (100%)	-0.66	0 100 100	91, 125, 144, 186	0
All	All	1662/1988 (83%)	-0.14	14 (0%) 83 28	55, 98, 159, 199	0

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	573	TYR	4.6
1	A	7	HIS	3.3
1	A	851	GLN	3.2
1	A	89	PRO	2.9
1	A	32	ASP	2.8

### 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

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, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
5	SAH	A	951	26/26	0.21	0.48	62,68,85,86	0
5	SAH	D	951	26/26	0.14	-0.85	70,74,86,87	0
4	MG	D	950	1/1	0.10	-1.78	36,36,36,36	0
4	MG	A	950	1/1	0.04	-4.21	60,60,60,60	0

## 6.5 Other polymers

There are no such residues in this entry.