wwPDB X-ray Structure Validation Summary Report (i)

Aug 2, 2023 – 10:19 AM EDT

PDB ID	:	10MH
Title	:	TERNARY STRUCTURE OF HHAI METHYLTRANSFERASE WITH ADOHCY AND HEMIMETHYLATED DNA CONTAINING 5,6-DIHYDR O-5-AZACYTOSINE AT THE TARGET
Authors	:	Sheikhnejad, G.; Brank, A.; Christman, J.K.; Goddard, A.; Alvarez, E.; Ford Junior, H.; Marquez, V.E.; Marasco, C.J.; Sufrin, J.R.; O'Gara, M.; Cheng, X.
Deposited on	:	1998-08-10
Resolution	:	2.55 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

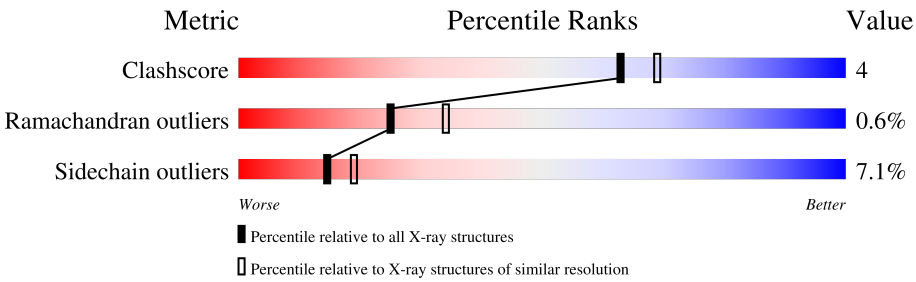
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.34

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 2.55 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	141614	1332 (2.56-2.52)
Ramachandran outliers	138981	1315 (2.56-2.52)
Sidechain outliers	138945	1315 (2.56-2.52)

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	B	12	75% 25%
2	C	12	42% 50% 8%
3	A	327	86% 13% •

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 3254 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 DNA chain called DNA (5'-D(P*CP*CP*AP*TP*GP*(5CM)P*GP*CP*TP*GP*AP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	B	12	244	116	44	72	12	0	0	0

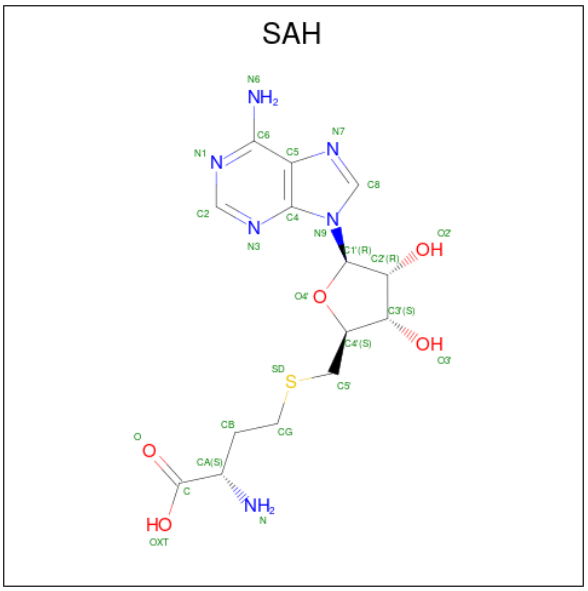
- Molecule 2 is a DNA chain called DNA (5'-D(P*GP*TP*CP*AP*GP*5NCP*GP*CP*AP*TP*GP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	C	12	249	116	49	72	12	0	0	0

- Molecule 3 is a protein called PROTEIN (CYTOSINE-SPECIFIC METHYLTRANSFERASE HHAI).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	A	327	2606	1662	444	487	13	0	0	0

- Molecule 4 is S-ADENOSYL-L-HOMOCYSTEINE (three-letter code: SAH) (formula: $C_{14}H_{20}N_6O_5S$).



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

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	B	6	6	6	0	0
			Total	O		
5	C	9	9	9	0	0
			Total	O		
5	A	114	114	114	0	0

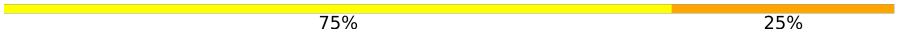
3 Residue-property plots (i)

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: DNA (5'-D(P*CP*CP*AP*TP*GP*(5CM)P*GP*CP*TP*GP*AP*C)-3')

Chain B:

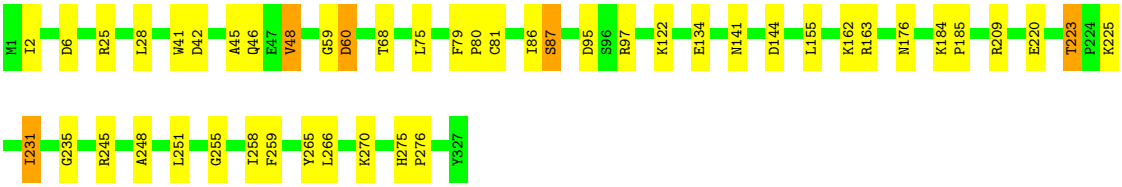


• Molecule 2: DNA (5'-D(P*GP*TP*CP*AP*GP*5NCP*GP*CP*AP*TP*GP*G)-3')

Chain C:

• Molecule 3: PROTEIN (CYTOSINE-SPECIFIC METHYLTRANSFERASE HHAI)

Chain A:



4 Data and refinement statistics (i)

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

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants a, b, c, α, β, γ	99.86Å 99.86Å 325.20Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	29.30 – 2.55	Depositor
% Data completeness (in resolution range)	76.6 (29.30-2.55)	Depositor
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	X-PLOR 3.1	Depositor
R, R_{free}	0.205 , (Not available)	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	3254	wwPDB-VP
Average B, all atoms (Å^2)	17.0	wwPDB-VP

5 Model quality (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: SAH, 5NC, 5CM

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	B	3.45	26/249 (10.4%)	4.66	85/379 (22.4%)
2	C	3.70	32/257 (12.5%)	3.94	62/393 (15.8%)
3	A	0.60	1/2661 (0.0%)	0.79	1/3586 (0.0%)
All	All	1.53	59/3167 (1.9%)	1.95	148/4358 (3.4%)

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	B	0	1
2	C	0	3
3	A	0	1
All	All	0	5

The worst 5 of 59 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	422	DG	C2'-C1'	-22.18	1.30	1.52
2	C	422	DG	O4'-C1'	21.37	1.67	1.42
1	B	412	DA	O3'-P	16.77	1.81	1.61
2	C	422	DG	P-O5'	13.05	1.72	1.59
2	C	429	DC	C5'-C4'	12.53	1.65	1.51

The worst 5 of 148 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	406	DG	O4'-C1'-C2'	-20.20	89.74	105.90
1	B	403	DC	O4'-C4'-C3'	-19.17	94.50	106.00

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	409	DC	O4'-C1'-C2'	-19.17	90.56	105.90
1	B	402	DC	O4'-C4'-C3'	-18.21	95.07	106.00
2	C	433	DG	O4'-C1'-C2'	-17.95	91.54	105.90

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	A	265	TYR	Sidechain
1	B	409	DC	Sidechain
2	C	428	DG	Sidechain
2	C	432	DG	Sidechain
2	C	433	DG	Sidechain

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	B	244	0	137	1	0
2	C	249	0	134	8	0
3	A	2606	0	2587	22	0
4	A	26	0	19	0	0
5	A	114	0	0	3	0
5	B	6	0	0	0	0
5	C	9	0	0	2	0
All	All	3254	0	2877	26	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 4.

The worst 5 of 26 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)
2:C:422:DG:O4'	2:C:422:DG:C1'	1.67	1.13
3:A:275:HIS:NE2	5:A:470:HOH:O	2.19	0.63

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:A:258:ILE:HG23	3:A:259:PHE:CD2	2.39	0.57
3:A:223:THR:HB	3:A:225:LYS:HG2	1.87	0.56
5:C:486:HOH:O	3:A:162:LYS:HD3	2.05	0.55

There are no symmetry-related clashes.

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
3	A	325/327 (99%)	307 (94%)	16 (5%)	2 (1%)	25 34

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	60	ASP
3	A	144	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
3	A	283/283 (100%)	263 (93%)	20 (7%)	14 19

5 of 20 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	231	ILE
3	A	266	LEU
3	A	276	PRO
3	A	270	LYS
3	A	68	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	141	ASN
3	A	237	GLN
3	A	297	GLN
3	A	65	ASN
3	A	63	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)

2 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	$\#\|Z\| > 2$
1	5CM	B	407	1,2	17,21,22	1.38	3 (17%)	24,30,33	3.16	13 (54%)
2	5NC	C	427	2	17,20,21	3.45	6 (35%)	24,28,31	1.97	8 (33%)

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
1	5CM	B	407	1,2	-	4/7/21/22	0/2/2/2
2	5NC	C	427	2	-	4/7/21/22	0/2/2/2

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	427	5NC	C6-N5	11.84	1.52	1.30
2	C	427	5NC	C6-N1	4.25	1.46	1.36
2	C	427	5NC	C1'-N1	3.59	1.57	1.48
1	B	407	5CM	C5'-C4'	3.24	1.61	1.51
2	C	427	5NC	C2'-C1'	2.39	1.59	1.52

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	407	5CM	C4'-O4'-C1'	-7.30	91.80	109.45
1	B	407	5CM	O4'-C4'-C3'	-5.14	93.67	105.67
1	B	407	5CM	O4'-C1'-C2'	-4.50	97.74	106.25
1	B	407	5CM	O5'-C5'-C4'	4.31	123.66	108.99
1	B	407	5CM	C2'-C1'-N1	4.17	123.36	113.77

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	B	407	5CM	O4'-C4'-C5'-O5'
1	B	407	5CM	C3'-C4'-C5'-O5'
2	C	427	5NC	C2'-C1'-N1-C6
2	C	427	5NC	O4'-C1'-N1-C6
2	C	427	5NC	C2'-C1'-N1-C2

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	C	427	5NC	2	0

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

1 ligand is 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	$\#\|Z\| > 2$
4	SAH	A	328	-	24,28,28	1.18	2 (8%)	25,40,40	1.90	6 (24%)

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
4	SAH	A	328	-	-	1/11/31/31	0/3/3/3

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	328	SAH	O-C	2.66	1.30	1.22
4	A	328	SAH	C5'-C4'	2.06	1.60	1.52

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	328	SAH	O4'-C1'-C2'	-6.19	97.88	106.93
4	A	328	SAH	O4'-C4'-C3'	-3.12	98.94	105.11
4	A	328	SAH	OXT-C-O	-2.59	118.21	124.09
4	A	328	SAH	C4-C5-N7	2.52	112.03	109.40
4	A	328	SAH	C1'-N9-C4	2.41	130.87	126.64

There are no chirality outliers.

All (1) torsion outliers are listed below:

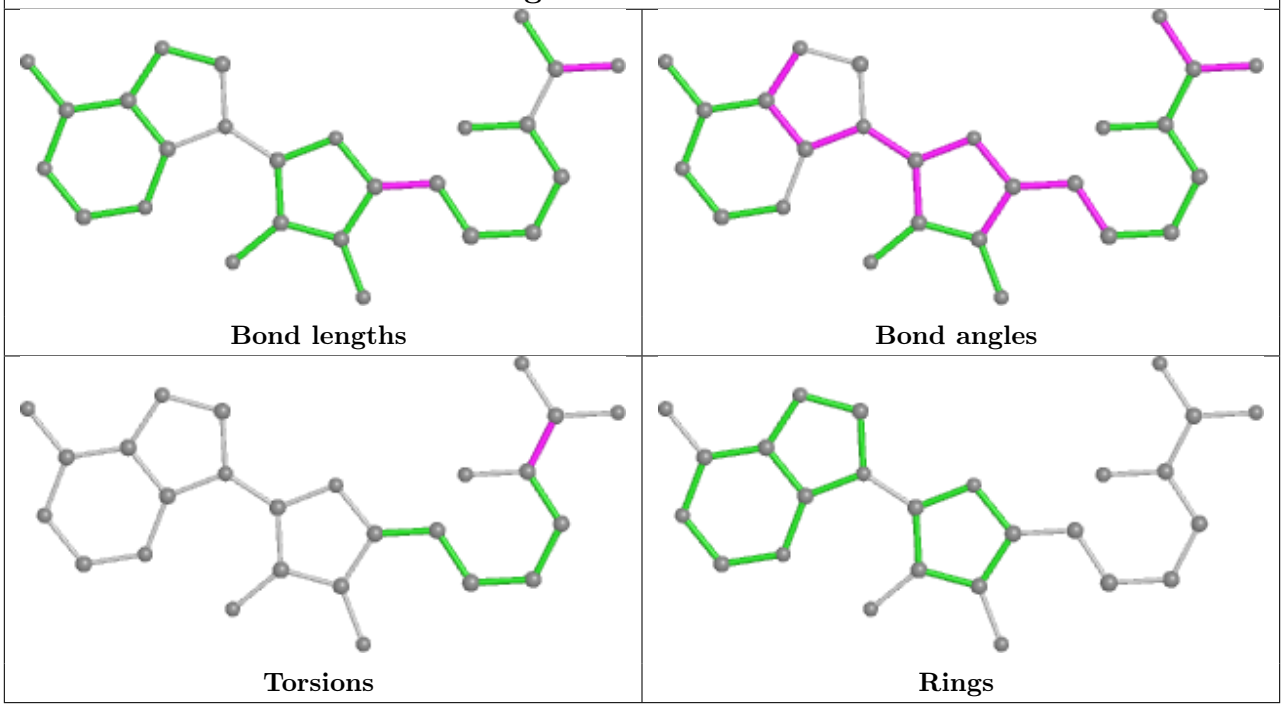
Mol	Chain	Res	Type	Atoms
4	A	328	SAH	O-C-CA-CB

There are no ring outliers.

No monomer is involved in short contacts.

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 328



5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	B	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	B	412:DA	O3'	413:DC	P	1.81

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates (i)

EDS was not executed - this section is therefore empty.

6.4 Ligands (i)

EDS was not executed - this section is therefore empty.

6.5 Other polymers (i)

EDS was not executed - this section is therefore empty.