



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 2, 2016 – 12:00 AM GMT

PDB ID : 7MHT  
Title : CYTOSINE-SPECIFIC METHYLTRANSFERASE HHAI/DNA COMPLEX  
Authors : O'Gara, M.; Horton, J.R.; Roberts, R.J.; Cheng, X.  
Deposited on : 1998-08-05  
Resolution : 2.87 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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/validation/2016/XrayValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity : 4.02b-467  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : **NOT EXECUTED**  
EDS : **NOT EXECUTED**  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865

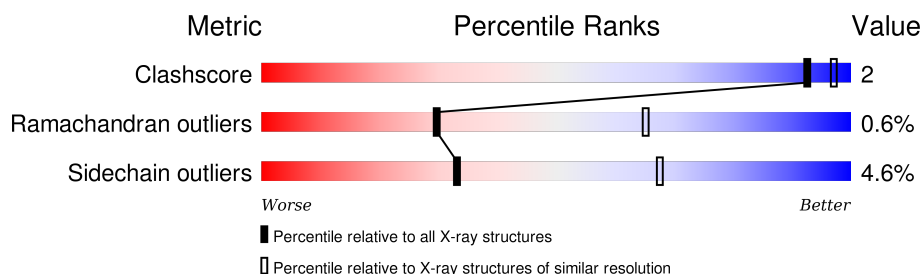
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.87 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	102246	2202 (2.90-2.86)
Ramachandran outliers	100387	2149 (2.90-2.86)
Sidechain outliers	100360	2152 (2.90-2.86)

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. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	C	12	<div> <div style="width: 92%;"></div> <div style="width: 8%;"></div> </div> <div>92% 8%</div>
2	D	12	<div> <div style="width: 67%;"></div> <div style="width: 25%;"></div> <div style="width: 8%;"></div> </div> <div>67% 25% 8%</div>
3	A	327	<div> <div style="width: 90%;"></div> <div style="width: 9%;"></div> </div> <div>90% 9% .</div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 3265 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 5'-D(P\*GP\*TP\*CP\*AP\*GP\*CP\*GP\*CP\*AP\*TP\*GP\*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	12	Total	C	N	O	P	0	0	0
			249	117	48	72	12			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	12	Total	C	N	O	P	0	0	0
			245	116	46	71	12			

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

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

- Molecule 4 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
4	A	1	Total	C	N	O	S	0	0
			26	14	6	5	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	116	Total O 116 116	0	0
5	C	15	Total O 15 15	0	0
5	D	8	Total O 8 8	0	0

### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

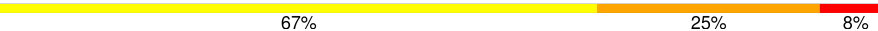
Note EDS was not executed.

- Molecule 1: 5'-D(P\*GP\*TP\*CP\*AP\*GP\*CP\*GP\*CP\*AP\*TP\*GP\*G)-3'

Chain C: 



- Molecule 2: 5'-D(P\*CP\*CP\*AP\*TP\*GP\*AP\*GP\*CP\*TP\*GP\*AP\*C)-3'

Chain D: 



- Molecule 3: CYTOSINE-SPECIFIC METHYLTRANSFERASE HHAI

Chain A: 





## 4 Data and refinement statistics

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

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	99.86 Å 99.86 Å 325.20 Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	23.00 – 2.87	Depositor
% Data completeness (in resolution range)	90.0 (23.00-2.87)	Depositor
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
Refinement program	X-PLOR 3.1	Depositor
R, $R_{free}$	0.156 , (Not available)	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	3265	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	13.0	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

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	C	2.27	9/279 (3.2%)	3.33	47/429 (11.0%)
2	D	2.47	13/274 (4.7%)	3.33	51/420 (12.1%)
3	A	0.46	0/2661	0.70	0/3586
All	All	1.07	22/3214 (0.7%)	1.59	98/4435 (2.2%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
2	D	0	2

All (22) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	422	DC	C5'-C4'	11.26	1.63	1.51
1	C	409	DC	N1-C6	-8.69	1.31	1.37
2	D	423	DC	P-O5'	7.82	1.67	1.59
2	D	428	DG	N9-C4	7.72	1.44	1.38
1	C	408	DG	N7-C5	6.98	1.43	1.39
2	D	422	DC	P-O5'	6.72	1.66	1.59
1	C	407	DC	N1-C6	6.58	1.41	1.37
2	D	433	DC	P-O5'	6.45	1.66	1.59
2	D	423	DC	C5'-C4'	6.24	1.58	1.51
1	C	403	DT	N1-C2	-6.13	1.33	1.38
2	D	426	DG	C5'-C4'	6.01	1.57	1.51
1	C	407	DC	O3'-P	6.00	1.68	1.61
1	C	408	DG	C4'-C3'	5.91	1.59	1.53
1	C	407	DC	C5'-C4'	5.50	1.57	1.51
2	D	429	DC	C5'-C4'	5.45	1.57	1.51

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	425	DT	C4'-O4'	-5.45	1.39	1.45
2	D	433	DC	C2'-C1'	5.37	1.57	1.52
2	D	424	DA	C6-N6	-5.31	1.29	1.33
2	D	427	DA	C6-N1	-5.28	1.31	1.35
2	D	427	DA	N9-C4	5.19	1.41	1.37
1	C	412	DG	O3'-P	5.15	1.67	1.61
1	C	402	DG	C6-N1	5.08	1.43	1.39

All (98) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	413	DG	C1'-O4'-C4'	-18.16	91.94	110.10
2	D	427	DA	O4'-C4'-C3'	-14.89	97.07	106.00
1	C	405	DA	O4'-C1'-C2'	-13.30	95.26	105.90
1	C	407	DC	O4'-C1'-C2'	-12.60	95.82	105.90
2	D	433	DC	O4'-C1'-N1	12.25	116.58	108.00
2	D	433	DC	O4'-C1'-C2'	-12.16	96.17	105.90
1	C	412	DG	O4'-C1'-N9	11.24	115.87	108.00
2	D	424	DA	P-O3'-C3'	10.86	132.74	119.70
2	D	430	DT	O4'-C1'-N1	10.74	115.52	108.00
2	D	425	DT	C6-C5-C7	-10.17	116.80	122.90
1	C	411	DT	C6-C5-C7	-9.88	116.97	122.90
2	D	432	DA	O4'-C4'-C3'	-9.78	100.13	106.00
1	C	409	DC	P-O3'-C3'	9.71	131.36	119.70
2	D	423	DC	O4'-C1'-N1	9.65	114.76	108.00
1	C	411	DT	P-O3'-C3'	9.59	131.21	119.70
2	D	431	DG	P-O3'-C3'	9.57	131.19	119.70
1	C	411	DT	O4'-C1'-C2'	-9.31	98.45	105.90
1	C	409	DC	N1-C2-O2	9.06	124.34	118.90
2	D	422	DC	N3-C4-C5	-9.00	118.30	121.90
2	D	429	DC	N1-C2-O2	8.99	124.30	118.90
2	D	429	DC	O4'-C1'-C2'	-8.79	98.87	105.90
2	D	426	DG	N9-C4-C5	8.75	108.90	105.40
2	D	428	DG	O4'-C1'-N9	8.62	114.03	108.00
1	C	413	DG	O4'-C1'-C2'	-8.53	99.08	105.90
2	D	427	DA	C5-C6-N1	-8.37	113.52	117.70
1	C	407	DC	N1-C2-O2	8.33	123.90	118.90
2	D	429	DC	N3-C2-O2	-8.20	116.16	121.90
2	D	428	DG	P-O3'-C3'	8.04	129.35	119.70
1	C	402	DG	N3-C2-N2	-8.03	114.28	119.90
2	D	429	DC	N3-C4-C5	-7.87	118.75	121.90
2	D	423	DC	N1-C2-O2	7.69	123.51	118.90

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	426	DG	N3-C2-N2	-7.69	114.52	119.90
2	D	431	DG	O4'-C1'-N9	7.58	113.31	108.00
1	C	413	DG	O4'-C4'-C3'	-7.53	101.48	106.00
1	C	405	DA	N9-C4-C5	7.39	108.76	105.80
2	D	426	DG	N1-C2-N2	7.31	122.78	116.20
2	D	427	DA	C4'-C3'-C2'	-7.31	96.52	103.10
1	C	410	DA	O4'-C4'-C3'	7.23	110.34	106.00
2	D	427	DA	N1-C2-N3	7.22	132.91	129.30
2	D	431	DG	O4'-C1'-C2'	-7.22	100.12	105.90
2	D	424	DA	O4'-C4'-C3'	-7.21	101.61	104.50
2	D	426	DG	C8-N9-C4	-7.18	103.53	106.40
1	C	404	DC	O4'-C1'-C2'	-7.12	100.20	105.90
2	D	426	DG	C2-N3-C4	7.10	115.45	111.90
2	D	429	DC	P-O3'-C3'	6.92	128.00	119.70
2	D	427	DA	C2-N3-C4	-6.89	107.16	110.60
1	C	403	DT	C1'-O4'-C4'	6.85	116.95	110.10
1	C	403	DT	N3-C4-C5	-6.79	111.13	115.20
1	C	402	DG	N1-C2-N2	6.77	122.30	116.20
2	D	430	DT	O3'-P-O5'	-6.76	91.15	104.00
1	C	405	DA	P-O3'-C3'	6.76	127.81	119.70
1	C	406	DG	O4'-C1'-C2'	-6.70	100.54	105.90
2	D	429	DC	O3'-P-O5'	-6.69	91.28	104.00
1	C	406	DG	C4'-C3'-C2'	-6.69	97.08	103.10
1	C	406	DG	N3-C4-C5	-6.57	125.31	128.60
1	C	403	DT	O4'-C1'-C2'	-6.49	100.71	105.90
1	C	406	DG	N3-C4-N9	6.48	129.89	126.00
1	C	412	DG	C4'-C3'-C2'	-6.43	97.31	103.10
1	C	408	DG	P-O3'-C3'	6.42	127.40	119.70
1	C	411	DT	N1-C2-N3	6.38	118.43	114.60
1	C	413	DG	N1-C6-O6	6.35	123.71	119.90
1	C	406	DG	N3-C2-N2	6.34	124.34	119.90
1	C	402	DG	N1-C6-O6	6.33	123.70	119.90
1	C	411	DT	C4-C5-C6	6.24	121.74	118.00
2	D	425	DT	N1-C2-N3	6.18	118.31	114.60
2	D	426	DG	P-O3'-C3'	6.13	127.05	119.70
2	D	430	DT	O4'-C1'-C2'	-6.12	101.01	105.90
1	C	409	DC	N3-C2-O2	-6.06	117.66	121.90
1	C	407	DC	C4-C5-C6	-6.04	114.38	117.40
2	D	423	DC	P-O5'-C5'	5.95	130.41	120.90
1	C	412	DG	C8-N9-C4	-5.92	104.03	106.40
2	D	425	DT	C4'-C3'-C2'	-5.90	97.79	103.10
1	C	405	DA	C4'-C3'-C2'	-5.86	97.83	103.10

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	403	DT	C4-C5-C6	5.83	121.50	118.00
1	C	409	DC	N3-C4-N4	5.83	122.08	118.00
1	C	403	DT	C4-C5-C7	-5.81	115.51	119.00
2	D	427	DA	C4-C5-C6	5.77	119.89	117.00
1	C	402	DG	C8-N9-C4	-5.66	104.14	106.40
1	C	405	DA	O3'-P-O5'	-5.65	93.27	104.00
1	C	409	DC	N3-C4-C5	-5.64	119.64	121.90
2	D	431	DG	O4'-C4'-C3'	-5.63	102.25	104.50
1	C	404	DC	N1-C2-O2	5.60	122.26	118.90
2	D	425	DT	C4-C5-C6	5.56	121.33	118.00
2	D	433	DC	C4'-C3'-C2'	-5.56	98.10	103.10
1	C	403	DT	N3-C4-O4	5.51	123.21	119.90
2	D	422	DC	N3-C4-N4	5.49	121.85	118.00
1	C	412	DG	O4'-C1'-C2'	-5.49	101.51	105.90
2	D	426	DG	N3-C4-C5	-5.47	125.86	128.60
1	C	404	DC	P-O3'-C3'	5.46	126.25	119.70
1	C	407	DC	C2-N3-C4	5.41	122.61	119.90
2	D	427	DA	P-O3'-C3'	5.40	126.18	119.70
2	D	426	DG	O4'-C1'-N9	5.31	111.71	108.00
2	D	422	DC	C1'-O4'-C4'	5.25	115.35	110.10
2	D	425	DT	C2-N3-C4	-5.24	124.05	127.20
2	D	426	DG	C4-C5-N7	-5.19	108.72	110.80
2	D	431	DG	C6-C5-N7	-5.14	127.32	130.40
2	D	432	DA	C4'-C3'-C2'	-5.07	98.53	103.10
2	D	428	DG	O3'-P-O5'	-5.03	94.44	104.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	D	425	DT	Sidechain
2	D	433	DC	Sidechain

## 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	C	249	0	135	1	0
2	D	245	0	135	2	0
3	A	2606	0	2587	11	0
4	A	26	0	19	0	0
5	A	116	0	0	0	0
5	C	15	0	0	0	0
5	D	8	0	0	1	0
All	All	3265	0	2876	12	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 2.

All (12) 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)
3:A:125:ALA:HB2	3:A:166:ILE:HD12	1.82	0.61
3:A:95:ASP:OD1	3:A:97:ARG:HD3	2.09	0.53
2:D:427:DA:C8	3:A:81:CYS:SG	2.94	0.51
2:D:425:DT:H2'	2:D:426:DG:C8	2.48	0.48
3:A:101:PHE:O	3:A:104:ILE:HB	2.15	0.47
1:C:406:DG:O6	3:A:256:GLY:HA3	2.14	0.47
3:A:25:ARG:HD2	3:A:53:PHE:CE1	2.51	0.46
5:D:353:HOH:O	3:A:80:PRO:HA	2.16	0.45
3:A:69:ILE:HA	3:A:70:PRO:HD3	1.81	0.44
3:A:238:GLY:HA3	3:A:257:GLY:HA3	2.00	0.43
3:A:184:LYS:HA	3:A:185:PRO:HD3	1.83	0.41
3:A:45:ALA:O	3:A:48:VAL:HG23	2.22	0.40

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%)	316 (97%)	7 (2%)	2 (1%)	30 65

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	60	ASP
3	A	261	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
3	A	283/283 (100%)	270 (95%)	13 (5%)	33 67

All (13) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	28	LEU
3	A	41	TRP
3	A	48	VAL
3	A	81	CYS
3	A	87	SER
3	A	117	PHE
3	A	220	GLU
3	A	223	THR
3	A	231	ILE
3	A	245	ARG
3	A	266	LEU
3	A	270	LYS
3	A	325	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
3	A	63	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
3	A	141	ASN

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

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

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

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	-	20,28,28	0.70	0	19,40,40	1.18	1 (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
4	SAH	A	328	-	-	0/7/31/31	0/3/3/3

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
4	A	328	SAH	N3-C2-N1	2.77	131.02	128.89

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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 [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

EDS was not executed - this section will therefore be empty.

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

EDS was not executed - this section will therefore be empty.

### 6.3 Carbohydrates [i](#)

EDS was not executed - this section will therefore be empty.

### 6.4 Ligands [i](#)

EDS was not executed - this section will therefore be empty.

### 6.5 Other polymers [i](#)

EDS was not executed - this section will therefore be empty.