wwPDB X-ray Structure Validation Summary Report (i)

May 22, 2020 – 03:52 am BST

PDB ID	:	1HAP
Title	:	COMPLEX OF HUMAN ALPHA-THROMBIN WITH A 15MER OLIGONUCLEOTIDE GGTTGGTGTGGTTGG (BASED ON X-RAY MODEL OF DNA)
Authors	:	Padmanabhan, K.; Tulinsky, A.
Deposited on	:	1995-10-03
Resolution	:	2.80 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:

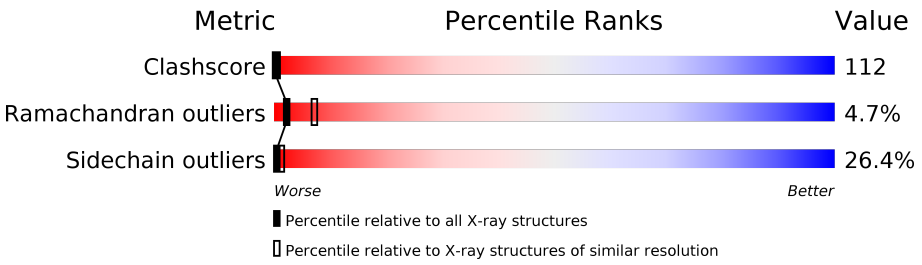
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 2.80 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	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)

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%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	D	15	20% 73% 7%
2	L	36	8% 31% 28% 8% 25%
3	H	259	10% 47% 32% 9% •

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	0G6	H	297	X	-	-	-

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 2771 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(*GP*GP*TP*TP*GP*GP*TP*GP*TP*GP*GP*TP*TP*GP*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	D	15	315	150	57	94	14	0	0	0

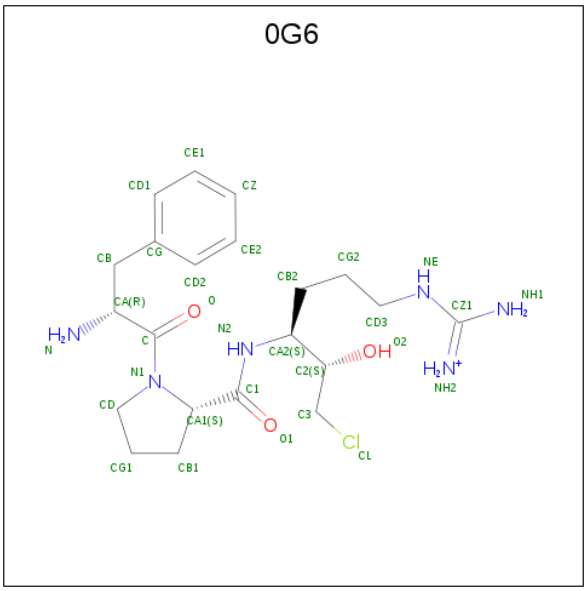
- Molecule 2 is a protein called Thrombin light chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	L	27	222	140	36	45	1	0	0	0

- Molecule 3 is a protein called Thrombin heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	H	253	2053	1310	362	367	14	0	0	0

- Molecule 4 is D-phenylalanyl-N-[(2S,3S)-6-{[amino(iminio)methyl]amino}-1-chloro-2-hydroxyhexan-3-yl]-L-prolinamide (three-letter code: 0G6) (formula: $C_{21}H_{34}ClN_6O_3$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
4	H	1	30	21	6	3	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	D	16	16	16	0	0
			Total	O		
5	L	10	10	10	0	0
			Total	O		
5	H	125	125	125	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. 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: 5'-D(*GP*GP*TP*TP*GP*GP*TP*GP*TP*GP*GP*TP*TP*GP*G)-3'

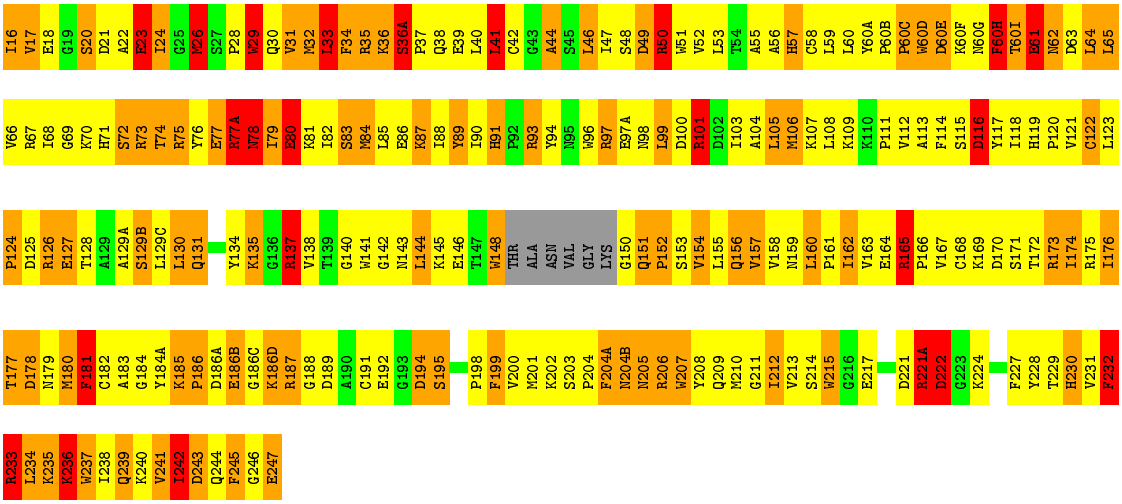
Chain D: 20% 73% 7%

- Molecule 2: Thrombin light chain

Chain L: 8% 31% 28% 8% 25%

- Molecule 3: Thrombin heavy chain

Chain H: 10% 47% 32% 9%



4 Data and refinement statistics (i)

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

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	58.28Å 77.61Å 100.02Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	10.00 – 2.80	Depositor
% Data completeness (in resolution range)	68.0 (10.00-2.80)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	NUCLIN, PROLSQ	Depositor
R, R_{free}	0.161 , (Not available)	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	2771	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: 0G6

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	D	1.71	4/353 (1.1%)	3.90	90/547 (16.5%)
2	L	1.05	0/224	2.43	12/298 (4.0%)
3	H	1.02	1/2107 (0.0%)	2.29	105/2846 (3.7%)
All	All	1.14	5/2684 (0.2%)	2.60	207/3691 (5.6%)

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	D	0	1
2	L	0	1
3	H	0	2
All	All	0	4

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	401	DG	O4'-C1'	6.06	1.49	1.42
1	D	409	DT	C4-O4	5.60	1.28	1.23
3	H	60(D)	TRP	C-N	-5.49	1.21	1.34
1	D	404	DT	C4-O4	5.46	1.28	1.23
1	D	413	DT	C4-O4	5.32	1.28	1.23

The worst 5 of 207 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	404	DT	P-O3'-C3'	26.73	151.78	119.70
1	D	405	DG	P-O3'-C3'	19.18	142.72	119.70

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	H	137	ARG	NE-CZ-NH1	17.40	129.00	120.30
1	D	406	DG	O4'-C1'-N9	17.18	120.02	108.00
1	D	411	DG	O4'-C1'-N9	13.28	117.30	108.00

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	401	DG	Sidechain
3	H	60(H)	PHE	Mainchain
3	H	78	ASN	Mainchain
2	L	14(D)	ARG	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	D	315	0	173	33	8
2	L	222	0	225	49	0
3	H	2053	0	2017	496	8
4	H	30	0	30	16	0
5	D	16	0	0	1	0
5	H	125	0	0	11	0
5	L	10	0	0	1	0
All	All	2771	0	2445	554	8

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 112.

The worst 5 of 554 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:H:162:ILE:HD11	3:H:199:PHE:CZ	1.66	1.30
3:H:172:THR:CG2	3:H:176:ILE:HD11	1.69	1.22
3:H:51:TRP:HE1	3:H:247:GLU:N	1.38	1.21

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:H:51:TRP:CZ2	3:H:246:GLY:HA2	1.78	1.19
3:H:50:ARG:HB2	3:H:247:GLU:HG2	1.21	1.16

The worst 5 of 8 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:D:413:DT:OP2	3:H:101:ARG:NH1[2_564]	1.15	1.05
1:D:413:DT:P	3:H:101:ARG:NH1[2_564]	1.50	0.70
1:D:413:DT:C5'	3:H:101:ARG:NE[2_564]	1.58	0.62
1:D:413:DT:O5'	3:H:101:ARG:NH1[2_564]	1.72	0.48
1:D:413:DT:C5'	3:H:101:ARG:CZ[2_564]	1.80	0.40

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
2	L	25/36 (69%)	17 (68%)	6 (24%)	2 (8%)	1 2
3	H	249/259 (96%)	197 (79%)	41 (16%)	11 (4%)	2 8
All	All	274/295 (93%)	214 (78%)	47 (17%)	13 (5%)	2 7

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	H	233	ARG
2	L	2	GLY
3	H	49	ASP
3	H	61	GLU
3	H	77	GLU

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
2	L	25/31 (81%)	18 (72%)	7 (28%)	0 1
3	H	221/225 (98%)	163 (74%)	58 (26%)	0 1
All	All	246/256 (96%)	181 (74%)	65 (26%)	0 1

5 of 65 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	H	80	GLU
3	H	116	ASP
3	H	237	TRP
3	H	83	SER
3	H	87	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
3	H	38	GLN
3	H	57	HIS
3	H	204(B)	ASN
3	H	244	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)

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	0G6	H	297	3	30,31,32	2.19	2 (6%)	37,41,42	1.47	6 (16%)

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	0G6	H	297	3	1/1/8/10	8/31/41/43	0/2/2/2

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	H	297	0G6	C3-C2	-8.86	1.25	1.51
4	H	297	0G6	O2-C2	-6.69	1.24	1.43

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	H	297	0G6	CA2-N2-C1	-3.03	117.72	123.07
4	H	297	0G6	CB2-CG2-CD3	2.84	120.58	112.05
4	H	297	0G6	O-C-N1	2.66	126.12	121.38
4	H	297	0G6	O2-C2-C3	2.45	116.99	109.74
4	H	297	0G6	C1-CA1-N1	-2.41	105.91	112.56

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
4	H	297	0G6	C2

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	H	297	0G6	O2-C2-CA2-CB2
4	H	297	0G6	C2-CA2-CB2-CG2
4	H	297	0G6	CA2-CB2-CG2-CD3
4	H	297	0G6	C3-C2-CA2-N2
4	H	297	0G6	O2-C2-CA2-N2

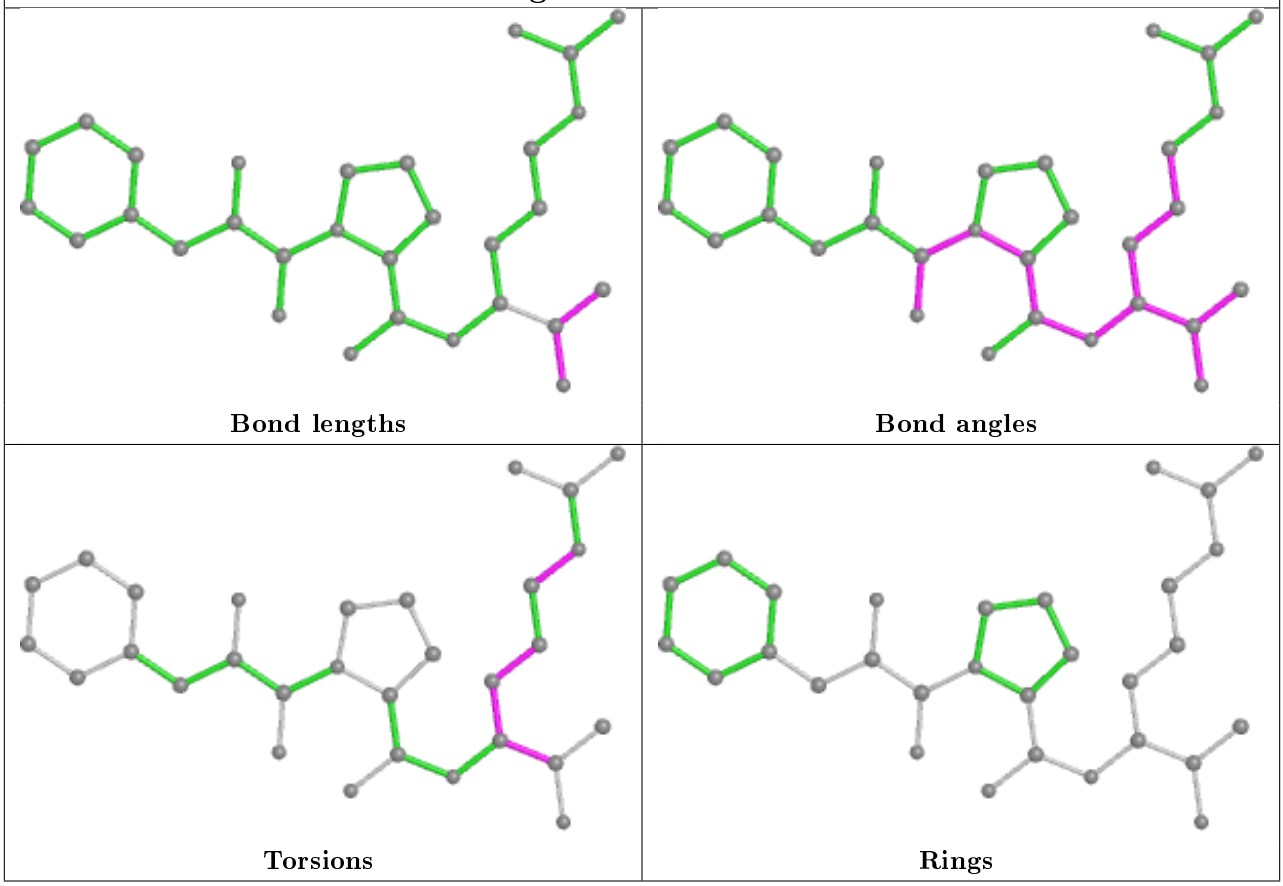
There are no ring outliers.

1 monomer is involved in 16 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	H	297	0G6	16	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 0G6 H 297



5.7 Other polymers (i)

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 (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.