W O R L D W I D E

wwPDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

May 21, 2020 – 05:44 am BST

PDB ID	:	5J9Q
Title	:	Crystal structure of the NuA4 core complex
Authors	:	Chen, Z.C.; Xu, P.
Deposited on	:	2016-04-11
Resolution	:	3.25 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 ⓘ symbol.

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

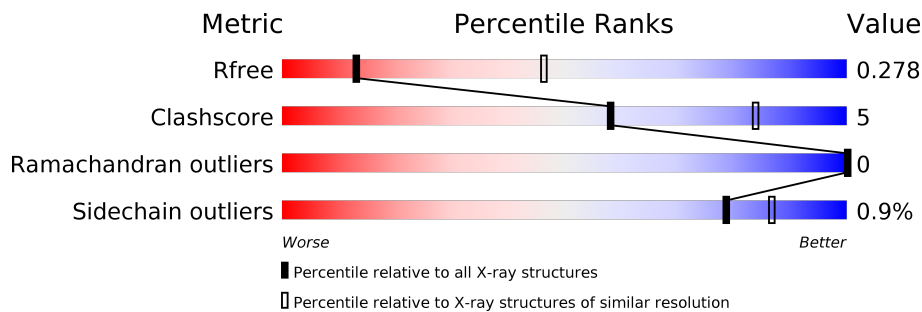|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.25 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
Rfree		0.278
Clashscore		5
Ramachandran outliers		0
Sidechain outliers		0.9%

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))
R_{free}	130704	1191 (3.30-3.22)
Clashscore	141614	1251 (3.30-3.22)
Ramachandran outliers	138981	1229 (3.30-3.22)
Sidechain outliers	138945	1228 (3.30-3.22)

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%

Mol	Chain	Length	Quality of chain
1	A	305	81% 15% •
1	E	305	84% 11% • •
1	I	305	80% 15% • •
2	B	113	61% 9% 30%
2	F	113	59% 7% 34%
2	J	113	55% 12% 33%
3	C	351	74% 9% 17%

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	G	351	73% · · · 10% · 16%
3	N	351	69% · · · 13% · 18%
4	D	120	92% · · · 8%
4	H	120	84% · · · 15% ·
4	K	120	84% · · · 16%
5	L	10	20% · 20% · 10% · 50%
5	M	10	10% · 20% · 10% · 60%
5	O	10	30% · 70%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 19671 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 acetyltransferase ESA1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	E	293	Total 2479	C 1601	N 420	O 447	S 11	0	0	0
1	A	292	Total 2470	C 1596	N 419	O 445	S 10	0	0	0
1	I	292	Total 2470	C 1596	N 419	O 445	S 10	0	0	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	338	GLN	GLU	engineered mutation	UNP Q08649
A	338	GLN	GLU	engineered mutation	UNP Q08649
I	338	GLN	GLU	engineered mutation	UNP Q08649

- Molecule 2 is a protein called Chromatin modification-related protein EAF6.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	F	75	Total 622	C 389	N 99	O 133	S 1	0	0	0
2	B	79	Total 660	C 413	N 108	O 138	S 1	0	0	0
2	J	76	Total 627	C 392	N 100	O 134	S 1	0	0	0

- Molecule 3 is a protein called Enhancer of polycomb-like protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	G	295	Total 2476	C 1562	N 437	O 469	S 8	0	0	0
3	N	289	Total 2426	C 1533	N 424	O 461	S 8	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	293	Total 2458	C 1554	N 429	O 467	S 8	0	0	0

- Molecule 4 is a protein called Chromatin modification-related protein YNG2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	120	Total 969	C 613	N 165	O 188	S 3	0	0	0
4	H	120	Total 969	C 613	N 165	O 188	S 3	0	0	0
4	K	120	Total 969	C 613	N 165	O 188	S 3	0	0	0

- Molecule 5 is a protein called Htz1.

| Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
5	L	5	Total 32	C 18	N 6	O 8	0	0	0
5	M	4	Total 26	C 15	N 5	O 6	0	0	0
5	O	3	Total 18	C 11	N 4	O 3	0	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 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.
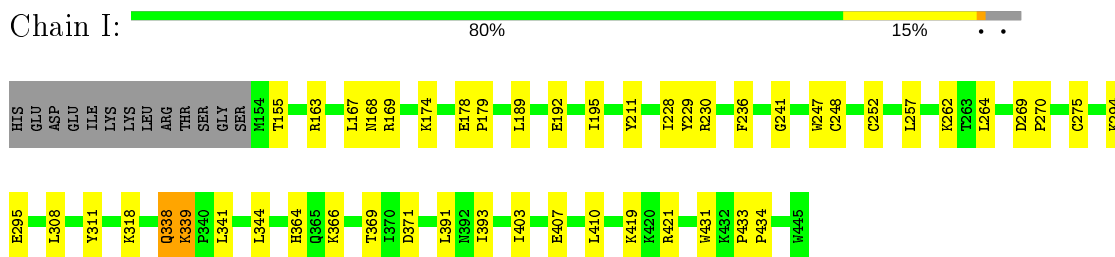
- Molecule 1: Histone acetyltransferase ESA1
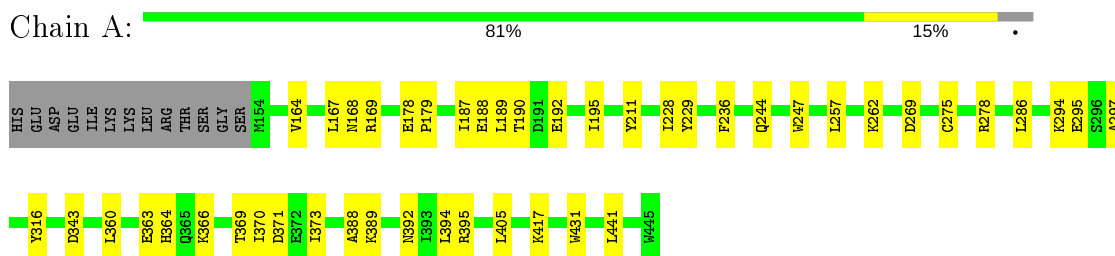
Chain E:

- Molecule 1: Histone acetyltransferase ESA1

Chain A:

- Molecule 1: Histone acetyltransferase ESA1

Chain I:

- Molecule 2: Chromatin modification-related protein EAF6

Chain F:

• Molecule 2: Chromatin modification-related protein EAF6

Chain B:

• Molecule 2: Chromatin modification-related protein EAF6

Chain J:

• Molecule 3: Enhancer of polycomb-like protein 1

Chain G:

• Molecule 3: Enhancer of polycomb-like protein 1

Chain N:

• Molecule 3: Enhancer of polycomb-like protein 1

Chain C:

HIS THR LYS HIS K118 I121 P122 T123 Y137 T153 D168 I191 A199 L220 S227 ASP MET ALA D231 R235 F249 I250 T251 P255 Q258 R262 P263 L264 F271 R282 P293 I306 R314 A344 L366 K367 I368 K379
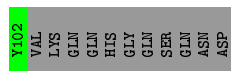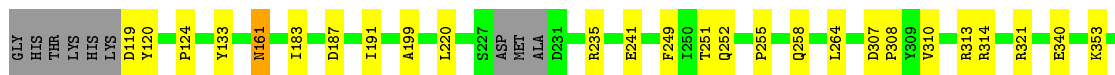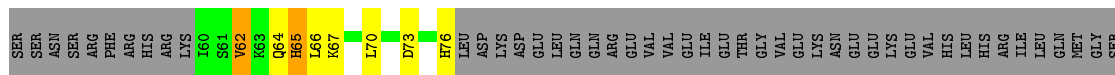
G386 K395 ARG LYS ARG PRO THR

- Molecule 4: Chromatin modification-related protein YNG2

Chain D:

92%   8%

M1 T10 V14 D32 H62 E65 L76 T92 H100 E112 L117 V120

- Molecule 4: Chromatin modification-related protein YNG2

Chain H:

84%   15%

M1 D2 L5 E8 F21 D32 L33 Y42 Q56 P60 K61 H62 E65 Q83 K86 T92 L111 E112 E113 D114 G115 V116 L117 V120

- Molecule 4: Chromatin modification-related protein YNG2

Chain K:

84%   16%

M1 T10 P18 L25 I28 D32 Y42 H50 R54 I59 H62 E65 L76 N102 K103 L104 E105 K106 L111 E112 L117 V120

- Molecule 5: Htz1

Chain L:

20%   20%   10%   50%

S1 G2 A3 K4 D5 SER GLY SER LEU ARG

- Molecule 5: Htz1

Chain M:

10%   20%   10%   60%

SER G2 A3 K4 D5 SER GLY SER LEU ARG

- Molecule 5: Htz1

Chain O:

30%   70%

SER G2 A3 K4 ASP SER GLY SER LEU ARG

4   Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	217.86Å 190.99Å 172.35Å 90.00° 113.56° 90.00°	Depositor
Resolution (Å)	30.00 – 3.25 46.74 – 3.25	Depositor EDS
% Data completeness (in resolution range)	78.9 (30.00-3.25) 78.7 (46.74-3.25)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.04 (at 3.25Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.250 , 0.273 0.252 , 0.278	Depositor DCC
R_{free} test set	2000 reflections (2.50%)	wwPDB-VP
Wilson B-factor (Å^2)	52.1	Xtriage
Anisotropy	0.035	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 26.0	EDS
L-test for twinning[2]	$<|L|> = 0.49$, $<L^2> = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.78	EDS
Total number of atoms	19671	wwPDB-VP
Average B, all atoms (Å^2)	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.45% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $<|L|>$, $<L^2>$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

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: ALY

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	0.29	0/2523	0.43	0/3411
1	E	0.27	0/2532	0.41	0/3422
1	I	0.31	1/2523 (0.0%)	0.43	0/3411
2	B	0.27	0/669	0.39	0/893
2	F	0.26	0/630	0.39	0/841
2	J	0.27	0/635	0.41	0/848
3	C	0.27	0/2509	0.41	0/3379
3	G	0.26	0/2528	0.42	0/3404
3	N	0.26	0/2477	0.40	0/3335
4	D	0.27	0/981	0.39	0/1317
4	H	0.27	0/981	0.38	0/1317
4	K	0.27	0/981	0.39	0/1317
5	L	0.33	0/31	0.81	0/39
5	M	0.29	0/25	0.92	0/31
5	O	0.24	0/17	0.45	0/20
All	All	0.28	1/20042 (0.0%)	0.41	0/26985

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	I	0	1
3	G	0	1
All	All	0	2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	I	339	LYS	C-N	6.50	1.46	1.34

There are no bond angle outliers.

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	G	62	VAL	Peptide
1	I	338	GLN	Peptide

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	A	2470	0	2446	30	0
1	E	2479	0	2458	25	0
1	I	2470	0	2445	35	0
2	B	660	0	619	7	0
2	F	622	0	579	7	0
2	J	627	0	581	15	0
3	C	2458	0	2414	27	0
3	G	2476	0	2427	26	0
3	N	2426	0	2374	36	0
4	D	969	0	1003	9	0
4	H	969	0	1003	15	0
4	K	969	0	1003	14	0
5	L	32	0	32	3	0
5	M	26	0	24	3	0
5	O	18	0	20	0	0
All	All	19671	0	19428	188	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 5.

The worst 5 of 188 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:I:407:GLU:HG2	3:C:63:LYS:HG3	1.67	0.74
1:E:338:GLN:NE2	5:L:2:GLY:O	2.23	0.71

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:G:379:LYS:NZ	3:G:386:GLY:O	2.18	0.71
2:J:41:THR:HG22	2:J:93:ARG:HH22	1.55	0.71
4:D:14:VAL:HG22	4:D:100:HIS:HD2	1.57	0.69

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	
1	A	289/305 (95%)	273 (94%)	16 (6%)	0	100	100
1	E	290/305 (95%)	275 (95%)	15 (5%)	0	100	100
1	I	289/305 (95%)	274 (95%)	15 (5%)	0	100	100
2	B	73/113 (65%)	70 (96%)	3 (4%)	0	100	100
2	F	69/113 (61%)	67 (97%)	2 (3%)	0	100	100
2	J	70/113 (62%)	66 (94%)	4 (6%)	0	100	100
3	C	287/351 (82%)	273 (95%)	14 (5%)	0	100	100
3	G	289/351 (82%)	278 (96%)	11 (4%)	0	100	100
3	N	283/351 (81%)	275 (97%)	8 (3%)	0	100	100
4	D	118/120 (98%)	117 (99%)	1 (1%)	0	100	100
4	H	118/120 (98%)	116 (98%)	2 (2%)	0	100	100
4	K	118/120 (98%)	117 (99%)	1 (1%)	0	100	100
5	L	3/10 (30%)	3 (100%)	0	0	100	100
5	M	2/10 (20%)	2 (100%)	0	0	100	100
5	O	1/10 (10%)	0	1 (100%)	0	100	100
All	All	2299/2697 (85%)	2206 (96%)	93 (4%)	0	100	100

There are no Ramachandran outliers to report.

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	
1	A	270/283 (95%)	270 (100%)	0	100	100
1	E	272/283 (96%)	270 (99%)	2 (1%)	84	90
1	I	270/283 (95%)	269 (100%)	1 (0%)	91	94
2	B	73/100 (73%)	71 (97%)	2 (3%)	44	70
2	F	69/100 (69%)	69 (100%)	0	100	100
2	J	69/100 (69%)	68 (99%)	1 (1%)	67	81
3	C	279/333 (84%)	277 (99%)	2 (1%)	84	90
3	G	280/333 (84%)	275 (98%)	5 (2%)	59	77
3	N	274/333 (82%)	272 (99%)	2 (1%)	84	90
4	D	111/111 (100%)	111 (100%)	0	100	100
4	H	111/111 (100%)	109 (98%)	2 (2%)	59	77
4	K	111/111 (100%)	110 (99%)	1 (1%)	78	87
5	L	3/7 (43%)	2 (67%)	1 (33%)	0	0
5	M	2/7 (29%)	1 (50%)	1 (50%)	0	0
5	O	1/7 (14%)	1 (100%)	0	100	100
All	All	2195/2502 (88%)	2175 (99%)	20 (1%)	78	87

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

Mol	Chain	Res	Type
3	N	250	ILE
3	C	65	HIS
4	H	116	VAL
1	I	369	THR
3	N	183	ILE

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

Mol	Chain	Res	Type
1	A	364	HIS
1	I	338	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)

3 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	ALY	A	262	1	10,11,12	0.79	0	7,12,14	1.17	1 (14%)
1	ALY	E	262	1	10,11,12	1.00	1 (10%)	7,12,14	1.02	0
1	ALY	I	262	1	10,11,12	0.83	0	7,12,14	0.89	0

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	ALY	A	262	1	-	3/9/10/12	-
1	ALY	E	262	1	-	4/9/10/12	-
1	ALY	I	262	1	-	3/9/10/12	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	262	ALY	CH-NZ	2.04	1.40	1.34

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	262	ALY	CE-NZ-CH	2.40	126.24	122.56

There are no chirality outliers.

5 of 10 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	E	262	ALY	O-C-CA-CB
1	I	262	ALY	CG-CD-CE-NZ
1	A	262	ALY	OH-CH-NZ-CE
1	A	262	ALY	CH3-CH-NZ-CE
1	E	262	ALY	OH-CH-NZ-CE

There are no ring outliers.

3 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	A	262	ALY	1	0
1	E	262	ALY	1	0
1	I	262	ALY	1	0

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

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)

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.