W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

May 16, 2020 – 02:42 am BST

PDB ID	:	5N94
Title	:	Crystal Structure of Drosophila DHX36 helicase in complex with polyU
Authors	:	Chen, W.-F.; Rety, S.; Hai-Lei Guo, H.-L.; Wu, W.-Q.; Liu, N.-N.; Liu, Q.-W.; Dai, Y.-X.; Xi, X.-G.
Deposited on	:	2017-02-24
Resolution	:	2.43 Å(reported)

The following versions of software and data (see references (1)) were used in the production of this report:

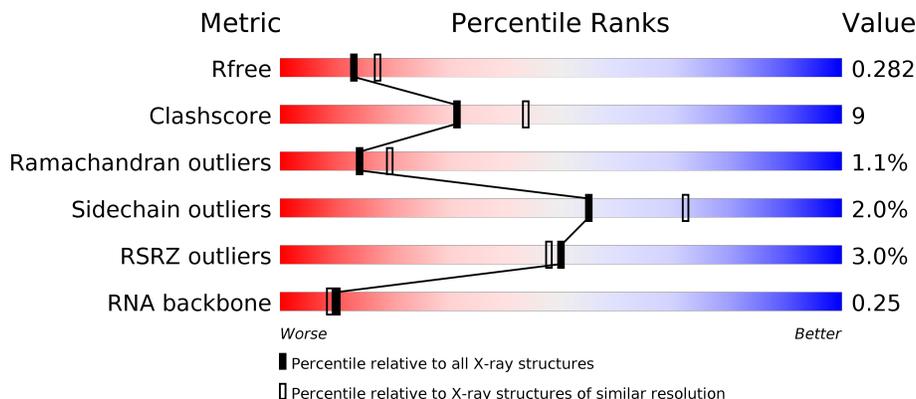|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.43 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		0.282
Clashscore		9
Ramachandran outliers		1.1%
Sidechain outliers		2.0%
RSRZ outliers		3.0%
RNA backbone		0.25

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	4647 (2.44-2.40)
Clashscore	141614	5161 (2.44-2.40)
Ramachandran outliers	138981	5073 (2.44-2.40)
Sidechain outliers	138945	5074 (2.44-2.40)
RSRZ outliers	127900	4543 (2.44-2.40)
RNA backbone	3102	1080 (2.80-2.04)

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% 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.

Mol	Chain	Length	Quality of chain
1	A	944	3% / 72% 17% · 10%
1	B	944	3% / 71% 17% · 10%
2	C	8	63% 25% 13%
2	D	8	38% 38% 25%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 14015 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 CG9323, isoform A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	849	Total 6814	C 4304	N 1200	O 1265	S 45	0	0	0
1	B	849	Total 6814	C 4304	N 1200	O 1265	S 45	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	943	VAL	-	expression tag	UNP Q8SWT2
A	944	ASP	-	expression tag	UNP Q8SWT2
B	943	VAL	-	expression tag	UNP Q8SWT2
B	944	ASP	-	expression tag	UNP Q8SWT2

- Molecule 2 is a RNA chain called RNA (5'-R(P*UP*UP*UP*UP*UP*UP*UP*U)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	8	Total 160	C 72	N 16	O 64	P 8	0	0	0
2	D	8	Total 160	C 72	N 16	O 64	P 8	0	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	30	Total 30	O 30	0	0
3	B	29	Total 29	O 29	0	0
3	C	4	Total 4	O 4	0	0
3	D	4	Total 4	O 4	0	0

3 Residue-property plots (i)

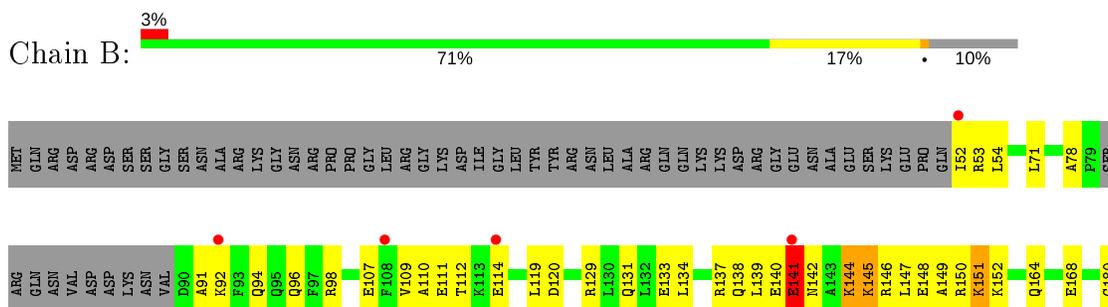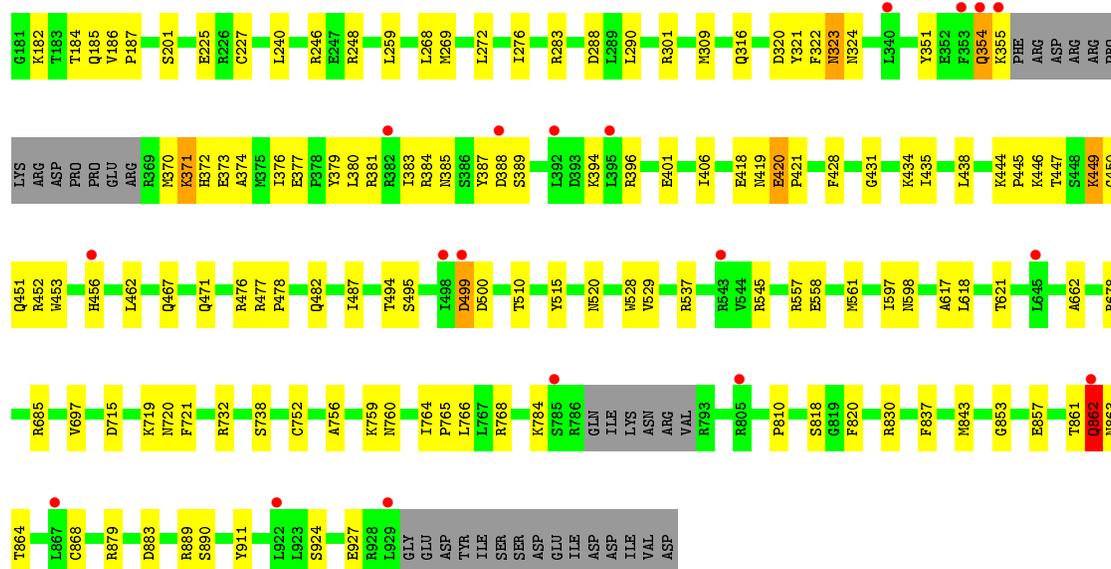These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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.
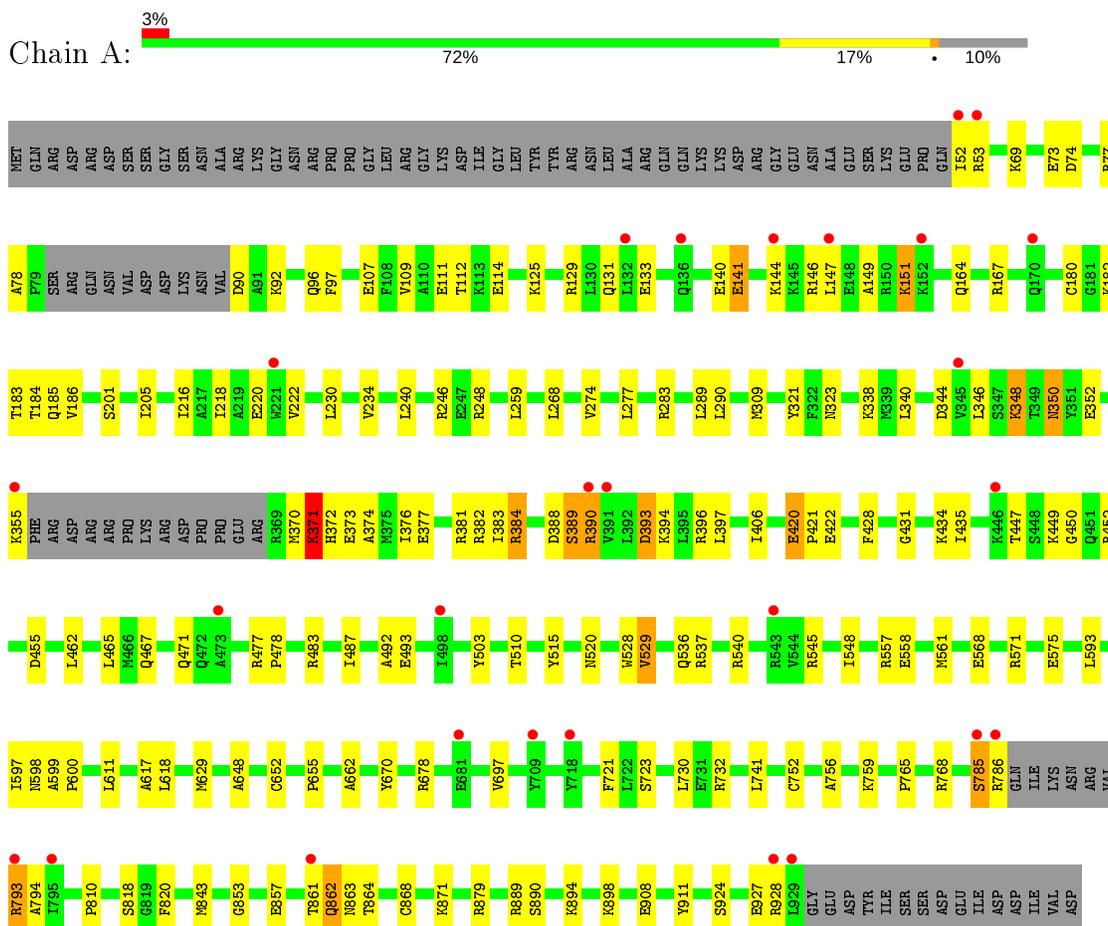
- Molecule 1: CG9323, isoform A

- Molecule 1: CG9323, isoform A

G181 K182 T183 T184 Q185 V186 P187 S201 E225 R226 C227 L240 R246 E247 R248 L259 L268 M269 L272 I276 R283 D288 L289 L290 R301 M309 Q316 D320 Y321 F322 N323 N324 L340 Y351 E352 F353 Q354 K355 PHE ARG ASP ARG PRO

LYS ARG ASP PRO GLU ARG R369 M370 K371 H372 E373 A374 M375 I376 E377 P378 Y379 L380 R381 R382 I383 R384 N385 S386 Y387 D388 S389 L392 D393 K394 L395 R396 E401 I406 E418 N419 E420 P421 F428 G431 K434 I435 L438 K444 P445 K446 T447 S448 K449 G450

Q451 R452 W453 H456 L462 Q467 Q471 R476 R477 P478 Q482 I487 T494 S495 I498 D499 D500 T510 Y515 N520 W528 V529 R537 R543 V544 R545 R557 E558 M561 I597 N598 A617 L618 T621 L645 A662 R678

R685 V697 D715 K719 N720 F721 R732 S738 C752 A756 K759 N760 I764 P765 L766 L767 R768 K784 S785 R786 GLN ILE LYS ASN ARG VAL R793 R805 P810 S818 G819 F820 R830 F837 M843 G853 E857 T861 Q862 N863

T864 I867 C868 R879 D883 R889 S890 Y911 L922 L923 S924 E927 R928 L929 GLY GLU ASP TYR ILE SER SER ASP GLU ILE ASP ASP ILE VAL ASP

● Molecule 2: RNA (5'-R(P*UP*UP*UP*UP*UP*UP*UP*U)-3')

Chain C:

63%   25%   13%

U1 U2 U3 U8

● Molecule 2: RNA (5'-R(P*UP*UP*UP*UP*UP*UP*UP*U)-3')

Chain D:

38%   38%   25%

U1 U2 U3 U4 U8

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	338.33Å 51.53Å 159.90Å 90.00° 118.15° 90.00°	Depositor
Resolution (Å)	47.00 – 2.43	Depositor
	47.00 – 2.43	EDS
% Data completeness (in resolution range)	90.8 (47.00-2.43)	Depositor
	90.5 (47.00-2.43)	EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.25 (at 2.42Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.212 , 0.278	Depositor
	0.219 , 0.282	DCC
R_{free} test set	4043 reflections (4.78%)	wwPDB-VP
Wilson B-factor (Å^2)	49.9	Xtriage
Anisotropy	0.747	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 27.1	EDS
L-test for twinning[2]	$<|L|> = 0.49$, $<L^2> = 0.32$	Xtriage
Estimated twinning fraction	0.469 for -h-2*l,-k,l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	14015	wwPDB-VP
Average B, all atoms (Å^2)	60.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.02% 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)

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	0.50	1/6933 (0.0%)	0.69	1/9353 (0.0%)
1	B	0.49	1/6933 (0.0%)	0.69	3/9353 (0.0%)
2	C	0.67	0/175	1.69	6/268 (2.2%)
2	D	0.71	0/175	1.60	7/268 (2.6%)
All	All	0.50	2/14216 (0.0%)	0.73	17/19242 (0.1%)

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	A	0	3
1	B	0	4
All	All	0	7

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	752	CYS	CB-SG	-9.31	1.66	1.82
1	B	752	CYS	CB-SG	-5.31	1.73	1.81

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	1	U	C5-C6-N1	10.71	128.05	122.70
2	D	1	U	C5-C6-N1	8.22	126.81	122.70
2	D	1	U	C2-N1-C1'	7.08	126.20	117.70
2	C	1	U	C2-N1-C1'	6.98	126.08	117.70
2	D	3	U	C2-N1-C1'	6.82	125.89	117.70

There are no chirality outliers.

5 of 7 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	141	GLU	Peptide
1	A	390	ARG	Peptide
1	A	90	ASP	Peptide
1	B	141	GLU	Peptide
1	B	322	PHE	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	6814	0	6913	126	0
1	B	6814	0	6913	123	0
2	C	160	0	81	6	0
2	D	160	0	81	4	0
3	A	30	0	0	3	0
3	B	29	0	0	1	0
3	C	4	0	0	0	0
3	D	4	0	0	1	0
All	All	14015	0	13988	250	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 9.

The worst 5 of 250 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:D:8:U:O3'	3:D:101:HOH:O	1.84	0.95
1:B:862:GLN:HB3	1:B:864:THR:H	1.33	0.91
1:B:283:ARG:NH2	1:B:598:ASN:OD1	2.02	0.91
1:A:373:GLU:HA	1:A:376:ILE:HG22	1.51	0.90
1:B:323:ASN:OD1	1:B:324:ASN:N	2.04	0.89

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	841/944 (89%)	792 (94%)	38 (4%)	11 (1%)	12	16
1	B	841/944 (89%)	793 (94%)	41 (5%)	7 (1%)	19	27
All	All	1682/1888 (89%)	1585 (94%)	79 (5%)	18 (1%)	14	19

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	350	ASN
1	A	371	LYS
1	A	389	SER
1	A	785	SER
1	B	323	ASN

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	757/842 (90%)	741 (98%)	16 (2%)	53	71
1	B	757/842 (90%)	742 (98%)	15 (2%)	55	72
All	All	1514/1684 (90%)	1483 (98%)	31 (2%)	55	72

5 of 31 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	898	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	119	LEU
1	B	738	SER
1	A	928	ARG
1	B	145	LYS

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

Mol	Chain	Res	Type
1	A	164	GLN
1	A	209	GLN
1	A	863	ASN
1	A	909	ASN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	C	8/8 (100%)	2 (25%)	1 (12%)
2	D	8/8 (100%)	2 (25%)	1 (12%)
All	All	16/16 (100%)	4 (25%)	2 (12%)

All (4) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	C	2	U
2	C	3	U
2	D	2	U
2	D	3	U

All (2) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	C	1	U
2	D	1	U

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)

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)

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, 95th 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^2)	Q<0.9
1	A	849/944 (89%)	0.30	27 (3%)	47	45	37, 57, 90, 112	0
1	B	849/944 (89%)	0.25	24 (2%)	53	50	36, 57, 90, 108	0
2	C	8/8 (100%)	-0.07	0	100	100	43, 56, 71, 83	0
2	D	8/8 (100%)	-0.36	0	100	100	44, 58, 73, 83	0
All	All	1714/1904 (90%)	0.27	51 (2%)	50	48	36, 57, 90, 112	0

The worst 5 of 51 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	795	ILE	4.9
1	A	53	ARG	4.8
1	B	929	LEU	4.7
1	B	785	SER	4.3
1	A	543	ARG	3.8

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

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

6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.