W O R L D W I D E
wwPDB
P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 8, 2022 – 04:04 pm GMT

|		
PDB ID	:	7Q3O
Title	:	Structure of CDX1 bound to hydroxymethylated DNA
Authors	:	Morgunova, E.; Yin, Y.; Popov, A.; Taipale, J.
Deposited on	:	2021-10-28
Resolution	:	2.78 Å(reported)

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 2.78 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.
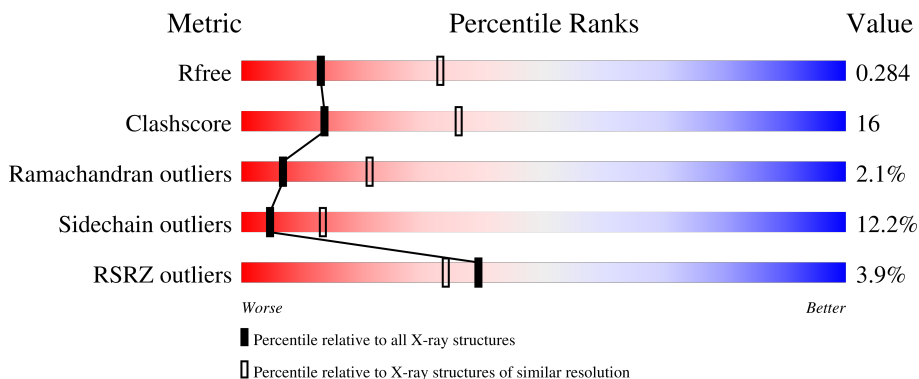
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4107 (2.80-2.76)
Clashscore	141614	4575 (2.80-2.76)
Ramachandran outliers	138981	4487 (2.80-2.76)
Sidechain outliers	138945	4489 (2.80-2.76)
RSRZ outliers	127900	4027 (2.80-2.76)

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% 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	18	6% · 39% · 56% · 6%
1	B	18	11% · 50% · 44% · 6%
1	D	18	50% · 44% · 6%
1	F	18	6% · 56% · 33% · 11%
2	H	18	6% · 44% · 44% · 11%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	I	18	6% 39% 56% 6%
2	J	18	6% 39% 56% 6%
2	L	18	6% 39% 56% 6%
3	C	65	2% 68% 26%
3	E	65	6% 68% 18% 5% 8%
3	G	65	2% 75% 14% 5% 6%
3	K	65	2% 72% 20% 6%

2   Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 5202 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 hydroxymethylated DNA (18-MER).

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	18	Total	C	N	O	P		0	0	0
			368	177	57	116	18				
1	B	18	Total	C	N	O	P		0	0	0
			368	177	57	116	18				
1	D	18	Total	C	N	O	P		0	0	0
			368	177	57	116	18				
1	F	18	Total	C	N	O	P		0	0	0
			368	177	57	116	18				

- Molecule 2 is a DNA chain called hydroxymethylated DNA (18-MER).

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	H	18	Total	C	N	O	P		0	0	0
			373	177	76	102	18				
2	I	18	Total	C	N	O	P		0	0	0
			373	177	76	102	18				
2	J	18	Total	C	N	O	P		0	0	0
			373	177	76	102	18				
2	L	18	Total	C	N	O	P		0	0	0
			373	177	76	102	18				

- Molecule 3 is a protein called Homeobox protein CDX-1.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	K	61	Total	C	N	O	0	0	0
			541	340	108	93			
3	C	63	Total	C	N	O	0	0	0
			556	349	111	96			
3	E	60	Total	C	N	O	0	0	0
			530	335	104	91			
3	G	61	Total	C	N	O	0	0	0
			541	340	108	93			

- Molecule 4 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	5	Total 5	O 5	0	0
4	H	5	Total 5	O 5	0	0
4	B	4	Total 4	O 4	0	0
4	I	3	Total 3	O 3	0	0
4	D	6	Total 6	O 6	0	0
4	J	8	Total 8	O 8	0	0
4	F	6	Total 6	O 6	0	0
4	L	4	Total 4	O 4	0	0
4	K	9	Total 9	O 9	0	0
4	C	11	Total 11	O 11	0	0
4	E	4	Total 4	O 4	0	0
4	G	5	Total 5	O 5	0	0

3 Residue-property plots ⓘ

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 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: hydroxymethylated DNA (18-MER)
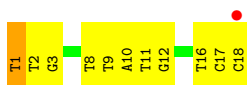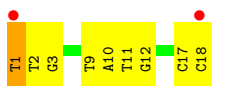
Chain A:

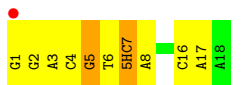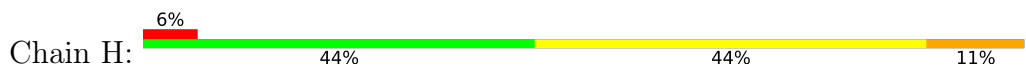- Molecule 1: hydroxymethylated DNA (18-MER)

Chain B:

- Molecule 1: hydroxymethylated DNA (18-MER)

Chain D:

- Molecule 1: hydroxymethylated DNA (18-MER)

Chain F:

- Molecule 2: hydroxymethylated DNA (18-MER)

Chain H:

- Molecule 2: hydroxymethylated DNA (18-MER)
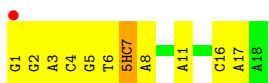
Chain I:

6%

39% 56% 6%

G1 G2 A3 C4 G5 T6 5HC7 A8 A11 C16 A17 A18

• Molecule 2: hydroxymethylated DNA (18-MER)

Chain J:

6%

39% 56% 6%

G1 G2 A3 C4 G5 T6 5HC7 A8 A11 C16 A17 A18

• Molecule 2: hydroxymethylated DNA (18-MER)

Chain L:

6%

39% 56% 6%

G1 G2 A3 C4 G5 T6 5HC7 A8 C16 A17 A18

• Molecule 3: Homeobox protein CDX-1

Chain K:

2%

72% 20% 6%

T183 R184 T185 K186 V191 H206 R209 L219 R228 Q229 I232 N236 R237 E241 R242 K243 VAL ASN LYS LYS

• Molecule 3: Homeobox protein CDX-1

Chain C:

2%

68% 26%

T183 R184 T185 K186 D187 K188 Y189 R190 R198 H206 R209 R214 L219 T226 E227 R228 Q235 N236 E241 K243 V244 N245 LYS LYS

• Molecule 3: Homeobox protein CDX-1

Chain E:

6%

68% 18% 5% 8%

THR ARG T185 K186 D187 K188 Y189 R190 H196 F205 H206 Y207 S208 R209 R214 R215 K216 L219 K231 K240 E241 R242 K243 V244 ASN LYS LYS

• Molecule 3: Homeobox protein CDX-1

Chain G:

2%

75% 14% 5% 6%

T183 R184 T185 K186 D195 R198 R209 R214 S217 E218 L219 T226 N236 E241 R242 K243 VAL ASN LYS LYS
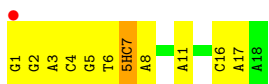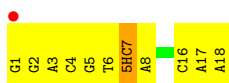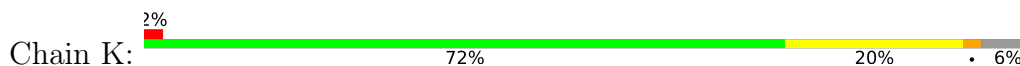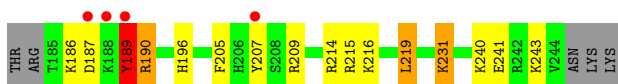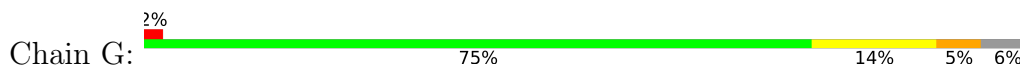
4   Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	45.37Å 64.67Å 67.03Å 112.37° 108.11° 93.93°	Depositor
Resolution (Å)	42.18 – 2.78 42.14 – 2.78	Depositor EDS
% Data completeness (in resolution range)	90.0 (42.18-2.78) 90.0 (42.14-2.78)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.15 (at 2.77Å)	Xtriage
Refinement program	REFMAC 5.8.0257	Depositor
R, R_{free}	0.217 , 0.286 0.220 , 0.284	Depositor DCC
R_{free} test set	709 reflections (4.77%)	wwPDB-VP
Wilson B-factor (Å^2)	63.4	Xtriage
Anisotropy	0.776	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< |L| > = 0.52, < L^2 > = 0.35$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	5202	wwPDB-VP
Average B, all atoms (Å^2)	85.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.07% 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: 5HC

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	0.69	1/409 (0.2%)	0.89	0/628
1	B	0.74	1/409 (0.2%)	0.91	0/628
1	D	0.68	1/409 (0.2%)	0.91	0/628
1	F	0.73	1/409 (0.2%)	0.90	1/628 (0.2%)
2	H	0.66	1/396 (0.3%)	0.95	0/606
2	I	0.58	0/396	0.96	0/606
2	J	0.55	0/396	0.93	0/606
2	L	0.65	0/396	0.95	0/606
3	C	0.68	0/565	0.97	0/754
3	E	0.69	0/539	1.00	0/719
3	G	0.72	0/550	1.01	0/733
3	K	0.70	0/550	1.01	0/733
All	All	0.68	5/5424 (0.1%)	0.95	1/7875 (0.0%)

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	1	DT	OP3-P	-10.65	1.48	1.61
1	F	1	DT	OP3-P	-10.18	1.49	1.61
1	D	1	DT	OP3-P	-9.64	1.49	1.61
1	A	1	DT	OP3-P	-9.10	1.50	1.61
2	H	5	DG	O3'-P	-5.81	1.54	1.61

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	F	10	DA	C1'-O4'-C4'	-5.18	104.92	110.10

There are no chirality outliers.

There are no planarity outliers.

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	A	368	0	208	14	0
1	B	368	0	208	13	0
1	D	368	0	208	12	0
1	F	368	0	208	10	0
2	H	373	0	203	15	1
2	I	373	0	203	15	1
2	J	373	0	203	15	0
2	L	373	0	203	17	0
3	C	556	0	571	16	0
3	E	530	0	545	10	0
3	G	541	0	556	7	0
3	K	541	0	556	18	0
4	A	5	0	0	1	0
4	B	4	0	0	0	0
4	C	11	0	0	1	0
4	D	6	0	0	2	0
4	E	4	0	0	1	0
4	F	6	0	0	0	0
4	G	5	0	0	0	0
4	H	5	0	0	0	0
4	I	3	0	0	2	0
4	J	8	0	0	0	0
4	K	9	0	0	0	0
4	L	4	0	0	2	0
All	All	5202	0	3872	143	1

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 16.

The worst 5 of 143 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:L:17:DA:N7	4:L:101:HOH:O	2.01	0.94
1:F:1:DT:H2'	1:F:2:DT:C6	2.04	0.93
1:B:1:DT:H2'	1:B:2:DT:C6	2.04	0.92

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:1:DT:H2'	1:D:2:DT:C6	2.05	0.92
1:A:1:DT:H2'	1:A:2:DT:C6	2.04	0.92

All (1) 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)
2:H:1:DG:O6	2:I:11:DA:OP1[1_554]	2.14	0.06

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	C	61/65 (94%)	58 (95%)	2 (3%)	1 (2%)	9	28
3	E	58/65 (89%)	54 (93%)	2 (3%)	2 (3%)	3	11
3	G	59/65 (91%)	56 (95%)	2 (3%)	1 (2%)	9	27
3	K	59/65 (91%)	57 (97%)	1 (2%)	1 (2%)	9	27
All	All	237/260 (91%)	225 (95%)	7 (3%)	5 (2%)	7	21

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	E	190	ARG
3	E	189	TYR
3	K	186	LYS
3	C	186	LYS
3	G	186	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	C	59/61 (97%)	50 (85%)	9 (15%)	2	7
3	E	56/61 (92%)	48 (86%)	8 (14%)	3	9
3	G	57/61 (93%)	50 (88%)	7 (12%)	4	13
3	K	57/61 (93%)	53 (93%)	4 (7%)	15	37
All	All	229/244 (94%)	201 (88%)	28 (12%)	5	13

5 of 28 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	E	187	ASP
3	G	219	LEU
3	E	219	LEU
3	G	209	ARG
3	E	214	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
3	E	236	ASN
3	G	235	GLN
3	G	236	ASN
3	C	206	HIS
3	C	236	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

4 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	#$|Z| > 2$
2	5HC	L	7	2,1	18,22,23	1.04	1 (5%)	23,31,34	1.63	4 (17%)
2	5HC	I	7	2,1	18,22,23	1.01	1 (5%)	23,31,34	1.58	3 (13%)
2	5HC	J	7	2,1	18,22,23	0.86	0	23,31,34	1.47	3 (13%)
2	5HC	H	7	2,1	18,22,23	1.18	3 (16%)	23,31,34	1.55	3 (13%)

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
2	5HC	L	7	2,1	-	5/9/23/24	0/2/2/2
2	5HC	I	7	2,1	-	5/9/23/24	0/2/2/2
2	5HC	J	7	2,1	-	4/9/23/24	0/2/2/2
2	5HC	H	7	2,1	-	4/9/23/24	0/2/2/2

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	H	7	5HC	O3'-C3'	3.12	1.50	1.43
2	L	7	5HC	O3'-C3'	2.60	1.48	1.43
2	I	7	5HC	O3'-C3'	2.55	1.48	1.43
2	H	7	5HC	C6-N1	-2.32	1.34	1.38
2	H	7	5HC	C6-C5	2.01	1.40	1.34

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	L	7	5HC	O3'-C3'-C4'	4.78	128.41	110.10
2	H	7	5HC	O3'-C3'-C4'	4.70	128.09	110.10
2	I	7	5HC	O3'-C3'-C4'	4.55	127.50	110.10
2	J	7	5HC	O3'-C3'-C4'	-4.19	94.08	110.10
2	L	7	5HC	O4'-C1'-N1	3.06	113.32	107.86

There are no chirality outliers.

5 of 18 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	H	7	5HC	C4-C5-C5M-O5
2	H	7	5HC	C6-C5-C5M-O5
2	I	7	5HC	C4-C5-C5M-O5
2	I	7	5HC	C6-C5-C5M-O5
2	H	7	5HC	C3'-C4'-C5'-O5'

There are no ring outliers.

4 monomers are involved in 11 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	L	7	5HC	3	0
2	I	7	5HC	3	0
2	J	7	5HC	2	0
2	H	7	5HC	3	0

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

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 ⓘ

6.1 Protein, DNA and RNA chains ⓘ

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	18/18 (100%)	-0.16	1 (5%)	24	19	56, 77, 148, 225	0
1	B	18/18 (100%)	-0.23	2 (11%)	5	3	65, 91, 144, 155	0
1	D	18/18 (100%)	-0.18	0	100	100	63, 104, 140, 156	0
1	F	18/18 (100%)	-0.24	1 (5%)	24	19	55, 72, 129, 201	0
2	H	17/18 (94%)	0.74	1 (5%)	22	17	63, 75, 124, 263	0
2	I	17/18 (94%)	-0.05	1 (5%)	22	17	74, 89, 120, 139	0
2	J	17/18 (94%)	-0.20	1 (5%)	22	17	74, 89, 130, 136	0
2	L	17/18 (94%)	0.44	1 (5%)	22	17	63, 73, 124, 260	0
3	C	63/65 (96%)	-0.15	1 (1%)	72	69	61, 76, 103, 118	0
3	E	60/65 (92%)	-0.07	4 (6%)	17	13	55, 74, 110, 125	0
3	G	61/65 (93%)	-0.04	1 (1%)	72	69	50, 67, 88, 102	0
3	K	61/65 (93%)	0.06	1 (1%)	72	69	51, 69, 91, 98	0
All	All	385/404 (95%)	-0.03	15 (3%)	39	34	50, 75, 117, 263	0

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	H	1	DG	21.7
2	L	1	DG	13.8
1	A	18	DC	5.7
2	I	1	DG	3.7
2	J	1	DG	3.7

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

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum,

median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	5HC	I	7	21/22	0.91	0.15	72,93,99,112	0
2	5HC	J	7	21/22	0.93	0.14	89,95,102,106	0
2	5HC	H	7	21/22	0.96	0.14	63,73,79,85	0
2	5HC	L	7	21/22	0.96	0.16	52,61,79,83	0

6.3 Carbohydrates (i)

There are no monosaccharides 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.