



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 9, 2026 – 09:20 PM UTC

PDB ID : 5B2J / pdb_00005b2j
Title : Human nucleosome containing CpG methylated DNA
Authors : Fujii, Y.; Wakamori, M.; Umehara, T.; Yokoyama, S.
Deposited on : 2016-01-18
Resolution : 2.60 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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

MolProbity	:	4-5-2 with Phenix2.0
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

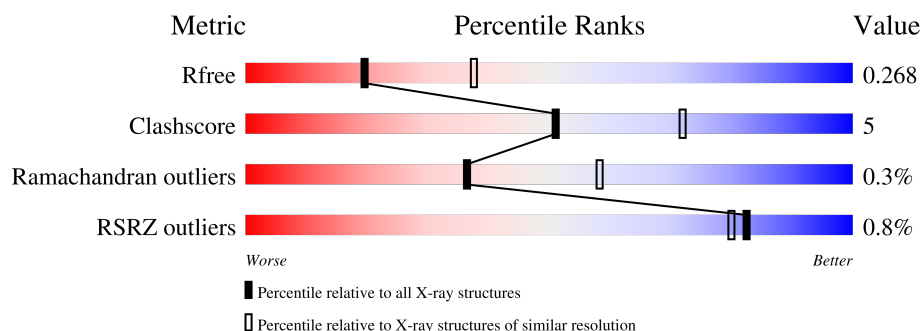
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.60 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.



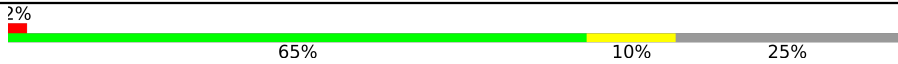
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	180053	4008 (2.60-2.60)
Clashscore	190562	4347 (2.60-2.60)
Ramachandran outliers	187476	4277 (2.60-2.60)
RSRZ outliers	180081	4008 (2.60-2.60)

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 $\leq 5\%$. 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.

Mol	Chain	Length	Quality of chain
1	A	139	
1	E	139	
2	B	106	
2	F	106	
3	C	133	
3	G	133	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	D	129	 2% 65% 10% 25%
4	H	129	 64% 9% 27%
5	I	146	 68% 32%
5	J	146	 % 66% 33% .

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 12171 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 H3.1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	99	Total	C	N	O	S	0	0	0
			815	514	158	139	4			
1	E	99	Total	C	N	O	S	0	0	0
			816	514	158	140	4			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	GLY	-	expression tag	UNP P68431
A	-2	SER	-	expression tag	UNP P68431
A	-1	HIS	-	expression tag	UNP P68431
E	-3	GLY	-	expression tag	UNP P68431
E	-2	SER	-	expression tag	UNP P68431
E	-1	HIS	-	expression tag	UNP P68431

- Molecule 2 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	82	Total	C	N	O	S	0	0	0
			653	412	127	113	1			
2	F	85	Total	C	N	O	S	0	0	0
			683	430	136	116	1			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-3	GLY	-	expression tag	UNP P62805
B	-2	SER	-	expression tag	UNP P62805
B	-1	HIS	-	expression tag	UNP P62805
F	-3	GLY	-	expression tag	UNP P62805
F	-2	SER	-	expression tag	UNP P62805
F	-1	HIS	-	expression tag	UNP P62805

- Molecule 3 is a protein called Histone H2A type 1-B/E.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	106	Total	C	N	O	0	0	0
			819	517	160	142			
3	G	105	Total	C	N	O	0	0	0
			810	511	158	141			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-3	GLY	-	expression tag	UNP P04908
C	-2	SER	-	expression tag	UNP P04908
C	-1	HIS	-	expression tag	UNP P04908
G	-3	GLY	-	expression tag	UNP P04908
G	-2	SER	-	expression tag	UNP P04908
G	-1	HIS	-	expression tag	UNP P04908

- Molecule 4 is a protein called Histone H2B type 1-J.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	97	Total	C	N	O	S	0	0	0
			765	480	142	141	2			
4	H	94	Total	C	N	O	S	0	0	0
			736	462	134	138	2			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	-6	GLY	-	expression tag	UNP P06899
D	-5	SER	-	expression tag	UNP P06899
D	-4	HIS	-	expression tag	UNP P06899
H	-6	GLY	-	expression tag	UNP P06899
H	-5	SER	-	expression tag	UNP P06899
H	-4	HIS	-	expression tag	UNP P06899

- Molecule 5 is a DNA chain called DNA (146-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	I	146	Total	C	N	O	P	0	0	0
			2994	1436	539	874	145			
5	J	146	Total	C	N	O	P	0	0	0
			2994	1436	539	874	145			

- Molecule 6 is MANGANESE (II) ION (CCD ID: MN) (formula: Mn).

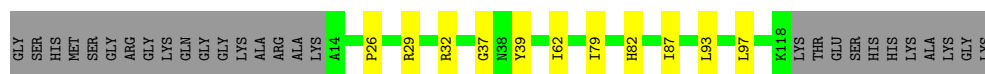
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	E	1	Total 1	Mn 1	0	0
6	I	3	Total 3	Mn 3	0	0
6	J	3	Total 3	Mn 3	0	0

- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	7	Total 7	O 7	0	0
7	B	4	Total 4	O 4	0	0
7	C	7	Total 7	O 7	0	0
7	D	8	Total 8	O 8	0	0
7	E	19	Total 19	O 19	0	0
7	F	15	Total 15	O 15	0	0
7	G	6	Total 6	O 6	0	0
7	H	2	Total 2	O 2	0	0
7	I	5	Total 5	O 5	0	0
7	J	6	Total 6	O 6	0	0

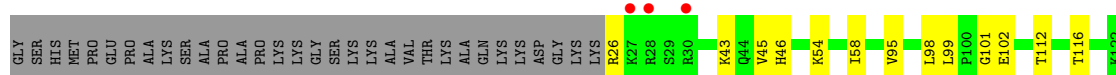
- Molecule 3: Histone H2A type 1-B/E

Chain G:  71% 8% 21%



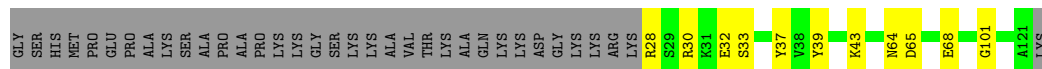
- Molecule 4: Histone H2B type 1-J

Chain D:  2% 65% 10% 25%



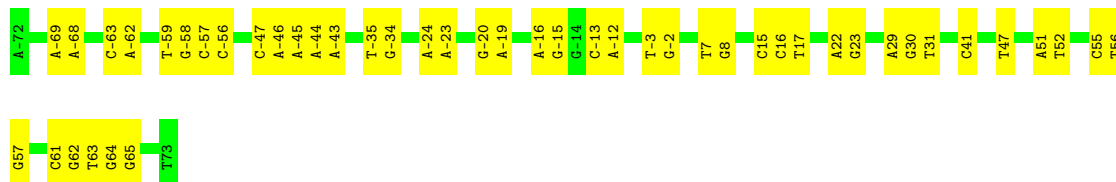
- Molecule 4: Histone H2B type 1-J

Chain H:  64% 9% 27%



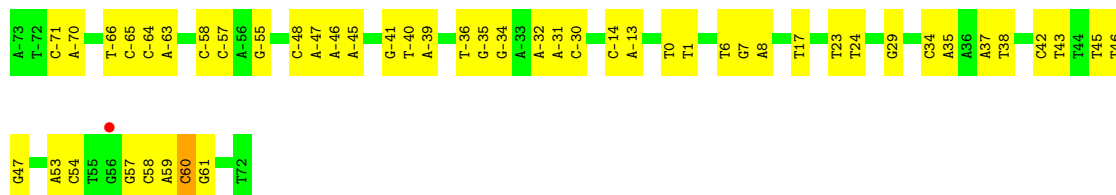
- Molecule 5: DNA (146-MER)

Chain I:  68% 32%



- Molecule 5: DNA (146-MER)

Chain J:  66% 33% 1%



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	106.54Å 109.67Å 181.35Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.90 – 2.60 19.90 – 2.60	Depositor EDS
% Data completeness (in resolution range)	99.5 (19.90-2.60) 91.4 (19.90-2.60)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	0.13	Depositor
$\langle I/\sigma(I) \rangle$ ¹	0.63 (at 2.51Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.197 , 0.252 (Not available) , 0.268	Depositor DCC
R_{free} test set	1988 reflections (2.69%)	wwPDB-VP
Wilson B-factor (Å ²)	53.5	Xtriage
Anisotropy	0.331	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 55.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.45$, $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	0.029 for k,h,-l	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	12171	wwPDB-VP
Average B, all atoms (Å ²)	96.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.35% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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: 5CM, MN

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	0.15	0/827	0.38	0/1109
1	E	0.17	0/828	0.39	0/1109
2	B	0.12	0/660	0.34	0/883
2	F	0.16	0/691	0.39	0/923
3	C	0.18	0/829	0.33	0/1118
3	G	0.12	0/820	0.32	0/1107
4	D	0.16	0/776	0.35	0/1040
4	H	0.13	0/747	0.35	0/1004
5	I	0.19	0/3266	0.41	0/5031
5	J	0.20	0/3266	0.42	0/5031
All	All	0.18	0/12710	0.39	0/18355

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	815	0	856	10	0
1	E	816	0	856	9	0
2	B	653	0	696	7	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	F	683	0	729	6	0
3	C	819	0	879	6	0
3	G	810	0	866	7	0
4	D	765	0	797	10	0
4	H	736	0	758	9	0
5	I	2994	0	1661	34	0
5	J	2994	0	1661	38	0
6	E	1	0	0	0	0
6	I	3	0	0	0	0
6	J	3	0	0	0	0
7	A	7	0	0	0	0
7	B	4	0	0	0	0
7	C	7	0	0	0	0
7	D	8	0	0	0	0
7	E	19	0	0	1	0
7	F	15	0	0	0	0
7	G	6	0	0	0	0
7	H	2	0	0	0	0
7	I	5	0	0	1	0
7	J	6	0	0	2	0
All	All	12171	0	9759	109	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 109 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)
5:I:29:DA:OP2	7:I:201:HOH:O	2.00	0.79
1:E:69:ARG:NH2	5:I:17:DT:OP2	2.16	0.79
5:J:-14:DC:H2"	5:J:-13:DA:C8	2.25	0.72
2:B:92:ARG:HH21	4:D:98:LEU:HD23	1.53	0.72
4:H:30:ARG:NH2	5:J:-45:DA:OP1	2.25	0.70

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	97/139 (70%)	97 (100%)	0	0	100	100
1	E	97/139 (70%)	96 (99%)	1 (1%)	0	100	100
2	B	80/106 (76%)	78 (98%)	2 (2%)	0	100	100
2	F	83/106 (78%)	80 (96%)	3 (4%)	0	100	100
3	C	104/133 (78%)	103 (99%)	1 (1%)	0	100	100
3	G	103/133 (77%)	101 (98%)	2 (2%)	0	100	100
4	D	95/129 (74%)	94 (99%)	0	1 (1%)	11	25
4	H	92/129 (71%)	90 (98%)	1 (1%)	1 (1%)	11	25
All	All	751/1014 (74%)	739 (98%)	10 (1%)	2 (0%)	36	58

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	H	101	GLY
4	D	101	GLY

5.3.2 Protein sidechains [i](#)

There are no protein residues with a non-rotameric sidechain to report in this entry.

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

8 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	$\# Z > 2$
5	5CM	I	61	5	18,21,22	0.80	0	24,30,33	0.66	0
5	5CM	I	-5	5	18,21,22	0.82	0	24,30,33	0.59	0
5	5CM	J	-62	5	18,21,22	0.81	0	24,30,33	0.78	0
5	5CM	J	60	5	18,21,22	0.76	0	24,30,33	0.80	1 (4%)
5	5CM	J	4	5	18,21,22	0.81	0	24,30,33	0.79	0
5	5CM	J	-6	5	18,21,22	0.70	0	24,30,33	0.62	0
5	5CM	I	5	5	18,21,22	0.71	0	24,30,33	0.70	0
5	5CM	I	-61	5	18,21,22	0.83	0	24,30,33	0.71	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
5	5CM	I	61	5	-	0/7/21/22	0/2/2/2
5	5CM	I	-5	5	-	6/7/21/22	0/2/2/2
5	5CM	J	-62	5	-	0/7/21/22	0/2/2/2
5	5CM	J	60	5	-	0/7/21/22	0/2/2/2
5	5CM	J	4	5	-	6/7/21/22	0/2/2/2
5	5CM	J	-6	5	-	6/7/21/22	0/2/2/2
5	5CM	I	5	5	-	6/7/21/22	0/2/2/2
5	5CM	I	-61	5	-	0/7/21/22	0/2/2/2

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	J	60	5CM	C4'-O4'-C1'	-2.13	104.45	109.51

There are no chirality outliers.

5 of 24 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	J	-6	5CM	C3'-C4'-C5'-O5'
5	J	-6	5CM	O4'-C4'-C5'-O5'
5	I	-5	5CM	C2'-C1'-N1-C6
5	I	5	5CM	C2'-C1'-N1-C6
5	J	-6	5CM	C2'-C1'-N1-C6

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	I	61	5CM	1	0
5	J	60	5CM	1	0

5.5 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 7 ligands modelled in this entry, 7 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

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 [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	99/139 (71%)	-0.20	1 (1%) 79 76	49, 71, 100, 145	0
1	E	99/139 (71%)	-0.41	1 (1%) 79 76	40, 56, 91, 120	0
2	B	82/106 (77%)	-0.19	1 (1%) 76 73	51, 67, 98, 143	0
2	F	85/106 (80%)	-0.48	1 (1%) 76 73	40, 55, 80, 113	0
3	C	106/133 (79%)	-0.44	0 100 100	36, 62, 91, 129	0
3	G	105/133 (78%)	-0.25	0 100 100	52, 73, 104, 124	0
4	D	97/129 (75%)	-0.23	3 (3%) 51 45	40, 63, 116, 140	0
4	H	94/129 (72%)	-0.21	0 100 100	50, 71, 103, 172	0
5	I	142/146 (97%)	0.12	0 100 100	66, 124, 178, 201	0
5	J	142/146 (97%)	0.23	1 (0%) 84 82	68, 123, 177, 224	0
All	All	1051/1306 (80%)	-0.17	8 (0%) 82 80	36, 72, 151, 224	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	102	GLY	4.2
2	F	102	GLY	4.1
4	D	27	LYS	3.1
1	E	135	ALA	2.7
1	A	39	HIS	2.5

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

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($\text{\AA}^2$)	Q<0.9
5	5CM	I	-61	20/21	0.67	0.16	123,150,166,170	0
5	5CM	J	-62	20/21	0.89	0.12	83,113,125,126	0
5	5CM	J	60	20/21	0.89	0.10	102,127,134,135	0
5	5CM	I	61	20/21	0.91	0.09	108,122,139,142	0
5	5CM	I	5	20/21	0.91	0.10	78,98,110,115	0
5	5CM	J	4	20/21	0.93	0.08	96,107,132,133	0
5	5CM	J	-6	20/21	0.94	0.09	76,91,114,119	0
5	5CM	I	-5	20/21	0.95	0.07	76,91,108,108	0

6.3 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

6.4 Ligands [i](#)

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($\text{\AA}^2$)	Q<0.9
6	MN	J	103	1/1	0.77	0.11	138,138,138,138	0
6	MN	J	101	1/1	0.80	0.08	140,140,140,140	0
6	MN	I	102	1/1	0.80	0.13	134,134,134,134	0
6	MN	I	101	1/1	0.89	0.15	113,113,113,113	0
6	MN	E	301	1/1	0.92	0.10	58,58,58,58	0
6	MN	I	103	1/1	0.92	0.09	115,115,115,115	0
6	MN	J	102	1/1	0.93	0.08	114,114,114,114	0

6.5 Other polymers [i](#)

There are no such residues in this entry.