



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 5, 2026 – 04:11 PM UTC

PDB ID : 4GJP / pdb_00004gjp
Title : Crystal structure of the TAL effector dHax3 bound to dsDNA containing repetitive methyl-CpG
Authors : Yan, N.; Deng, D.; Yan, C.Y.; Yin, P.; Pan, X.J.; Shi, Y.G.
Deposited on : 2012-08-10
Resolution : 1.94 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

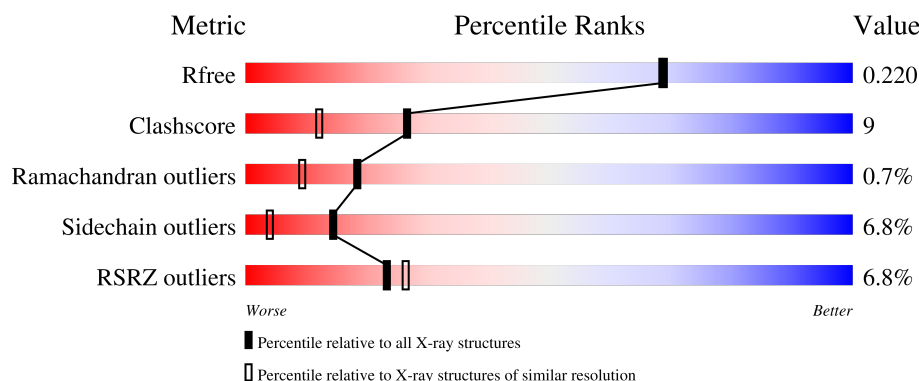
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.94 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	1452 (1.94-1.94)
Clashscore	190562	1494 (1.94-1.94)
Ramachandran outliers	187476	1479 (1.94-1.94)
Sidechain outliers	187428	1479 (1.94-1.94)
RSRZ outliers	180081	1453 (1.94-1.94)

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	499	8% 82% 14% .
1	B	499	6% 81% 15% ..
2	G	17	6% 65% 24% 12%
2	I	17	88% 12%
3	H	17	6% 65% 29% 6%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	J	17	 82%18%

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 9216 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 Hax3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	497	Total	C	N	O	S	0	0	0
			3596	2246	671	667	12			
1	B	489	Total	C	N	O	S	0	0	0
			3537	2209	659	657	12			

There are 52 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	230	MET	-	expression tag	UNP Q3ZD72
A	300	HIS	ASN	engineered mutation	UNP Q3ZD72
A	301	ASP	ILE	engineered mutation	UNP Q3ZD72
A	368	HIS	ASN	engineered mutation	UNP Q3ZD72
A	369	ASP	ILE	engineered mutation	UNP Q3ZD72
A	402	ASN	HIS	engineered mutation	UNP Q3ZD72
A	403	GLY	ASP	engineered mutation	UNP Q3ZD72
A	436	ASN	HIS	engineered mutation	UNP Q3ZD72
A	437	GLY	ASP	engineered mutation	UNP Q3ZD72
A	470	ASN	HIS	engineered mutation	UNP Q3ZD72
A	471	GLY	ASP	engineered mutation	UNP Q3ZD72
A	505	ASN	SER	engineered mutation	UNP Q3ZD72
A	539	GLY	SER	engineered mutation	UNP Q3ZD72
A	573	ASN	SER	engineered mutation	UNP Q3ZD72
A	606	ASN	HIS	engineered mutation	UNP Q3ZD72
A	607	GLY	ASP	engineered mutation	UNP Q3ZD72
A	640	HIS	ASN	engineered mutation	UNP Q3ZD72
A	641	ASP	ILE	engineered mutation	UNP Q3ZD72
A	721	LEU	-	expression tag	UNP Q3ZD72
A	722	GLU	-	expression tag	UNP Q3ZD72
A	723	HIS	-	expression tag	UNP Q3ZD72
A	724	HIS	-	expression tag	UNP Q3ZD72
A	725	HIS	-	expression tag	UNP Q3ZD72
A	726	HIS	-	expression tag	UNP Q3ZD72
A	727	HIS	-	expression tag	UNP Q3ZD72

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	728	HIS	-	expression tag	UNP Q3ZD72
B	230	MET	-	expression tag	UNP Q3ZD72
B	300	HIS	ASN	engineered mutation	UNP Q3ZD72
B	301	ASP	ILE	engineered mutation	UNP Q3ZD72
B	368	HIS	ASN	engineered mutation	UNP Q3ZD72
B	369	ASP	ILE	engineered mutation	UNP Q3ZD72
B	402	ASN	HIS	engineered mutation	UNP Q3ZD72
B	403	GLY	ASP	engineered mutation	UNP Q3ZD72
B	436	ASN	HIS	engineered mutation	UNP Q3ZD72
B	437	GLY	ASP	engineered mutation	UNP Q3ZD72
B	470	ASN	HIS	engineered mutation	UNP Q3ZD72
B	471	GLY	ASP	engineered mutation	UNP Q3ZD72
B	505	ASN	SER	engineered mutation	UNP Q3ZD72
B	539	GLY	SER	engineered mutation	UNP Q3ZD72
B	573	ASN	SER	engineered mutation	UNP Q3ZD72
B	606	ASN	HIS	engineered mutation	UNP Q3ZD72
B	607	GLY	ASP	engineered mutation	UNP Q3ZD72
B	640	HIS	ASN	engineered mutation	UNP Q3ZD72
B	641	ASP	ILE	engineered mutation	UNP Q3ZD72
B	721	LEU	-	expression tag	UNP Q3ZD72
B	722	GLU	-	expression tag	UNP Q3ZD72
B	723	HIS	-	expression tag	UNP Q3ZD72
B	724	HIS	-	expression tag	UNP Q3ZD72
B	725	HIS	-	expression tag	UNP Q3ZD72
B	726	HIS	-	expression tag	UNP Q3ZD72
B	727	HIS	-	expression tag	UNP Q3ZD72
B	728	HIS	-	expression tag	UNP Q3ZD72

- Molecule 2 is a DNA chain called DNA (5'-D(*TP*GP*TP*CP*CP*CP*TP*TP*(5CM)P*GP*(5CM)P*GP*TP*CP*TP*CP*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	I	17	Total	C	N	O	P	0	0	0
			337	164	50	107	16			
2	G	15	Total	C	N	O	P	0	0	0
			299	146	45	94	14			

- Molecule 3 is a DNA chain called DNA (5'-D(*AP*GP*AP*GP*AP*CP*GP*CP*GP*AP*AP*GP*GP*GP*AP*CP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	J	17	Total	C	N	O	P	0	0	0
			355	167	79	93	16			
3	H	16	Total	C	N	O	P	0	0	0
			337	157	74	90	16			

- Molecule 4 is MAGNESIUM ION (CCD ID: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	J	1	Total	Mg	0	0
			1	1		
4	B	1	Total	Mg	0	0
			1	1		

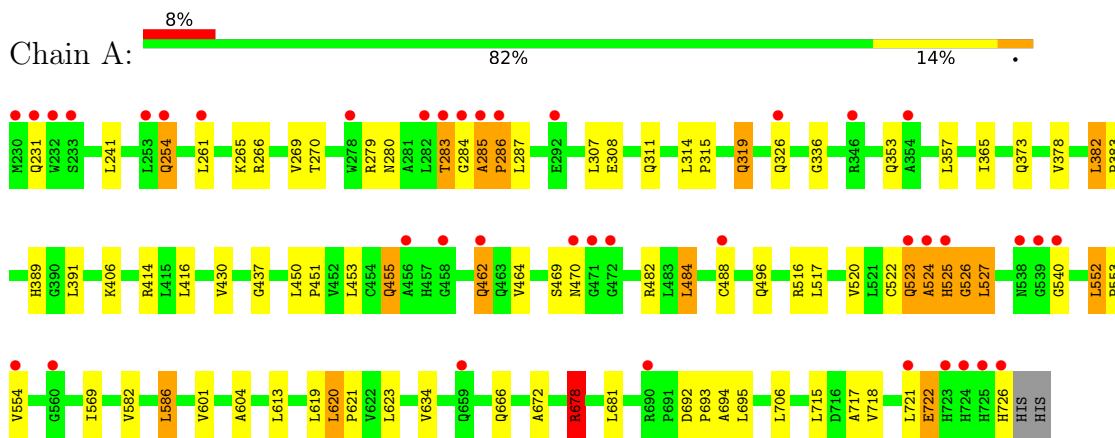
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	279	Total	O	0	0
			279	279		
5	I	63	Total	O	0	0
			63	63		
5	J	39	Total	O	0	0
			39	39		
5	B	290	Total	O	0	0
			290	290		
5	G	53	Total	O	0	0
			53	53		
5	H	29	Total	O	0	0
			29	29		

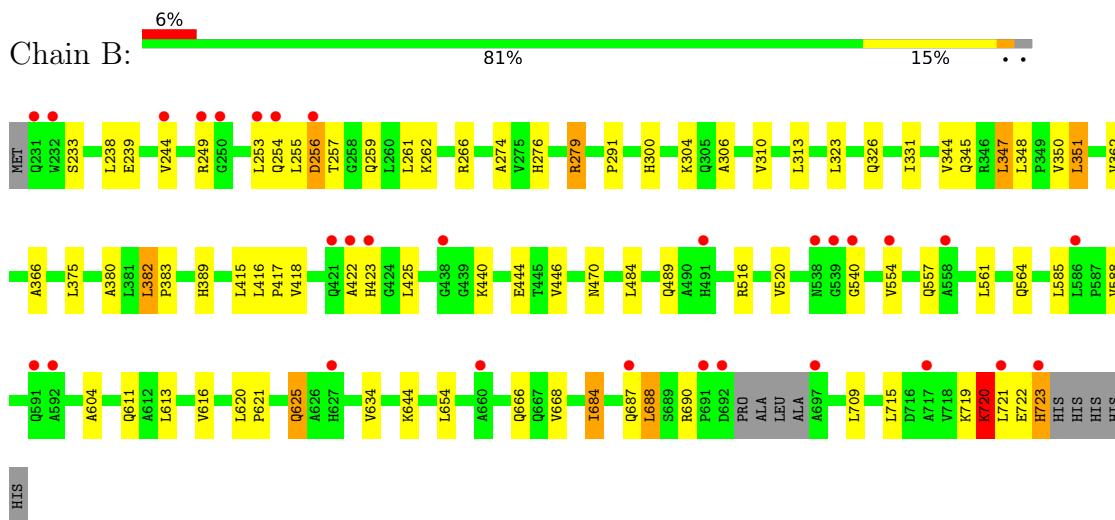
3 Residue-property plots [i](#)

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.

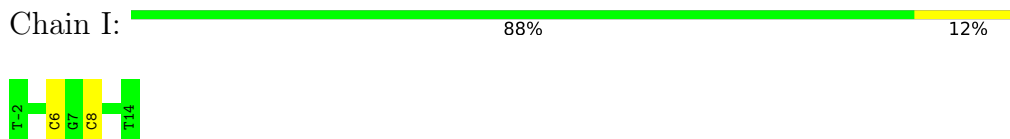
• Molecule 1: Hax3



• Molecule 1: Hax3



• Molecule 2: DNA (5'-D(*TP*GP*TP*CP*CP*CP*TP*TP*(5CM)P*GP*(5CM)P*GP*TP*C P*TP*CP*T)-3')



- Molecule 2: DNA (5'-D(*TP*GP*TP*CP*CP*CP*TP*TP*(5CM)P*GP*(5CM)P*GP*TP*CP*TP*CP*T)-3')

Chain G: 



- Molecule 3: DNA (5'-D(*AP*GP*AP*GP*AP*CP*GP*CP*GP*AP*AP*GP*GP*GP*AP*CP*A)-3')

Chain J: 



- Molecule 3: DNA (5'-D(*AP*GP*AP*GP*AP*CP*GP*CP*GP*AP*AP*GP*GP*GP*AP*CP*A)-3')

Chain H: 



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	81.20Å 87.11Å 88.15Å 90.00° 102.85° 90.00°	Depositor
Resolution (Å)	39.44 – 1.94 39.44 – 1.94	Depositor EDS
% Data completeness (in resolution range)	99.0 (39.44-1.94) 99.2 (39.44-1.94)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.87 (at 1.94Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.3_928)	Depositor
R, R_{free}	0.201 , 0.225 0.199 , 0.220	Depositor DCC
R_{free} test set	4413 reflections (4.96%)	wwPDB-VP
Wilson B-factor (Å ²)	33.7	Xtriage
Anisotropy	0.119	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 38.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	9216	wwPDB-VP
Average B, all atoms (Å ²)	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.58% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: MG, 5CM

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	0.63	0/3649	0.85	4/4985 (0.1%)
1	B	0.57	0/3586	0.82	3/4896 (0.1%)
2	G	0.35	0/285	0.94	0/433
2	I	0.33	0/327	0.97	0/497
3	H	0.32	0/381	0.93	1/587 (0.2%)
3	J	0.31	0/402	0.90	1/620 (0.2%)
All	All	0.56	0/8630	0.85	9/12018 (0.1%)

There are no bond length outliers.

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	678	ARG	CA-C-N	-6.70	112.80	119.56
1	A	678	ARG	C-N-CA	-6.70	112.80	119.56
3	H	-4	DA	P-O5'-C5'	-6.49	110.26	120.00
1	B	484	LEU	CA-C-N	-5.73	114.09	120.45
1	B	484	LEU	C-N-CA	-5.73	114.09	120.45

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	3596	0	3729	87	0
1	B	3537	0	3678	66	0
2	G	299	0	177	1	0
2	I	337	0	197	0	0
3	H	337	0	177	8	0
3	J	355	0	189	1	0
4	B	1	0	0	0	0
4	J	1	0	0	0	0
5	A	279	0	0	16	0
5	B	290	0	0	17	0
5	G	53	0	0	1	0
5	H	29	0	0	7	0
5	I	63	0	0	0	0
5	J	39	0	0	1	0
All	All	9216	0	8147	154	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 154 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:A:285:ALA:HB1	1:A:286:PRO:CD	1.66	1.25
1:A:717:ALA:O	1:A:721:LEU:HD23	1.39	1.19
1:A:285:ALA:HB1	1:A:286:PRO:HD2	1.22	1.18
3:H:1:DC:H2'	5:H:127:HOH:O	1.41	1.17
1:A:285:ALA:CB	1:A:286:PRO:HD2	1.83	1.06

There are no symmetry-related clashes.

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	
1	A	495/499 (99%)	480 (97%)	9 (2%)	6 (1%)	10	3
1	B	485/499 (97%)	467 (96%)	17 (4%)	1 (0%)	43	37
All	All	980/998 (98%)	947 (97%)	26 (3%)	7 (1%)	18	9

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	285	ALA
1	A	523	GLN
1	A	526	GLY
1	A	286	PRO
1	A	525	HIS

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	
1	A	378/383 (99%)	352 (93%)	26 (7%)	14	3
1	B	373/383 (97%)	348 (93%)	25 (7%)	15	4
All	All	751/766 (98%)	700 (93%)	51 (7%)	14	4

5 of 51 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	257	THR
1	B	351	LEU
1	B	720	LYS
1	B	261	LEU
1	B	323	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 34 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	505	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	543	GLN
1	B	627	HIS
1	A	543	GLN
1	A	496	GLN

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	5CM	I	8	3,2	18,21,22	1.58	3 (16%)	24,30,33	1.22	3 (12%)
2	5CM	G	8	3,2	18,21,22	1.69	3 (16%)	24,30,33	1.28	3 (12%)
2	5CM	I	6	3,2	18,21,22	1.65	3 (16%)	24,30,33	1.19	2 (8%)
2	5CM	G	6	3,2	18,21,22	1.67	3 (16%)	24,30,33	1.28	3 (12%)

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	5CM	I	8	3,2	-	0/7/21/22	0/2/2/2
2	5CM	G	8	3,2	-	0/7/21/22	0/2/2/2
2	5CM	I	6	3,2	-	0/7/21/22	0/2/2/2
2	5CM	G	6	3,2	-	0/7/21/22	0/2/2/2

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	G	8	5CM	C5-C4	5.94	1.48	1.44
2	G	6	5CM	C5-C4	5.74	1.48	1.44
2	I	6	5CM	C5-C4	5.68	1.48	1.44
2	I	8	5CM	C5-C4	5.36	1.48	1.44
2	G	6	5CM	C6-N1	-2.65	1.33	1.38

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	G	6	5CM	C5-C6-N1	-3.42	119.60	123.31
2	G	8	5CM	C5-C6-N1	-3.39	119.63	123.31
2	I	8	5CM	C5-C6-N1	-3.24	119.80	123.31
2	I	6	5CM	C5-C6-N1	-3.22	119.81	123.31
2	G	8	5CM	C5-C4-N3	-2.90	118.78	121.75

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 2 ligands modelled in this entry, 2 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 ⓘ

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	497/499 (99%)	0.51	39 (7%) 19 21	22, 35, 62, 91	0
1	B	489/499 (97%)	0.43	30 (6%) 27 30	23, 36, 65, 119	0
2	G	13/17 (76%)	-0.36	1 (7%) 19 22	26, 30, 41, 58	0
2	I	15/17 (88%)	-0.31	0 100 100	24, 28, 67, 102	0
3	H	16/17 (94%)	0.39	1 (6%) 26 29	36, 45, 95, 103	0
3	J	17/17 (100%)	0.31	0 100 100	31, 41, 75, 92	0
All	All	1047/1066 (98%)	0.45	71 (6%) 23 26	22, 36, 65, 119	0

The worst 5 of 71 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	524	ALA	6.8
1	B	697	ALA	5.5
1	A	525	HIS	4.5
1	A	285	ALA	4.3
1	A	230	MET	4.2

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(Å ²)	Q<0.9
2	5CM	I	8	20/21	0.97	0.06	26,26,27,27	0
2	5CM	G	6	20/21	0.97	0.06	25,26,26,26	0
2	5CM	I	6	20/21	0.98	0.05	25,25,25,26	0
2	5CM	G	8	20/21	0.98	0.06	28,28,29,29	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
4	MG	B	801	1/1	0.60	0.16	75,75,75,75	0
4	MG	J	101	1/1	0.98	0.10	43,43,43,43	0

6.5 Other polymers [i](#)

There are no such residues in this entry.