wwPDB X-ray Structure Validation Summary Report (i)

Jun 13, 2024 – 09:34 AM EDT

PDB ID	:	4GJR
Title	:	Crystal structure of the TAL effector dHax3 bound to methylated dsDNA
Authors	:	Yan, N.; Deng, D.; Yan, C.Y.; Yin, P.; Pan, X.J.; Shi, Y.G.
Deposited on	:	2012-08-10
Resolution	:	1.85 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 (i) 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 (i)) were used in the production of this report:

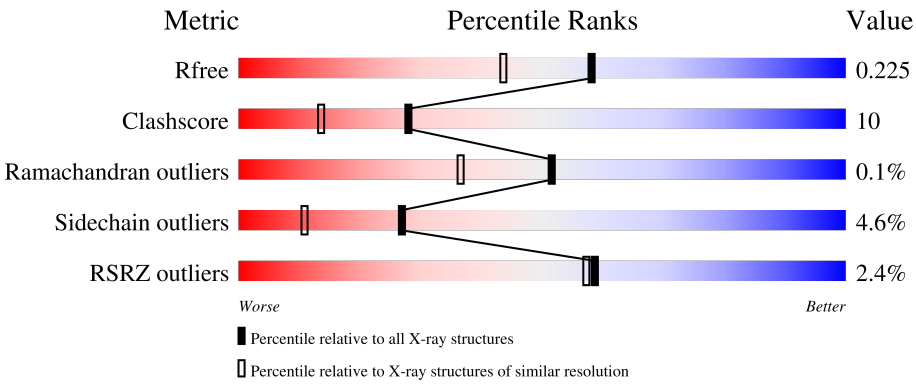
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.36.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 1.85 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	2469 (1.86-1.86)
Clashscore	141614	2625 (1.86-1.86)
Ramachandran outliers	138981	2592 (1.86-1.86)
Sidechain outliers	138945	2592 (1.86-1.86)
RSRZ outliers	127900	2436 (1.86-1.86)

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	499	3% 77% 19% • •
1	B	499	2% 82% 15% • •
2	G	17	6% 59% 35% 6%
2	I	17	59% 41%
3	H	17	6% 53% 47%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	J	17	6% 59% 35% 6%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 9340 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
			Total	C	N	O	S			
1	A	489	3546	2220	656	658	12	0	1	0
1	B	487	3575	2232	666	664	13	0	7	0

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	539	GLY	SER	engineered mutation	UNP Q3ZD72
A	572	HIS	ASN	engineered mutation	UNP Q3ZD72
A	573	ASP	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	539	GLY	SER	engineered mutation	UNP Q3ZD72
B	572	HIS	ASN	engineered mutation	UNP Q3ZD72
B	573	ASP	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*(5CM)P*TP*AP*(5CM)P*CP*TP*CP*(5CM)P*CP*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	I	17	334	164	49	105	16	0	0	0
			Total	C	N	O	P			
2	G	17	333	163	49	105	16	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	J	17	360	169	80	95	16	0	0	0
3	H	17	360	169	80	95	16	0	0	0

- Molecule 4 is water.

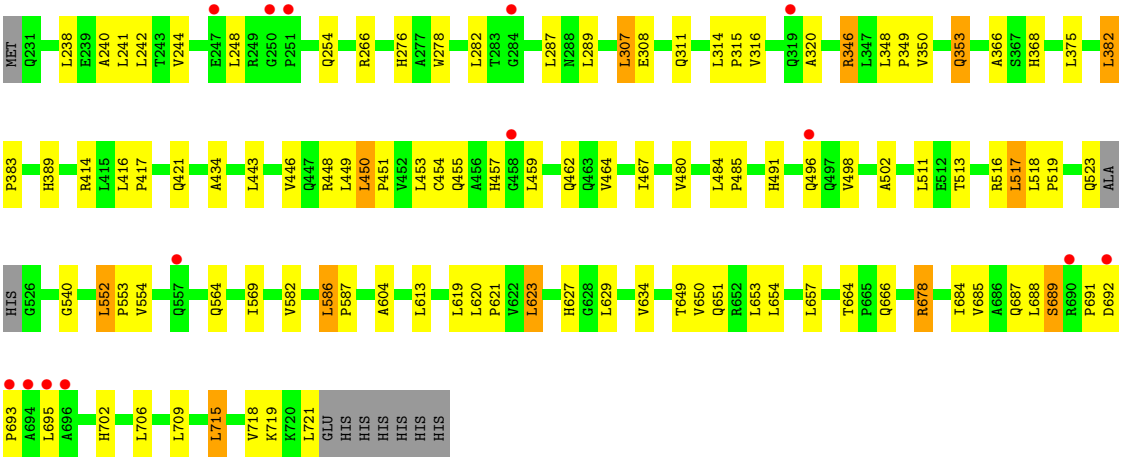
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	296	296	296	0	0
4	I	52	52	52	0	0
4	J	49	49	49	0	0
4	B	321	321	321	0	0
4	G	62	62	62	0	0
4	H	52	52	52	0	0

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.

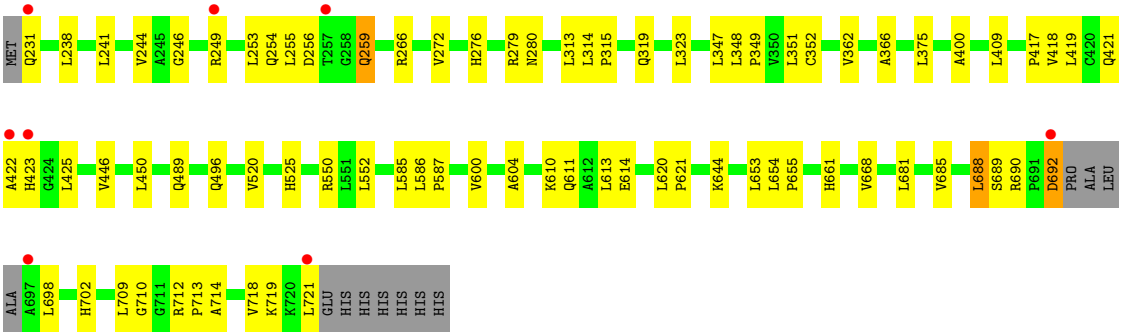
• Molecule 1: Hax3

Chain A:



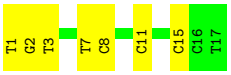
• Molecule 1: Hax3

Chain B:



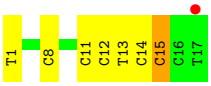
• Molecule 2: DNA (5'-D(*TP*GP*TP*CP*CP*CP*TP*(5CM)P*TP*AP*(5CM)P*CP*TP*CP *(5CM)P*CP*T)-3')

Chain I:



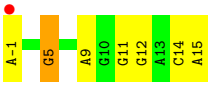
• Molecule 2: DNA (5'-D(*TP*GP*TP*CP*CP*CP*TP*(5CM)P*TP*AP*(5CM)P*CP*TP*CP*(5CM)P*CP*T)-3')

Chain G:



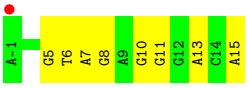
• Molecule 3: DNA (5'-D(*AP*GP*GP*GP*AP*GP*GP*TP*AP*GP*AP*GP*GP*GP*AP*CP*A)-3')

Chain J:



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

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	81.84Å 87.62Å 88.45Å 90.00° 102.85° 90.00°	Depositor
Resolution (Å)	32.37 – 1.85 32.37 – 1.85	Depositor EDS
% Data completeness (in resolution range)	99.2 (32.37-1.85) 99.2 (32.37-1.85)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.66 (at 1.85Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.1_743)	Depositor
R, R_{free}	0.198 , 0.230 0.195 , 0.225	Depositor DCC
R_{free} test set	5156 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å^2)	27.7	Xtriage
Anisotropy	0.070	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 48.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	9340	wwPDB-VP
Average B, all atoms (Å^2)	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.59% 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: 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	0.31	0/3597	0.48	0/4913
1	B	0.34	0/3625	0.50	0/4947
2	G	0.78	0/299	1.43	2/450 (0.4%)
2	I	0.70	0/300	1.39	2/452 (0.4%)
3	H	0.69	0/408	1.25	3/631 (0.5%)
3	J	0.68	0/408	1.27	2/631 (0.3%)
All	All	0.41	0/8637	0.71	9/12024 (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(^{o})	Ideal(^{o})
3	J	5	DG	O4'-C4'-C3'	-8.61	100.83	106.00
3	J	5	DG	C1'-O4'-C4'	-6.45	103.65	110.10
2	G	13	DT	O4'-C1'-N1	-6.31	103.59	108.00
3	H	10	DG	O4'-C1'-N9	-6.12	103.71	108.00
3	H	13	DA	O4'-C1'-N9	-5.98	103.81	108.00

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	3546	0	3692	93	0
1	B	3575	0	3715	64	0
2	G	333	0	198	6	0
2	I	334	0	201	1	0
3	H	360	0	190	8	0
3	J	360	0	190	6	0
4	A	296	0	0	8	1
4	B	321	0	0	4	1
4	G	62	0	0	1	0
4	H	52	0	0	2	0
4	I	52	0	0	0	0
4	J	49	0	0	4	0
All	All	9340	0	8186	162	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 10.

The worst 5 of 162 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:692:ASP:OD1	1:A:693:PRO:HD2	1.47	1.15
1:A:484:LEU:HB3	1:A:485:PRO:HD3	1.49	0.93
1:A:721:LEU:HD12	1:B:709:LEU:HD21	1.50	0.92
1:A:552:LEU:O	1:A:552:LEU:HD13	1.69	0.91
1:A:709:LEU:HD23	1:B:721:LEU:CD1	2.06	0.86

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)
4:A:947:HOH:O	4:B:888:HOH:O[1_656]	2.07	0.13

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	486/499 (97%)	467 (96%)	18 (4%)	1 (0%)	47	33
1	B	490/499 (98%)	479 (98%)	11 (2%)	0	100	100
All	All	976/998 (98%)	946 (97%)	29 (3%)	1 (0%)	51	36

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	691	PRO

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	374/383 (98%)	352 (94%)	22 (6%)	19	6
1	B	377/383 (98%)	365 (97%)	12 (3%)	39	22
All	All	751/766 (98%)	717 (96%)	34 (4%)	27	11

5 of 34 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	347	LEU
1	B	351	LEU
1	B	688	LEU
1	A	552	LEU
1	A	523	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	564	GLN
1	B	447	GLN
1	B	702	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	470	ASN
1	A	462	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)

6 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	G	15	3,2	18,21,22	1.66	3 (16%)	24,30,33	1.16	2 (8%)
2	5CM	I	15	3,2	18,21,22	1.66	3 (16%)	24,30,33	1.18	2 (8%)
2	5CM	G	11	3,2	18,21,22	1.67	3 (16%)	24,30,33	1.24	2 (8%)
2	5CM	I	8	3,2	18,21,22	1.71	3 (16%)	24,30,33	1.20	2 (8%)
2	5CM	G	8	3,2	18,21,22	1.72	3 (16%)	24,30,33	1.19	2 (8%)
2	5CM	I	11	3,2	18,21,22	1.62	3 (16%)	24,30,33	1.21	2 (8%)

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	G	15	3,2	-	0/7/21/22	0/2/2/2
2	5CM	I	15	3,2	-	0/7/21/22	0/2/2/2
2	5CM	G	11	3,2	-	0/7/21/22	0/2/2/2
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	11	3,2	-	0/7/21/22	0/2/2/2

The worst 5 of 18 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	G	8	5CM	C5-C4	6.13	1.48	1.44
2	I	8	5CM	C5-C4	6.05	1.48	1.44
2	G	11	5CM	C5-C4	5.85	1.48	1.44
2	I	15	5CM	C5-C4	5.75	1.48	1.44
2	G	15	5CM	C5-C4	5.71	1.48	1.44

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	G	15	5CM	C5-C6-N1	-3.62	119.38	123.31
2	G	11	5CM	C5-C6-N1	-3.56	119.45	123.31
2	I	8	5CM	C5-C6-N1	-3.51	119.50	123.31
2	I	15	5CM	C5-C6-N1	-3.48	119.53	123.31
2	G	8	5CM	C5-C6-N1	-3.45	119.57	123.31

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	G	15	5CM	1	0

5.5 Carbohydrates (i)

There are no monosaccharides 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	489/499 (97%)	-0.03	14 (2%)	51	50	19, 33, 59, 102	0
1	B	487/499 (97%)	-0.11	8 (1%)	72	72	16, 29, 55, 88	0
2	G	14/17 (82%)	-0.25	1 (7%)	16	15	17, 22, 48, 80	0
2	I	14/17 (82%)	-0.54	0	100	100	19, 21, 62, 70	0
3	H	17/17 (100%)	0.02	1 (5%)	22	22	27, 32, 79, 86	0
3	J	17/17 (100%)	-0.10	1 (5%)	22	22	23, 34, 65, 71	0
All	All	1038/1066 (97%)	-0.08	25 (2%)	59	57	16, 30, 59, 102	0

The worst 5 of 25 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	695	LEU	5.0
1	B	423[A]	HIS	4.9
1	A	690	ARG	4.4
1	A	693	PRO	4.4
1	B	257	THR	3.8

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	G	15	20/21	0.93	0.10	29,30,35,38	0
2	5CM	I	15	20/21	0.94	0.10	29,30,33,36	0
2	5CM	I	8	20/21	0.97	0.10	18,19,20,21	0
2	5CM	I	11	20/21	0.97	0.08	19,22,24,24	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	5CM	G	11	20/21	0.98	0.11	17,20,22,23	0
2	5CM	G	8	20/21	0.98	0.10	16,17,18,20	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.