



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 22, 2016 – 08:14 AM EDT

PDB ID : 5JPN  
Title : Structure of human complement C4 rebuilt using iMDFP  
Authors : Croll, T.I.; Andersen, G.R.  
Deposited on : 2016-05-03  
Resolution : 3.60 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](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity : 4.02b-467  
Mogul : 1.7.1 (RC1), CSD as537be (2016)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20027939  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20027939

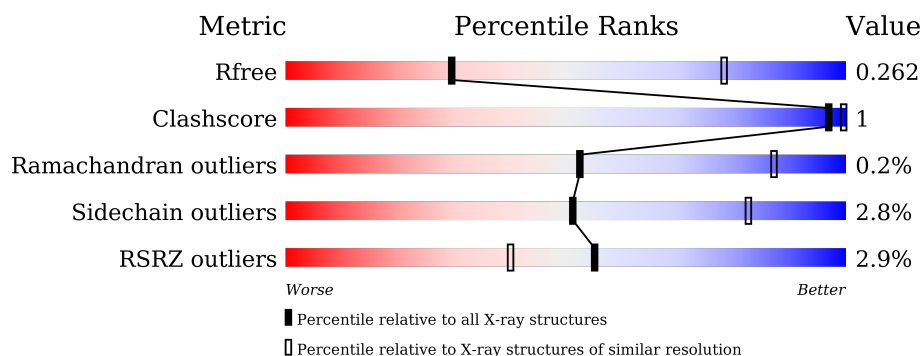
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.60 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))
$R_{free}$	91344	1408 (3.80-3.40)
Clashscore	102246	1010 (3.74-3.46)
Ramachandran outliers	100387	1007 (3.76-3.44)
Sidechain outliers	100360	1007 (3.76-3.44)
RSRZ outliers	91569	1003 (3.78-3.42)

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  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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	656	<div> <div>4%</div> <div> <div></div> <div>93%</div> <div>6%</div> </div> </div>
2	B	767	<div> <div>3%</div> <div> <div></div> <div>87%</div> <div>9%</div> </div> </div>
3	C	290	<div> <div>%</div> <div> <div></div> <div>92%</div> <div>8%</div> </div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
8	A2G	B	1508	-	-	-	X

## 2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 12891 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 Complement C4-A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	651	Total	C	N	O	S	0	0	0
			5012	3185	872	939	16			

- Molecule 2 is a protein called Complement C4-A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	697	Total	C	N	O	S	0	0	0
			5361	3371	943	1023	24			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1201	SER	THR	variant	UNP P0C0L4

- Molecule 3 is a protein called Complement C4-A.

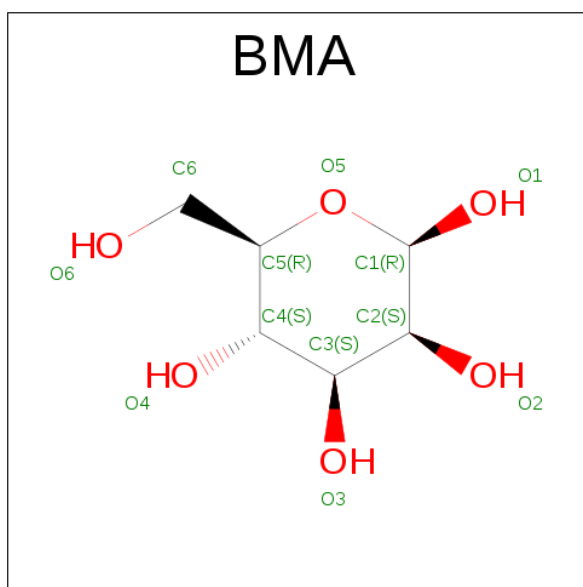
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	290	Total	C	N	O	S	0	0	0
			2311	1451	411	432	17			

- Molecule 4 is N-ACETYL-D-GLUCOSAMINE (three-letter code: NAG) (formula: C<sub>8</sub>H<sub>15</sub>NO<sub>6</sub>).



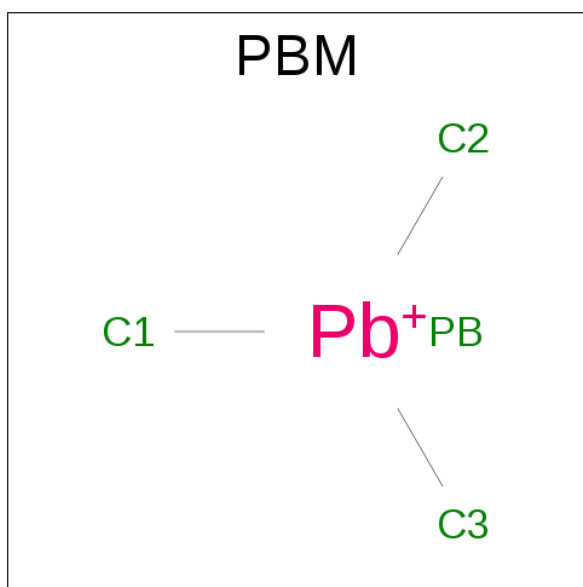
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total	C	N	O	0	0
			14	8	1	5		
4	A	1	Total	C	N	O	0	0
			14	8	1	5		
4	B	1	Total	C	N	O	0	0
			14	8	1	5		
4	B	1	Total	C	N	O	0	0
			14	8	1	5		
4	B	1	Total	C	N	O	0	0
			14	8	1	5		
4	B	1	Total	C	N	O	0	0
			14	8	1	5		
4	B	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 5 is BETA-D-MANNOSE (three-letter code: BMA) (formula: C<sub>6</sub>H<sub>12</sub>O<sub>6</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	C	O	0	0
			11	6	5		
5	B	1	Total	C	O	0	0
			11	6	5		
5	B	1	Total	C	O	0	0
			11	6	5		

- Molecule 6 is TRIMETHYL LEAD ION (three-letter code: PBM) (formula:  $C_3H_9Pb$ ).



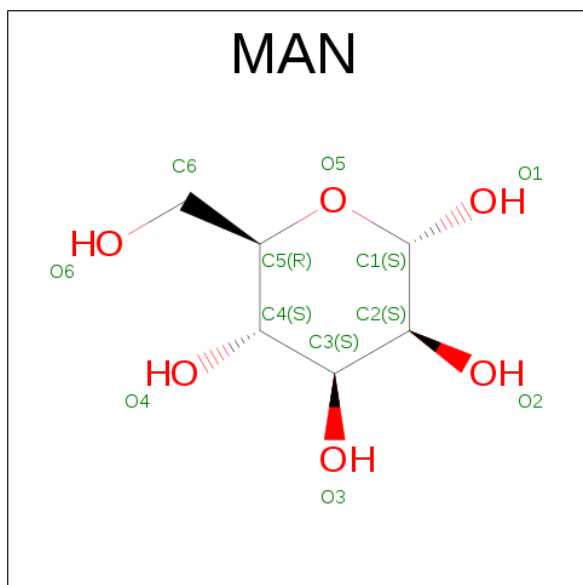
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	1	Total	Pb	0	0
			1	1		

*Continued on next page...*

Continued from previous page...

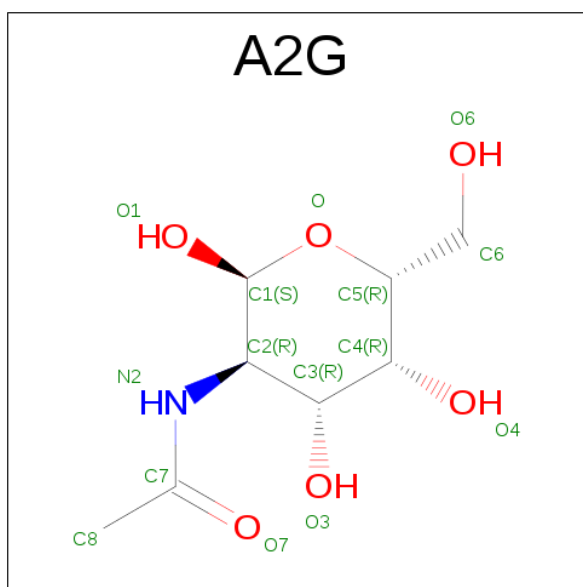
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	B	1	Total	Pb	0	0
			1	1		
6	B	1	Total	Pb	0	0
			1	1		
6	C	1	Total	Pb	0	0
			1	1		

- Molecule 7 is ALPHA-D-MANNOSE (three-letter code: MAN) (formula: C<sub>6</sub>H<sub>12</sub>O<sub>6</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
7	B	1	Total	C	O	0	0
			11	6	5		
7	B	1	Total	C	O	0	0
			11	6	5		
7	B	1	Total	C	O	0	0
			11	6	5		
7	B	1	Total	C	O	0	0
			11	6	5		

- Molecule 8 is N-ACETYL-2-DEOXY-2-AMINO-GALACTOSE (three-letter code: A2G) (formula: C<sub>8</sub>H<sub>15</sub>NO<sub>6</sub>).



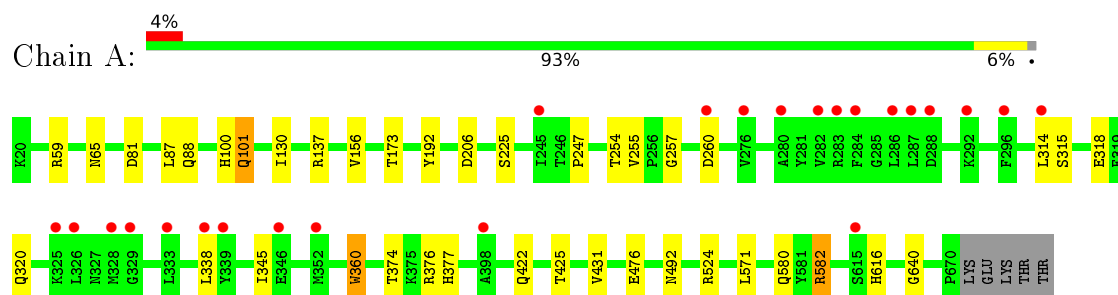
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
8	B	1	14	8	1	5	0	0



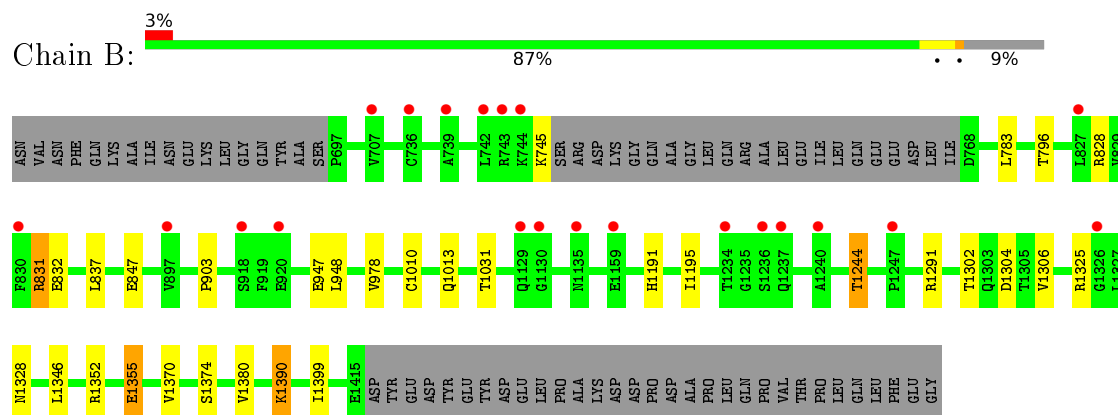
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

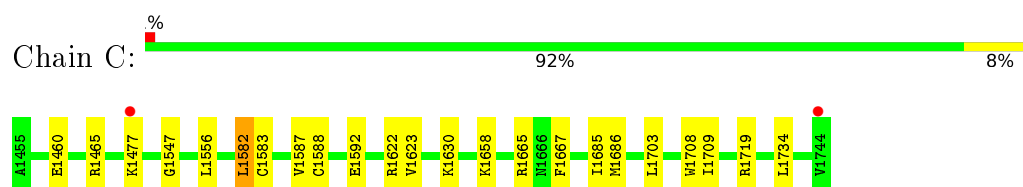
#### • Molecule 1: Complement C4-A



#### • Molecule 2: Complement C4-A



#### • Molecule 3: Complement C4-A



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	85.45Å 103.31Å 256.03Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.90 – 3.60 45.90 – 3.60	Depositor EDS
% Data completeness (in resolution range)	99.6 (45.90-3.60) 99.7 (45.90-3.60)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	9.70	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.48 (at 3.57Å)	Xtriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.215 , 0.263 0.213 , 0.262	Depositor DCC
$R_{free}$ test set	1347 reflections (5.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	113.2	Xtriage
Anisotropy	0.501	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 85.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	12891	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	130.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.66% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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: PBM, BMA, NAG, A2G, MAN

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	0.25	0/5128	0.46	0/6961
2	B	0.24	0/5472	0.44	1/7436 (0.0%)
3	C	0.24	0/2359	0.48	0/3188
All	All	0.25	0/12959	0.45	1/17585 (0.0%)

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	1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	783	LEU	CA-CB-CG	6.06	129.23	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	582	ARG	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	5012	0	5025	16	0
2	B	5361	0	5310	12	0
3	C	2311	0	2253	7	0
4	A	28	0	24	0	0
4	B	84	0	73	0	0
5	A	11	0	10	0	0
5	B	22	0	16	0	0
6	A	1	0	0	0	0
6	B	2	0	0	0	0
6	C	1	0	0	0	0
7	B	44	0	40	0	0
8	B	14	0	13	0	0
All	All	12891	0	12764	34	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 1.

The worst 5 of 34 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:476:GLU:HB3	1:A:492:ASN:HB2	1.71	0.72
1:A:100:HIS:CG	1:A:101:GLN:H	2.08	0.71
2:B:1325:ARG:HB3	2:B:1374:SER:HB2	1.89	0.55
1:A:81:ASP:N	1:A:81:ASP:OD1	2.39	0.53
2:B:1013:GLN:N	2:B:1013:GLN:OE1	2.38	0.49

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	649/656 (99%)	618 (95%)	30 (5%)	1 (0%)	52	87
2	B	693/767 (90%)	661 (95%)	30 (4%)	2 (0%)	46	83
3	C	288/290 (99%)	272 (94%)	15 (5%)	1 (0%)	46	83
All	All	1630/1713 (95%)	1551 (95%)	75 (5%)	4 (0%)	52	87

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	1390	LYS
1	A	101	GLN
3	C	1622	ARG
2	B	1244	THR

### 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	557/562 (99%)	542 (97%)	15 (3%)	52	83
2	B	581/641 (91%)	568 (98%)	13 (2%)	60	86
3	C	248/248 (100%)	237 (96%)	11 (4%)	35	74
All	All	1386/1451 (96%)	1347 (97%)	39 (3%)	51	82

5 of 39 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	828	ARG
2	B	1031	THR
3	C	1665	ARG
2	B	831	ARG
2	B	837	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 20 ligands modelled in this entry, 4 are modelled with single atom - leaving 16 for Mogul analysis.

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	$\# Z  > 2$
4	NAG	A	701	1,4	14,14,15	0.29	0	15,19,21	0.33	0
4	NAG	A	702	5,4	14,14,15	0.21	0	15,19,21	0.54	0
5	BMA	A	703	4	11,11,12	0.58	0	15,15,17	0.73	0
4	NAG	B	1501	2,4	14,14,15	0.61	1 (7%)	15,19,21	0.65	0
4	NAG	B	1502	4	14,14,15	0.29	0	15,19,21	0.36	0
4	NAG	B	1503	2,4	14,14,15	0.28	0	15,19,21	0.47	0
4	NAG	B	1504	5,4	14,14,15	0.38	0	15,19,21	0.49	0
5	BMA	B	1505	4,7	11,11,12	0.69	0	15,15,17	0.93	0
7	MAN	B	1506	5	11,11,12	0.77	0	15,15,17	0.83	1 (6%)
7	MAN	B	1507	5	11,11,12	0.63	0	15,15,17	0.93	2 (13%)
8	A2G	B	1508	2	14,14,15	1.68	4 (28%)	15,19,21	1.18	1 (6%)
4	NAG	B	1509	2,4	14,14,15	0.42	0	15,19,21	0.45	0
4	NAG	B	1510	5,4	14,14,15	0.19	0	15,19,21	0.63	1 (6%)
5	BMA	B	1511	4,7	11,11,12	0.52	0	15,15,17	0.76	0
7	MAN	B	1512	5	11,11,12	0.70	0	15,15,17	0.87	1 (6%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
7	MAN	B	1513	5	11,11,12	0.70	0	15,15,17	0.87	1 (6%)

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
4	NAG	A	701	1,4	-	0/6/23/26	0/1/1/1
4	NAG	A	702	5,4	-	0/6/23/26	0/1/1/1
5	BMA	A	703	4	-	0/2/19/22	0/1/1/1
4	NAG	B	1501	2,4	-	0/6/23/26	0/1/1/1
4	NAG	B	1502	4	-	0/6/23/26	0/1/1/1
4	NAG	B	1503	2,4	-	0/6/23/26	0/1/1/1
4	NAG	B	1504	5,4	-	0/6/23/26	0/1/1/1
5	BMA	B	1505	4,7	-	0/2/19/22	0/1/1/1
7	MAN	B	1506	5	-	0/2/19/22	0/1/1/1
7	MAN	B	1507	5	-	0/2/19/22	0/1/1/1
8	A2G	B	1508	2	-	0/6/23/26	0/1/1/1
4	NAG	B	1509	2,4	-	0/6/23/26	0/1/1/1
4	NAG	B	1510	5,4	-	0/6/23/26	0/1/1/1
5	BMA	B	1511	4,7	-	0/2/19/22	0/1/1/1
7	MAN	B	1512	5	-	0/2/19/22	0/1/1/1
7	MAN	B	1513	5	-	0/2/19/22	0/1/1/1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
8	B	1508	A2G	C3-C2	-2.05	1.47	1.52
4	B	1501	NAG	O5-C1	-2.03	1.40	1.43
8	B	1508	A2G	O3-C3	2.13	1.48	1.43
8	B	1508	A2G	O-C1	2.28	1.47	1.43
8	B	1508	A2G	C7-N2	2.93	1.45	1.34

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
8	B	1508	A2G	C2-N2-C7	-2.54	119.80	123.11
7	B	1506	MAN	O2-C2-C3	-2.27	105.61	110.19
7	B	1513	MAN	O2-C2-C3	-2.26	105.64	110.19
7	B	1512	MAN	O2-C2-C3	-2.24	105.67	110.19
7	B	1507	MAN	O2-C2-C3	-2.19	105.77	110.19

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

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	651/656 (99%)	0.15	24 (3%) 45 32	77, 131, 198, 234	0
2	B	697/767 (90%)	0.04	21 (3%) 54 38	76, 118, 190, 231	0
3	C	290/290 (100%)	-0.05	2 (0%) 89 81	79, 118, 162, 191	0
All	All	1638/1713 (95%)	0.07	47 (2%) 55 40	76, 123, 191, 234	0

The worst 5 of 47 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	918	SER	6.0
2	B	743	ARG	5.2
2	B	1237	GLN	4.0
2	B	739	ALA	3.5
1	A	615	SER	3.5

### 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](#)

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
8	A2G	B	1508	14/15	0.76	0.38	3.68	157,178,183,184	14
4	NAG	B	1503	14/15	0.90	0.26	0.76	153,155,163,175	0
4	NAG	B	1502	14/15	0.70	0.37	-	218,233,237,240	0
4	NAG	A	702	14/15	0.85	0.26	-	181,187,192,199	0
6	PBM	B	1515	1/4	0.49	0.33	-	137,137,137,137	1
6	PBM	A	704	1/4	0.94	0.20	-	134,134,134,134	1
5	BMA	A	703	11/12	0.85	0.21	-	171,178,184,185	0
4	NAG	A	701	14/15	0.94	0.17	-	138,151,167,172	0
4	NAG	B	1501	14/15	0.83	0.18	-	210,214,220,228	0
7	MAN	B	1513	11/12	0.71	0.29	-	192,205,209,211	0
4	NAG	B	1510	14/15	0.92	0.38	-	178,183,185,188	0
6	PBM	B	1514	1/4	0.95	0.07	-	142,142,142,142	1
7	MAN	B	1507	11/12	0.72	0.74	-	176,188,194,195	0
5	BMA	B	1511	11/12	0.77	0.36	-	189,196,206,206	0
7	MAN	B	1512	11/12	0.55	0.60	-	206,211,223,225	0
7	MAN	B	1506	11/12	0.71	0.52	-	174,183,191,194	0
4	NAG	B	1509	14/15	0.86	0.34	-	171,182,193,197	0
4	NAG	B	1504	14/15	0.83	0.26	-	162,194,205,210	0
5	BMA	B	1505	11/12	0.66	0.27	-	190,197,204,208	0
6	PBM	C	1801	1/4	0.88	0.21	-	123,123,123,123	1

## 6.5 Other polymers

There are no such residues in this entry.