



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 05:09 PM GMT

PDB ID : 4HC1  
Title : Crystal structure of a loop deleted mutant of human MAdCAM-1 D1D2 complexed with Fab 10G3  
Authors : Springer, T.; Yu, Y.; Zhu, J.  
Deposited on : 2012-09-28  
Resolution : 2.87 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 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
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) : trunk26865

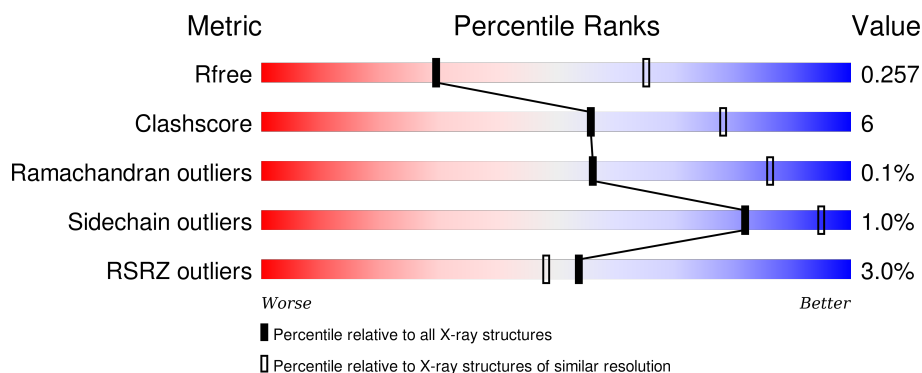
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.87 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	1945 (2.90-2.86)
Clashscore	102246	2202 (2.90-2.86)
Ramachandran outliers	100387	2149 (2.90-2.86)
Sidechain outliers	100360	2152 (2.90-2.86)
RSRZ outliers	91569	1950 (2.90-2.86)

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 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	206	<div> <div>3%</div> <div>81%</div> <div>16%</div> <div>.</div> </div>
1	B	206	<div> <div>3%</div> <div>79%</div> <div>17%</div> <div>.</div> </div>
2	H	220	<div> <div>86%</div> <div>14%</div> </div>
2	M	220	<div> <div>2%</div> <div>89%</div> <div>11%</div> </div>
3	L	214	<div> <div>2%</div> <div>84%</div> <div>15%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain	
			6%	10%
3	N	214		

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 18948 atoms, of which 9300 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 Mucosal addressin cell adhesion molecule 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	198	Total	C	H	N	O	S	0	1	0
			2939	922	1468	265	277	7			
1	B	198	Total	C	H	N	O	S	0	1	0
			2918	922	1448	264	277	7			

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	94	ASN	ASP	ENGINEERED MUTATION	UNP Q13477
A	148	GLN	GLU	ENGINEERED MUTATION	UNP Q13477
A	?	-	GLU	DELETION	UNP Q13477
A	?	-	GLU	DELETION	UNP Q13477
A	?	-	GLU	DELETION	UNP Q13477
A	?	-	PRO	DELETION	UNP Q13477
A	?	-	GLN	DELETION	UNP Q13477
A	?	-	GLY	DELETION	UNP Q13477
A	?	-	GLN	DELETION	UNP Q13477
A	?	-	GLY	DELETION	UNP Q13477
A	150	PRO	-	INSERTION	UNP Q13477
A	151	ILE	-	INSERTION	UNP Q13477
A	152	GLY	-	INSERTION	UNP Q13477
A	153	GLY	-	INSERTION	UNP Q13477
A	203	GLY	-	EXPRESSION TAG	UNP Q13477
A	204	GLY	-	EXPRESSION TAG	UNP Q13477
A	205	GLU	-	EXPRESSION TAG	UNP Q13477
A	206	ASN	-	EXPRESSION TAG	UNP Q13477
A	207	LEU	-	EXPRESSION TAG	UNP Q13477
A	208	TYR	-	EXPRESSION TAG	UNP Q13477
A	209	PHE	-	EXPRESSION TAG	UNP Q13477
A	210	GLN	-	EXPRESSION TAG	UNP Q13477
B	94	ASN	ASP	ENGINEERED MUTATION	UNP Q13477
B	148	GLN	GLU	ENGINEERED MUTATION	UNP Q13477
B	?	-	GLU	DELETION	UNP Q13477

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	?	-	GLU	DELETION	UNP Q13477
B	?	-	GLU	DELETION	UNP Q13477
B	?	-	PRO	DELETION	UNP Q13477
B	?	-	GLN	DELETION	UNP Q13477
B	?	-	GLY	DELETION	UNP Q13477
B	?	-	GLN	DELETION	UNP Q13477
B	?	-	GLY	DELETION	UNP Q13477
B	150	PRO	-	INSERTION	UNP Q13477
B	151	ILE	-	INSERTION	UNP Q13477
B	152	GLY	-	INSERTION	UNP Q13477
B	153	GLY	-	INSERTION	UNP Q13477
B	203	GLY	-	EXPRESSION TAG	UNP Q13477
B	204	GLY	-	EXPRESSION TAG	UNP Q13477
B	205	GLU	-	EXPRESSION TAG	UNP Q13477
B	206	ASN	-	EXPRESSION TAG	UNP Q13477
B	207	LEU	-	EXPRESSION TAG	UNP Q13477
B	208	TYR	-	EXPRESSION TAG	UNP Q13477
B	209	PHE	-	EXPRESSION TAG	UNP Q13477
B	210	GLN	-	EXPRESSION TAG	UNP Q13477

- Molecule 2 is a protein called 10G3 heavy chain.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	H	220	Total	C	H	N	O	S	0	1	0
			3288	1061	1612	276	333	6			
2	M	220	Total	C	H	N	O	S	0	1	0
			3295	1061	1621	275	332	6			

- Molecule 3 is a protein called 10G3 light chain.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	L	213	Total	C	H	N	O	S	0	0	0
			3218	1030	1569	277	335	7			
3	N	213	Total	C	H	N	O	S	0	0	0
			3218	1030	1569	277	335	7			

- Molecule 4 is SUGAR (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	H	N	O	0	0
			27	8	13	1	5		
4	B	1	Total	C	N	O		0	0
			14	8	1	5			

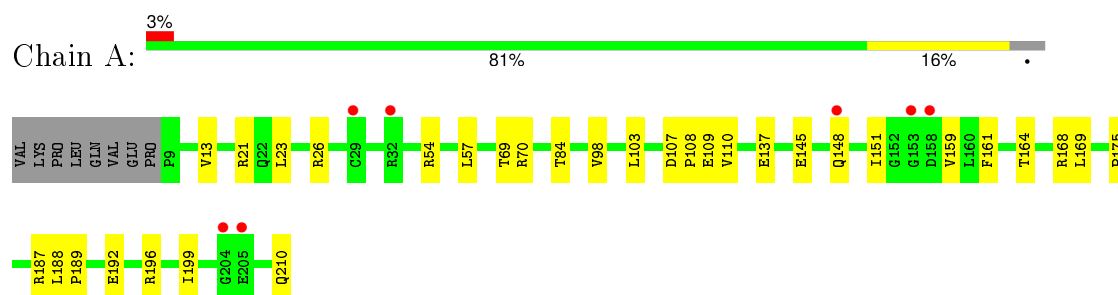
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	5	Total	O	0	0
			5	5		
5	H	1	Total	O	0	0
			1	1		
5	L	7	Total	O	0	0
			7	7		
5	B	5	Total	O	0	0
			5	5		
5	M	6	Total	O	0	0
			6	6		
5	N	7	Total	O	0	0
			7	7		

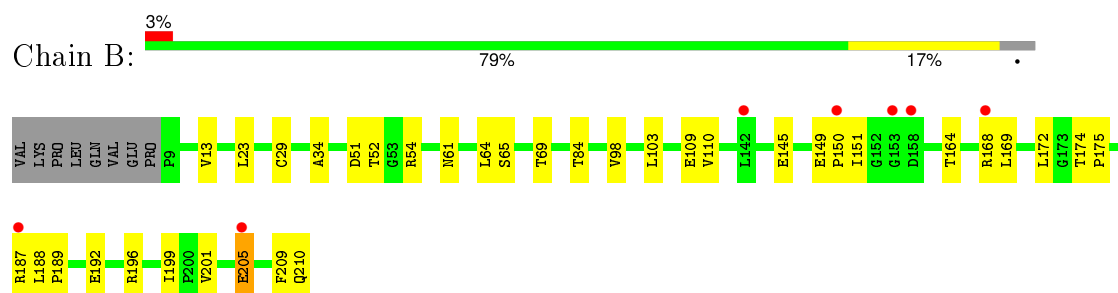
### 3 Residue-property plots [i](#)

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.

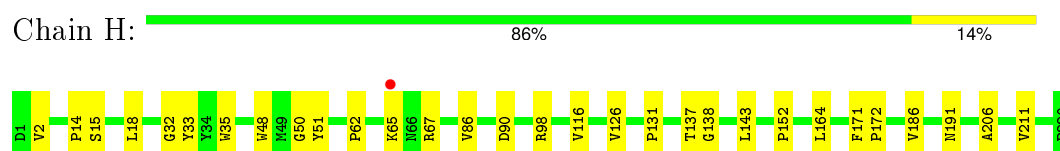
- Molecule 1: Mucosal addressin cell adhesion molecule 1



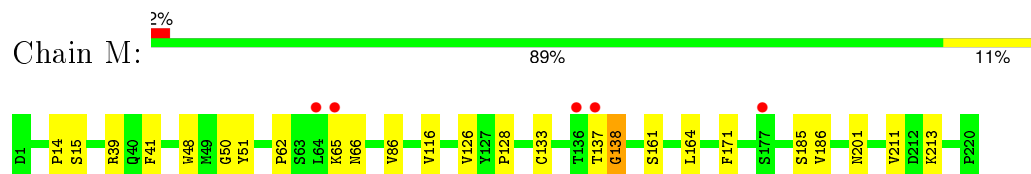
- Molecule 1: Mucosal addressin cell adhesion molecule 1



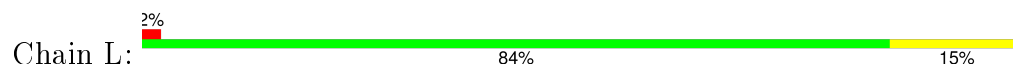
- Molecule 2: 10G3 heavy chain

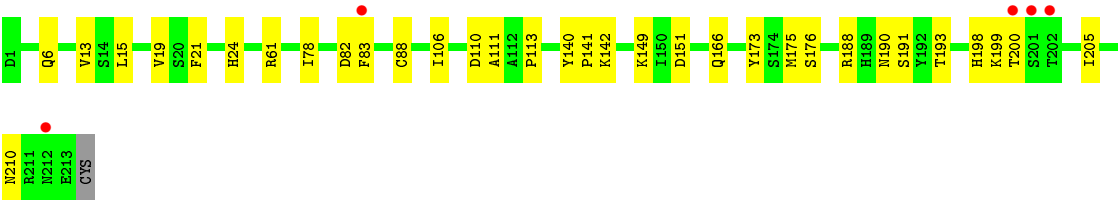


- Molecule 2: 10G3 heavy chain

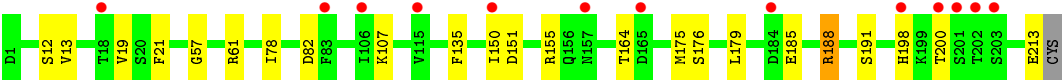
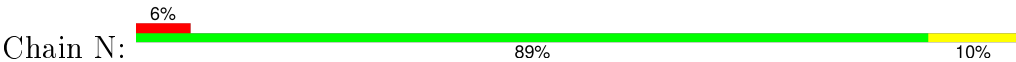


- Molecule 3: 10G3 light chain





● Molecule 3: 10G3 light chain





## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	317.85Å 71.73Å 70.56Å 90.00° 93.29° 90.00°	Depositor
Resolution (Å)	49.33 – 2.87 49.33 – 2.87	Depositor EDS
% Data completeness (in resolution range)	98.0 (49.33-2.87) 98.0 (49.33-2.87)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.41 (at 2.86Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, $R_{free}$	0.200 , 0.253 0.202 , 0.257	Depositor DCC
$R_{free}$ test set	1789 reflections (4.99%)	DCC
Wilson B-factor (Å <sup>2</sup> )	49.0	Xtriage
Anisotropy	0.565	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 39.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	1 of 35873 reflections (0.003%)	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	18948	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	60.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.81% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: NAG

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	0/1501	0.51	0/2049
1	B	0.25	0/1501	0.50	0/2052
2	H	0.24	0/1722	0.49	0/2359
2	M	0.24	0/1723	0.52	1/2360 (0.0%)
3	L	0.26	0/1687	0.46	0/2284
3	N	0.25	0/1687	0.46	0/2284
All	All	0.25	0/9821	0.49	1/13388 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
2	M	138	GLY	C-N-CA	8.32	142.50	121.70

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	1471	1468	1471	22	1
1	B	1470	1448	1468	29	0
2	H	1676	1612	1624	22	1
2	M	1674	1621	1627	16	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	L	1649	1569	1573	21	1
3	N	1649	1569	1573	14	0
4	A	14	13	12	0	0
4	B	14	0	13	2	0
5	A	5	0	0	1	0
5	B	5	0	0	0	0
5	H	1	0	0	0	0
5	L	7	0	0	0	0
5	M	6	0	0	0	0
5	N	7	0	0	0	0
All	All	9648	9300	9361	117	2

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 6.

The worst 5 of 117 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:H:2:VAL:HG21	2:H:98:ARG:HH12	1.38	0.89
3:L:151:ASP:HA	3:L:191:SER:OG	1.75	0.87
1:A:187:ARG:NH1	5:A:405:HOH:O	2.15	0.78
2:H:2:VAL:HG21	2:H:98:ARG:NH1	2.00	0.77
3:L:151:ASP:OD1	3:L:191:SER:OG	2.04	0.75

All (2) 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:191[A]:ASN:OD1	3:L:24:HIS:NE2[4_555]	2.14	0.06
1:A:26:ARG:HH12	1:A:210[A]:GLN:O[1_554]	1.59	0.01

## 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	196/206 (95%)	183 (93%)	12 (6%)	1 (0%)	34	69
1	B	197/206 (96%)	185 (94%)	12 (6%)	0	100	100
2	H	219/220 (100%)	212 (97%)	7 (3%)	0	100	100
2	M	219/220 (100%)	209 (95%)	10 (5%)	0	100	100
3	L	211/214 (99%)	202 (96%)	9 (4%)	0	100	100
3	N	211/214 (99%)	203 (96%)	8 (4%)	0	100	100
All	All	1253/1280 (98%)	1194 (95%)	58 (5%)	1 (0%)	56	86

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	159	VAL

### 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	158/165 (96%)	156 (99%)	2 (1%)	76	93
1	B	158/165 (96%)	154 (98%)	4 (2%)	55	84
2	H	192/191 (100%)	192 (100%)	0	100	100
2	M	192/191 (100%)	191 (100%)	1 (0%)	92	97
3	L	187/188 (100%)	185 (99%)	2 (1%)	80	94
3	N	187/188 (100%)	184 (98%)	3 (2%)	70	90
All	All	1074/1088 (99%)	1062 (99%)	12 (1%)	82	94

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	196	ARG
1	B	205[A]	GLU
3	N	21	PHE

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	54	ARG
2	M	133	CYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	83	HIS
2	H	58	ASN
2	H	176	GLN
3	N	50	HIS

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

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

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

2 ligands 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	$\# Z  > 2$
4	NAG	A	2061	1	14,14,15	0.90	1 (7%)	15,19,21	0.73	0
4	NAG	B	2061	1	14,14,15	0.46	0	15,19,21	1.16	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	2061	1	-	0/6/23/26	0/1/1/1
4	NAG	B	2061	1	-	0/6/23/26	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	2061	NAG	O3-C3	-2.93	1.35	1.43

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	B	2061	NAG	C1-O5-C5	3.67	116.91	112.25

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	B	2061	NAG	2	0

## 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, 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	198/206 (96%)	0.25	7 (3%)	48	41	23, 45, 108, 177	0
1	B	198/206 (96%)	0.37	7 (3%)	48	41	25, 54, 118, 171	0
2	H	220/220 (100%)	0.15	1 (0%)	91	90	21, 42, 77, 135	0
2	M	220/220 (100%)	0.22	5 (2%)	64	60	19, 48, 126, 172	0
3	L	213/214 (99%)	0.16	5 (2%)	64	60	18, 44, 93, 150	0
3	N	213/214 (99%)	0.48	13 (6%)	25	18	19, 59, 112, 151	0
All	All	1262/1280 (98%)	0.27	38 (3%)	54	48	18, 49, 106, 177	0

The worst 5 of 38 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	L	201	SER	6.0
1	A	158	ASP	5.1
3	N	202	THR	4.9
3	L	83	PHE	4.2
1	B	153	GLY	4.1

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

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

### 6.3 Carbohydrates ⓘ

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
4	NAG	A	2061	14/15	0.86	0.19	-	77,97,121,122	0
4	NAG	B	2061	14/15	0.77	0.38	-	37,44,48,49	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.