



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 19, 2016 – 10:56 PM GMT

PDB ID : 5F1B  
Title : Structural basis of Ebola virus entry: viral glycoprotein bound to its endosomal receptor Niemann-Pick C1  
Authors : Wang, H.; Shi, Y.; Song, J.; Qi, J.; Lu, G.; Yan, J.; Gao, G.F.  
Deposited on : 2015-11-30  
Resolution : 2.30 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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-20026982  
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-20026982

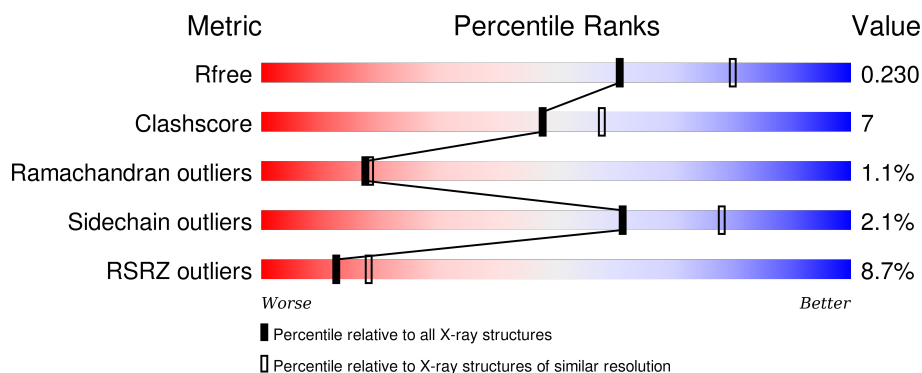
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.30 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	3852 (2.30-2.30)
Clashscore	102246	4452 (2.30-2.30)
Ramachandran outliers	100387	4410 (2.30-2.30)
Sidechain outliers	100360	4409 (2.30-2.30)
RSRZ outliers	91569	3857 (2.30-2.30)

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	158	<div> <div>12%</div> <div>85%</div> <div>13%</div> </div>
2	B	130	<div> <div>6%</div> <div>56%</div> <div>9%</div> <div>33%</div> </div>
3	C	256	<div> <div>5%</div> <div>71%</div> <div>11%</div> <div>16%</div> </div>

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 3828 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 GP1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	158	Total	C	N	O	S	0	0	0
			1194	757	209	223	5			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	31	ARG	-	expression tag	UNP P87666
A	42	VAL	THR	engineered mutation	UNP P87666

- Molecule 2 is a protein called GP2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	87	Total	C	N	O	S	0	0	0
			687	441	121	122	3			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	633	HIS	-	expression tag	UNP P87666
B	634	HIS	-	expression tag	UNP P87666
B	635	HIS	-	expression tag	UNP P87666
B	636	HIS	-	expression tag	UNP P87666
B	637	HIS	-	expression tag	UNP P87666
B	638	HIS	-	expression tag	UNP P87666

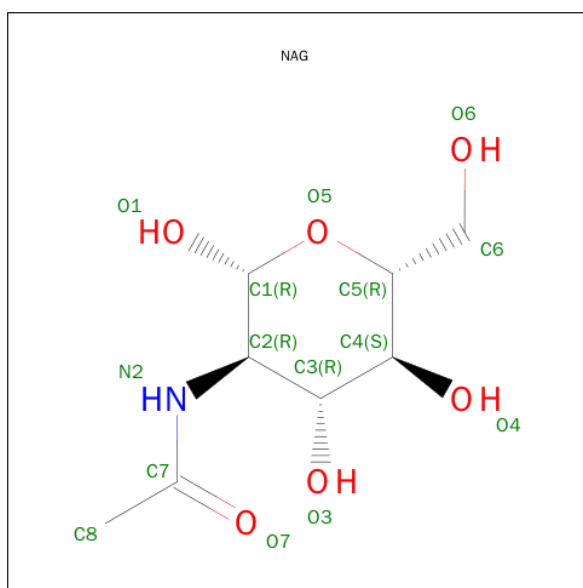
- Molecule 3 is a protein called Niemann-Pick C1 protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	215	Total	C	N	O	S	0	0	0
			1748	1134	281	329	4			

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	1	MET	-	expression tag	UNP O15118
C	249	LEU	-	expression tag	UNP O15118
C	250	GLU	-	expression tag	UNP O15118
C	251	HIS	-	expression tag	UNP O15118
C	252	HIS	-	expression tag	UNP O15118
C	253	HIS	-	expression tag	UNP O15118
C	254	HIS	-	expression tag	UNP O15118
C	255	HIS	-	expression tag	UNP O15118
C	256	HIS	-	expression tag	UNP O15118

- 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	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		

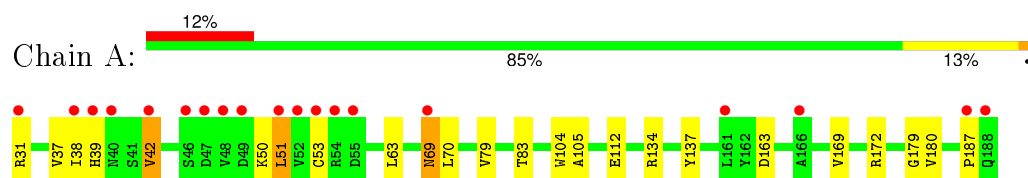
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	77	Total	O	0	0
			77	77		
5	B	31	Total	O	0	0
			31	31		
5	C	63	Total	O	0	0
			63	63		

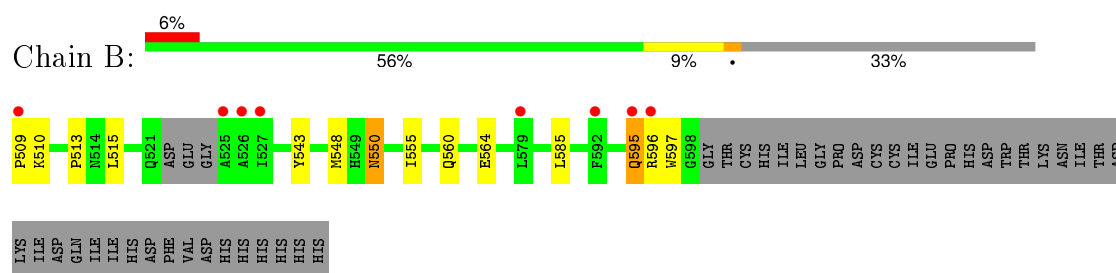
### 3 Residue-property plots [i](#)

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.

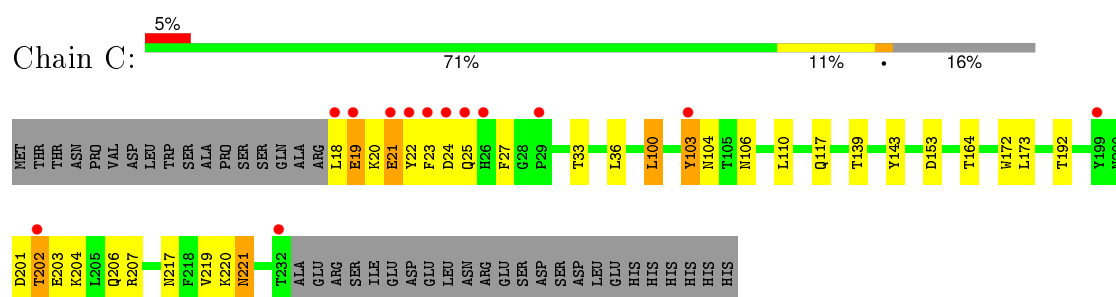
#### • Molecule 1: GP1



#### • Molecule 2: GP2



#### • Molecule 3: Niemann-Pick C1 protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 63	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	107.26 Å 107.26 Å 93.78 Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	41.86 – 2.30 41.86 – 2.30	Depositor EDS
% Data completeness (in resolution range)	100.0 (41.86-2.30) 100.0 (41.86-2.30)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.70 (at 2.29 Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.4_1496)	Depositor
R, $R_{free}$	0.181 , 0.230 0.180 , 0.230	Depositor DCC
$R_{free}$ test set	1367 reflections (5.01%)	DCC
Wilson B-factor (Å <sup>2</sup> )	41.5	Xtriage
Anisotropy	0.189	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 41.0	EDS
Estimated twinning fraction	0.046 for h,-h-k,-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 27310 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	3828	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	51.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.77% 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

### 5.1 Standard geometry

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	0.42	0/1223	0.59	1/1664 (0.1%)
2	B	0.42	0/702	0.62	0/952
3	C	0.39	0/1803	0.55	0/2469
All	All	0.40	0/3728	0.58	1/5085 (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(°)
1	A	42	VAL	CA-CB-CG2	5.22	118.73	110.90

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	69	ASN	Sidechain

### 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	1194	0	1153	18	1
2	B	687	0	679	13	1
3	C	1748	0	1656	24	1
4	B	28	0	25	1	0
5	A	77	0	0	1	1
5	B	31	0	0	1	1
5	C	63	0	0	2	1
All	All	3828	0	3513	53	3

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 7.

All (53) 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:B:510:LYS:NZ	4:B:701:NAG:O6	1.86	1.08
2:B:595:GLN:HE21	2:B:595:GLN:HA	1.25	0.98
1:A:39:HIS:HB2	1:A:42:VAL:HG22	1.60	0.83
2:B:595:GLN:NE2	2:B:595:GLN:HA	1.92	0.82
3:C:21:GLU:O	3:C:25:GLN:HG2	1.79	0.82
3:C:220:LYS:HG3	3:C:221:ASN:OD1	1.78	0.82
3:C:203:GLU:HG3	3:C:207:ARG:HH12	1.51	0.75
1:A:70:LEU:HD13	1:A:179:GLY:HA2	1.70	0.72
2:B:564:GLU:O	5:B:801:HOH:O	2.08	0.71
1:A:39:HIS:HD2	1:A:42:VAL:HG21	1.58	0.68
2:B:595:GLN:HE21	2:B:595:GLN:CA	2.05	0.67
2:B:509:PRO:HB2	2:B:510:LYS:HG3	1.77	0.67
3:C:201:ASP:OD2	3:C:204:LYS:HD3	1.97	0.65
3:C:117:GLN:NE2	5:C:301:HOH:O	2.20	0.58
1:A:39:HIS:CD2	1:A:42:VAL:HG21	2.38	0.57
1:A:31:ARG:N	5:A:203:HOH:O	2.38	0.55
1:A:112:GLU:OE2	1:A:172:ARG:HG3	2.07	0.55
1:A:69:ASN:HD21	1:A:105:ALA:HB2	1.72	0.54
3:C:106:ASN:HB3	3:C:164:THR:HG23	1.88	0.54
1:A:39:HIS:CD2	1:A:42:VAL:CG2	2.92	0.53
3:C:19:GLU:HG3	3:C:23:PHE:CE2	2.44	0.52
3:C:19:GLU:HG3	3:C:23:PHE:HE2	1.75	0.52
3:C:203:GLU:HG3	3:C:207:ARG:NH1	2.22	0.52
2:B:595:GLN:HG3	2:B:596:ARG:NH1	2.25	0.51

*Continued on next page...*



*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:69:ASN:ND2	1:A:105:ALA:HB2	2.25	0.51
1:A:163:ASP:OD1	2:B:543:TYR:OH	2.24	0.50
2:B:515:LEU:HB3	2:B:548:MET:HB2	1.93	0.50
2:B:595:GLN:CA	2:B:595:GLN:NE2	2.72	0.49
3:C:221:ASN:N	3:C:221:ASN:OD1	2.44	0.49
3:C:23:PHE:CD1	3:C:33:THR:HG21	2.48	0.48
3:C:23:PHE:HD1	3:C:27:PHE:HD2	1.60	0.48
1:A:38:ILE:HG13	1:A:187:PRO:HD3	1.95	0.48
3:C:100:LEU:HB3	3:C:104:ASN:HB3	1.95	0.48
1:A:104:TRP:CZ2	1:A:134:ARG:HD2	2.50	0.47
3:C:110:LEU:HB2	3:C:192:THR:HB	1.97	0.47
2:B:560:GLN:O	2:B:564:GLU:HG3	2.16	0.45
3:C:36:LEU:HD22	3:C:219:VAL:HG11	1.98	0.45
3:C:202:THR:O	3:C:206:GLN:HG2	2.16	0.45
1:A:37:VAL:HG21	1:A:51:LEU:HD11	1.97	0.45
3:C:20:LYS:O	3:C:24:ASP:OD2	2.33	0.45
1:A:79:VAL:O	1:A:83:THR:HG23	2.17	0.44
1:A:51:LEU:HA	1:A:51:LEU:HD12	1.68	0.44
1:A:70:LEU:HD11	1:A:180:VAL:HG22	2.00	0.44
3:C:22:TYR:CD1	3:C:173:LEU:HD11	2.52	0.44
3:C:143:TYR:CE2	3:C:153:ASP:HA	2.53	0.42
1:A:39:HIS:CB	1:A:42:VAL:HG22	2.40	0.42
3:C:100:LEU:O	3:C:103:TYR:HB3	2.20	0.42
3:C:22:TYR:HE2	3:C:27:PHE:CZ	2.37	0.42
1:A:63:LEU:HB3	2:B:585:LEU:HD22	2.01	0.41
3:C:23:PHE:HE1	3:C:192:THR:HG21	1.84	0.41
3:C:139:THR:OG1	5:C:302:HOH:O	2.21	0.41
2:B:513:PRO:O	2:B:555:ILE:HD13	2.21	0.41
3:C:18:LEU:HD13	3:C:172:TRP:CD1	2.57	0.40

All (3) 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)
5:A:251:HOH:O	5:B:815:HOH:O[3_655]	2.00	0.20
2:B:550:ASN:ND2	5:C:329:HOH:O[5_555]	2.07	0.13
1:A:137:TYR:OH	3:C:217:ASN:ND2[5_555]	2.19	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	156/158 (99%)	142 (91%)	12 (8%)	2 (1%)	15	15
2	B	83/130 (64%)	77 (93%)	4 (5%)	2 (2%)	7	5
3	C	213/256 (83%)	204 (96%)	8 (4%)	1 (0%)	34	41
All	All	452/544 (83%)	423 (94%)	24 (5%)	5 (1%)	17	18

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	103	TYR
1	A	50	LYS
1	A	51	LEU
2	B	597	TRP
2	B	550	ASN

### 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	126/130 (97%)	124 (98%)	2 (2%)	70	84
2	B	70/111 (63%)	69 (99%)	1 (1%)	74	86
3	C	194/232 (84%)	189 (97%)	5 (3%)	54	71
All	All	390/473 (82%)	382 (98%)	8 (2%)	61	78

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	53	CYS
1	A	169	VAL
2	B	595	GLN
3	C	19	GLU
3	C	21	GLU
3	C	100	LEU
3	C	202	THR
3	C	221	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	39	HIS
2	B	595	GLN
3	C	25	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](#)

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	B	701	2,4	14,14,15	0.68	0	15,19,21	0.69	0
4	NAG	B	702	4	14,14,15	0.29	0	15,19,21	0.47	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
4	NAG	B	701	2,4	-	0/6/23/26	0/1/1/1
4	NAG	B	702	4	-	0/6/23/26	0/1/1/1

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.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	B	701	NAG	1	0

## 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

### 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	158/158 (100%)	0.46	19 (12%) 6 9	25, 37, 123, 157	0
2	B	87/130 (66%)	0.48	8 (9%) 11 16	31, 48, 98, 143	0
3	C	215/256 (83%)	0.07	13 (6%) 25 33	31, 43, 111, 188	0
All	All	460/544 (84%)	0.28	40 (8%) 13 18	25, 41, 117, 188	0

All (40) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	525	ALA	12.5
3	C	21	GLU	11.2
3	C	18	LEU	9.7
3	C	25	GLN	7.2
1	A	53	CYS	6.4
3	C	24	ASP	6.3
2	B	526	ALA	6.1
2	B	527	ILE	6.1
3	C	26	HIS	5.3
3	C	22	TYR	5.1
1	A	51	LEU	4.8
1	A	54	ARG	4.7
3	C	103	TYR	4.6
2	B	509	PRO	4.5
1	A	49	ASP	4.2
1	A	39	HIS	4.0
3	C	232	THR	3.8
1	A	48	VAL	3.7
1	A	47	ASP	3.3
3	C	199	TYR	3.2
1	A	31	ARG	2.9
3	C	23	PHE	2.9
2	B	596	ARG	2.8

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	38	ILE	2.8
1	A	55	ASP	2.6
1	A	187	PRO	2.6
1	A	40	ASN	2.6
1	A	46	SER	2.6
3	C	202	THR	2.5
1	A	42	VAL	2.5
3	C	19	GLU	2.5
1	A	161	LEU	2.5
1	A	52	VAL	2.4
2	B	592	PHE	2.3
2	B	579	LEU	2.2
2	B	595	GLN	2.2
1	A	166	ALA	2.1
1	A	69	ASN	2.1
3	C	29	PRO	2.1
1	A	188	GLN	2.1

## 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(Å <sup>2</sup> )	Q<0.9
4	NAG	B	701	14/15	0.80	0.17	-0.22	61,92,104,111	0
4	NAG	B	702	14/15	0.65	0.39	-	124,133,140,140	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.