



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 26, 2017 – 11:55 AM EDT

PDB ID : 1CF5  
Title : BETA-MOMORCHARIN STRUCTURE AT 2.55 Å  
Authors : Yuan, Y.-R.; He, Y.-N.; Xiong, J.-P.; Xia, Z.-X.  
Deposited on : unknown  
Resolution : 2.55 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 : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20030345  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
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-20030345

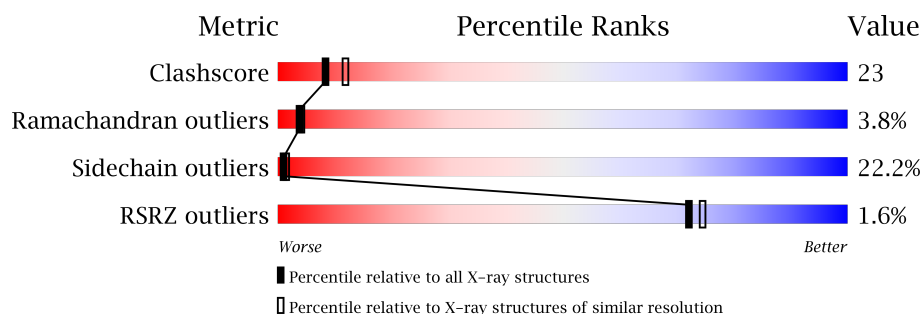
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.55 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))
Clashscore	112137	5755 (2.58-2.50)
Ramachandran outliers	110173	5652 (2.58-2.50)
Sidechain outliers	110143	5654 (2.58-2.50)
RSRZ outliers	101464	5026 (2.58-2.50)

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	249	<div> <div>2%</div> <div>46%</div> <div>40%</div> <div>12%</div> <div>.</div> </div>
1	B	249	<div> <div>2%</div> <div>43%</div> <div>42%</div> <div>14%</div> <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
2	FUC	A	901	X	-	-	-
2	FUC	B	901	X	-	-	-

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	NAG	B	902	-	-	-	X
6	XYP	A	906	-	-	-	X
6	XYP	B	906	-	-	-	X

## 2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 4188 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.

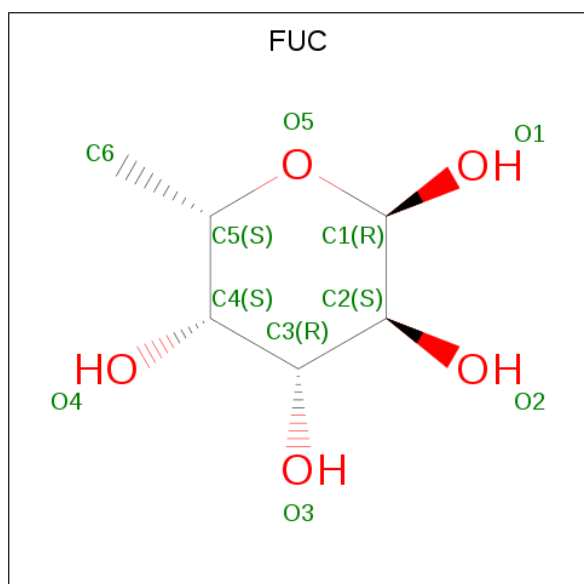
- Molecule 1 is a protein called PROTEIN (BETA-MOMORCHARIN).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
1	A	249	Total	C	N	O	0	0	0
			1991	1282	333	376			
1	B	249	Total	C	N	O	0	0	0
			1991	1282	333	376			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	51	ASN	ASP	CONFLICT	UNP P29339
A	219	GLN	GLU	CONFLICT	UNP P29339

- Molecule 2 is ALPHA-L-FUCOSE (three-letter code: FUC) (formula: C<sub>6</sub>H<sub>12</sub>O<sub>5</sub>).



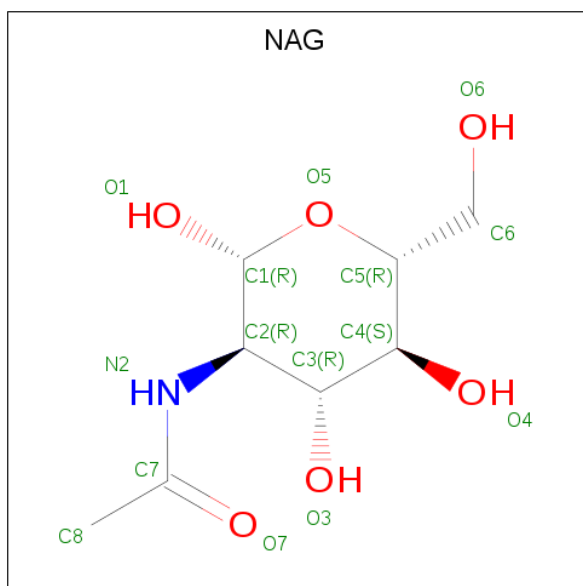
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			10	6	4		

*Continued on next page...*

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	B	1	Total	C	O	0	0
			10	6	4		

- Molecule 3 is N-ACETYL-D-GLUCOSAMINE (three-letter code: NAG) (formula:  $C_8H_{15}NO_6$ ).



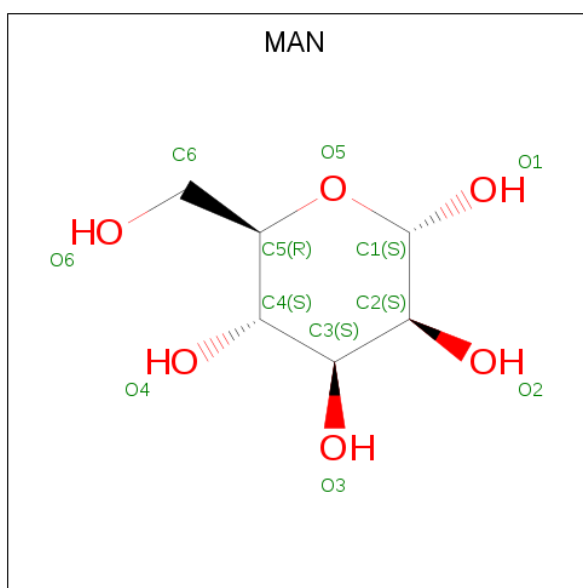
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	B	1	Total	C	N	O	0	0
			14	8	1	5		
3	B	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 4 is BETA-D-MANNOSE (three-letter code: BMA) (formula:  $C_6H_{12}O_6$ ).



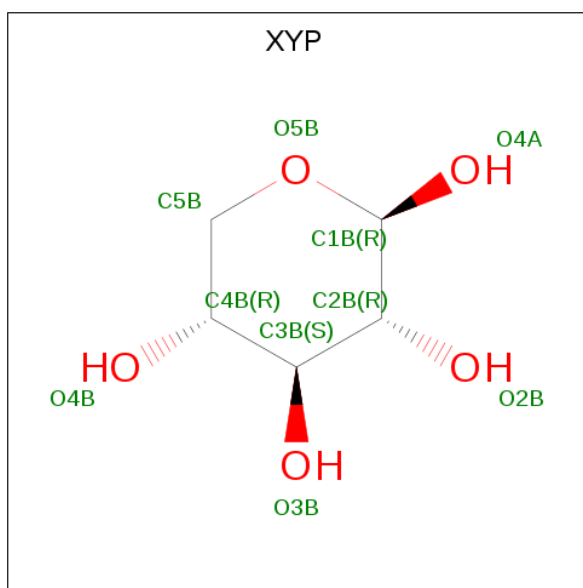
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			11	6	5		
4	B	1	Total	C	O	0	0
			11	6	5		

- Molecule 5 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
5	A	1	Total	C	O	0	0
			11	6	5		
5	B	1	Total	C	O	0	0
			11	6	5		

- Molecule 6 is BETA-D-XYLOPYRANOSE (three-letter code: XYP) (formula:  $C_5H_{10}O_5$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	A	1	Total	C	O	0	0
			9	5	4		
6	B	1	Total	C	O	0	0
			9	5	4		

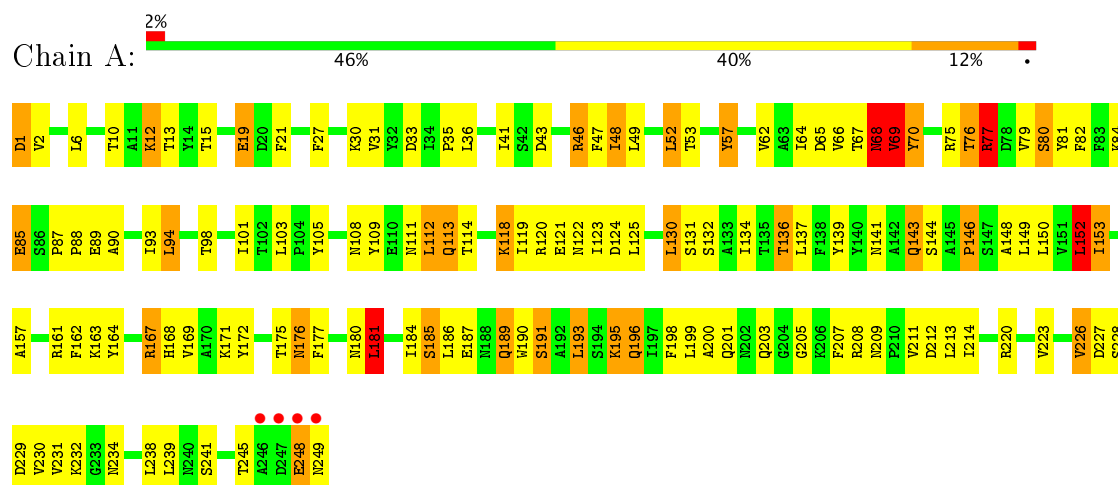
- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	31	Total	O	0	0
			31	31		
7	B	37	Total	O	0	0
			37	37		

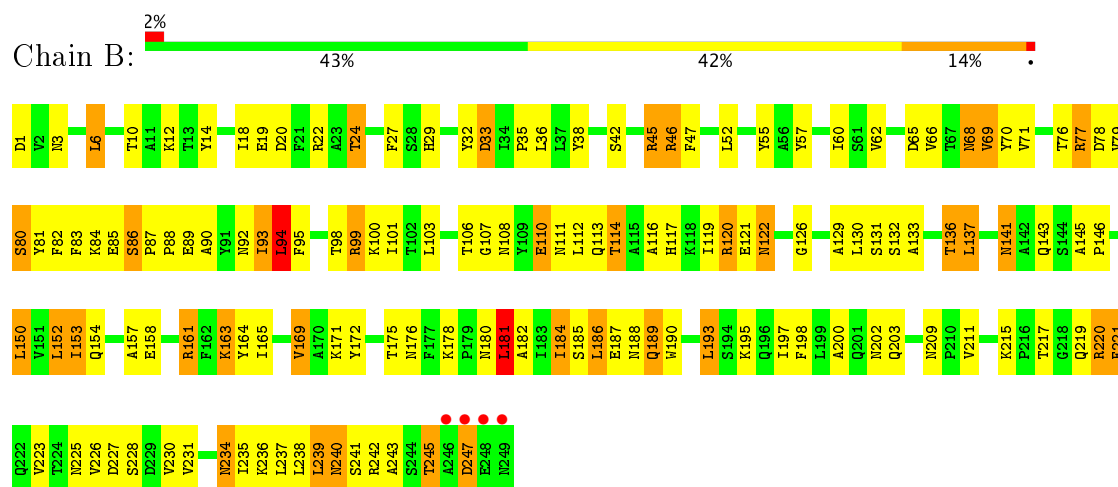
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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: PROTEIN (BETA-MOMORCHARIN)



#### • Molecule 1: PROTEIN (BETA-MOMORCHARIN)





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	49.09 Å   50.58 Å   61.12 Å 72.98°   78.39°   76.97°	Depositor
Resolution (Å)	10.00 – 2.55 32.81 – 2.43	Depositor EDS
% Data completeness (in resolution range)	79.3 (10.00-2.55) 75.0 (32.81-2.43)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.82 (at 2.42 Å)	Xtriage
Refinement program	X-PLOR 3.1	Depositor
R, $R_{free}$	0.172   ,   0.278 0.191   ,   (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	DCC
Wilson B-factor (Å <sup>2</sup> )	27.2	Xtriage
Anisotropy	0.669	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 94.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	4188	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.98% 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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: XYP, MAN, BMA, NAG, FUC

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	0.61	0/2034	0.77	2/2767 (0.1%)
1	B	0.61	0/2034	0.78	2/2767 (0.1%)
All	All	0.61	0/4068	0.77	4/5534 (0.1%)

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	7
1	B	0	10
All	All	0	17

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	B	152	LEU	CA-CB-CG	6.64	130.58	115.30
1	A	152	LEU	CA-CB-CG	6.40	130.01	115.30
1	A	181	LEU	CA-CB-CG	6.03	129.18	115.30
1	B	181	LEU	CA-CB-CG	5.93	128.93	115.30

There are no chirality outliers.

5 of 17 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	105	TYR	Sidechain
1	A	139	TYR	Sidechain

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Group
1	A	168	HIS	Sidechain
1	A	70	TYR	Sidechain
1	A	77	ARG	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	1991	0	2003	96	0
1	B	1991	0	2003	95	0
2	A	10	0	10	0	0
2	B	10	0	10	0	0
3	A	28	0	23	1	0
3	B	28	0	23	1	0
4	A	11	0	8	0	0
4	B	11	0	8	0	0
5	A	11	0	10	0	0
5	B	11	0	10	0	0
6	A	9	0	8	0	0
6	B	9	0	8	0	0
7	A	31	0	0	0	0
7	B	37	0	0	1	0
All	All	4188	0	4124	191	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 23.

The worst 5 of 191 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:B:226:VAL:HG23	1:B:231:VAL:HG11	1.25	1.17
1:B:189:GLN:HE21	1:B:189:GLN:HA	1.32	0.94
1:A:47:PHE:HD2	1:A:93:ILE:HD11	1.33	0.92
1:A:241:SER:O	1:A:245:THR:HG22	1.73	0.88
1:B:180:ASN:O	1:B:181:LEU:HB2	1.74	0.88

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	247/249 (99%)	213 (86%)	24 (10%)	10 (4%)	3	4
1	B	247/249 (99%)	219 (89%)	19 (8%)	9 (4%)	4	4
All	All	494/498 (99%)	432 (87%)	43 (9%)	19 (4%)	4	4

5 of 19 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	85	GLU
1	A	191	SER
1	B	77	ARG
1	B	228	SER
1	A	68	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	218/218 (100%)	172 (79%)	46 (21%)	1	2
1	B	218/218 (100%)	167 (77%)	51 (23%)	1	1
All	All	436/436 (100%)	339 (78%)	97 (22%)	1	1

5 of 97 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	232	LYS
1	B	62	VAL
1	B	221	PHE
1	A	248	GLU
1	B	24	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	29	HIS
1	B	117	HIS
1	B	234	ASN
1	B	68	ASN
1	B	113	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](#)

12 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
2	FUC	A	901	3	9,10,11	1.74	1 (11%)	13,14,16	1.29	2 (15%)
3	NAG	A	902	1,3,2	14,14,15	1.27	2 (14%)	15,19,21	2.06	5 (33%)
3	NAG	A	903	3,4	14,14,15	1.77	3 (21%)	15,19,21	3.12	8 (53%)
4	BMA	A	904	3,5,6	11,11,12	2.92	5 (45%)	13,15,17	1.80	4 (30%)
5	MAN	A	905	4	11,11,12	3.08	6 (54%)	13,15,17	1.57	2 (15%)
6	XYP	A	906	4	9,9,10	3.29	6 (66%)	10,12,14	1.88	3 (30%)
2	FUC	B	901	3	9,10,11	1.60	2 (22%)	13,14,16	1.24	1 (7%)
3	NAG	B	902	1,3,2	14,14,15	1.96	5 (35%)	15,19,21	3.35	9 (60%)
3	NAG	B	903	3,4	14,14,15	2.83	4 (28%)	15,19,21	1.72	4 (26%)
4	BMA	B	904	3,5,6	11,11,12	2.97	6 (54%)	13,15,17	3.15	6 (46%)
5	MAN	B	905	4	11,11,12	3.40	3 (27%)	13,15,17	2.43	5 (38%)
6	XYP	B	906	4	9,9,10	3.32	2 (22%)	10,12,14	1.61	2 (20%)

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	FUC	A	901	3	4/4/4/5	0/0/17/20	0/1/1/1
3	NAG	A	902	1,3,2	-	0/6/23/26	0/1/1/1
3	NAG	A	903	3,4	-	0/6/23/26	0/1/1/1
4	BMA	A	904	3,5,6	-	0/2/19/22	0/1/1/1
5	MAN	A	905	4	-	0/2/19/22	1/1/1/1
6	XYP	A	906	4	-	0/0/14/17	0/1/1/1
2	FUC	B	901	3	4/4/4/5	0/0/17/20	0/1/1/1
3	NAG	B	902	1,3,2	-	0/6/23/26	0/1/1/1
3	NAG	B	903	3,4	-	0/6/23/26	0/1/1/1
4	BMA	B	904	3,5,6	-	0/2/19/22	0/1/1/1
5	MAN	B	905	4	-	0/2/19/22	0/1/1/1
6	XYP	B	906	4	-	0/0/14/17	0/1/1/1

The worst 5 of 45 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	903	NAG	O5-C1	-4.19	1.36	1.43
2	B	901	FUC	O5-C1	2.04	1.47	1.43
6	A	906	XYP	O5B-C1B	2.04	1.47	1.42
6	A	906	XYP	O5B-C5B	2.06	1.47	1.42

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	905	MAN	C4-C5	2.11	1.57	1.53

The worst 5 of 51 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	B	904	BMA	C1-C2-C3	-7.43	100.24	109.65
3	A	903	NAG	O4-C4-C5	-3.72	99.91	109.28
3	A	903	NAG	O7-C7-C8	-3.56	115.58	122.06
4	A	904	BMA	C1-C2-C3	-3.25	105.53	109.65
3	B	903	NAG	O4-C4-C5	-3.17	101.30	109.28

5 of 8 chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	B	901	FUC	C2
2	B	901	FUC	C3
2	B	901	FUC	C1
2	B	901	FUC	C4
2	A	901	FUC	C2

There are no torsion outliers.

All (1) ring outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	A	905	MAN	C1-C2-C3-C4-C5-O5

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	903	NAG	1	0
3	B	902	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 [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	249/249 (100%)	-0.24	4 (1%) 72 74	3, 17, 49, 96	0
1	B	249/249 (100%)	-0.39	4 (1%) 72 74	2, 16, 51, 100	0
All	All	498/498 (100%)	-0.32	8 (1%) 72 74	2, 16, 50, 100	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	249	ASN	9.6
1	A	247	ASP	8.0
1	A	248	GLU	7.2
1	B	248	GLU	6.2
1	B	249	ASN	4.4

### 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
6	XYP	B	906	9/10	0.57	0.43	8.52	89,94,96,98	0
3	NAG	B	902	14/15	0.80	0.22	3.29	54,62,72,79	0
3	NAG	A	902	14/15	0.84	0.16	1.71	57,70,78,79	0
6	XYP	A	906	9/10	0.62	0.49	0.40	103,104,106,107	0
5	MAN	A	905	11/12	0.52	0.41	-	97,101,103,103	0
2	FUC	B	901	10/11	0.79	0.19	-	86,89,93,94	0
4	BMA	A	904	11/12	0.70	0.29	-	94,99,102,103	0
4	BMA	B	904	11/12	0.68	0.26	-	95,98,101,102	0
2	FUC	A	901	10/11	0.69	0.39	-	85,92,96,98	0
5	MAN	B	905	11/12	0.76	0.33	-	93,98,100,101	0
3	NAG	A	903	14/15	0.74	0.27	-	86,93,95,96	0
3	NAG	B	903	14/15	0.76	0.24	-	75,79,87,90	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.