



# Full wwPDB X-ray Structure Validation Report ⓘ

Mar 8, 2018 – 06:45 pm GMT

PDB ID : 4AGT  
Title : Crystal Structure of Fucose binding lectin from *Aspergillus Fumigatus* (AFL) in complex with Fuc1-6GlcNAc.  
Authors : Houser, J.; Komarek, J.; Kostlanova, N.; Lahmann, M.; Cioci, G.; Varrot, A.; Imberty, A.; Wimmerova, M.  
Deposited on : 2012-01-31  
Resolution : 2.00 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

<https://www.wwpdb.org/validation/2017/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 : 1.7.3 (157068), CSD as539be (2018)  
Xtriage (Phenix) : 1.13  
EDS : trunk30967  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Refmac : 5.8.0158  
CCP4 : 7.0 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk30967

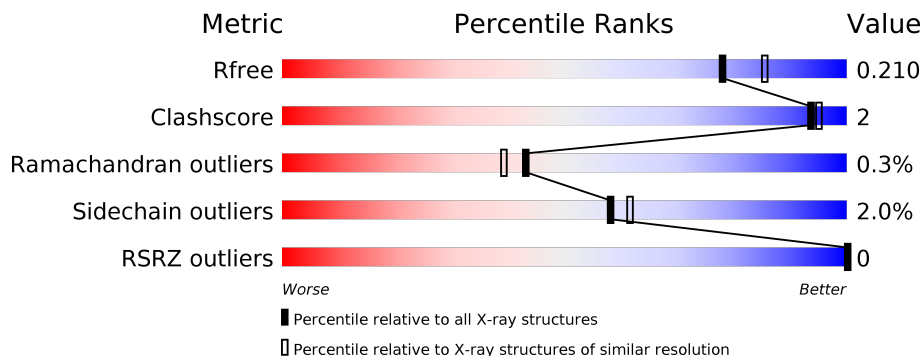
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.00 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))
$R_{free}$	111664	7193 (2.00-2.00)
Clashscore	122126	8267 (2.00-2.00)
Ramachandran outliers	120053	8166 (2.00-2.00)
Sidechain outliers	120020	8165 (2.00-2.00)
RSRZ outliers	108989	7011 (2.00-2.00)

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	315	94% 5%
1	B	315	92% 7%

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 5482 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 FUCOSE-SPECIFIC LECTIN FLEA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	314	2443	1557	422	460	4	0	0	0
1	B	314	2440	1556	422	458	4	0	0	0

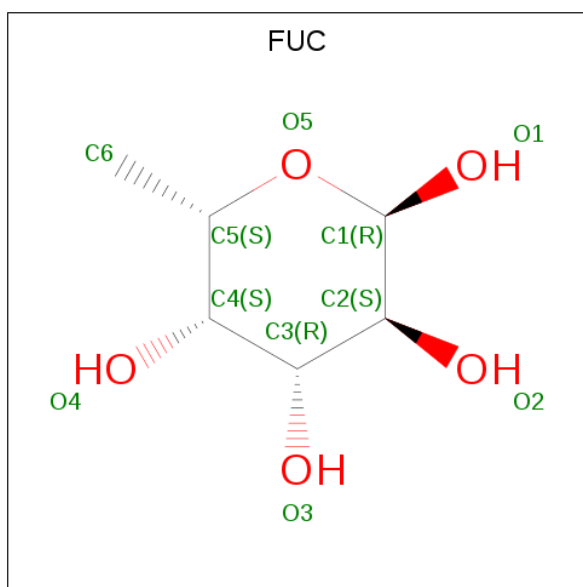
There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	20	SER	LEU	conflict	UNP Q4WW81
A	111	CYS	ARG	conflict	UNP Q4WW81
B	20	SER	LEU	conflict	UNP Q4WW81
B	111	CYS	ARG	conflict	UNP Q4WW81

- Molecule 2 is SODIUM ION (three-letter code: NA) (formula: Na).

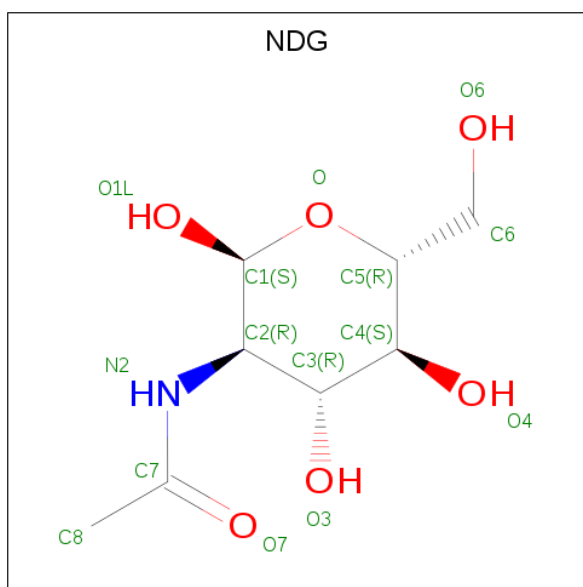
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Na	0	0
			1	1		

- Molecule 3 is ALPHA-L-FUCOSE (three-letter code: FUC) (formula: C<sub>6</sub>H<sub>12</sub>O<sub>5</sub>).



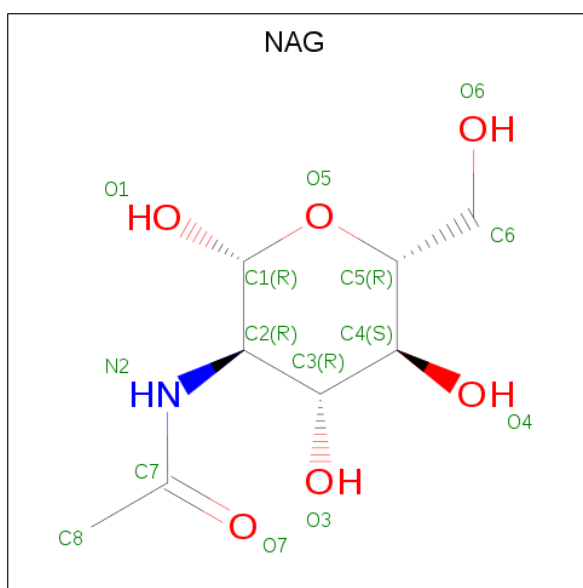
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 11 6 5	0	0
3	A	1	Total C O 10 6 4	0	0
3	A	1	Total C O 11 6 5	0	0
3	A	1	Total C O 11 6 5	0	0
3	A	1	Total C O 11 6 5	0	0
3	A	1	Total C O 11 6 5	0	0
3	B	1	Total C O 11 6 5	0	0
3	B	1	Total C O 11 6 5	0	0
3	B	1	Total C O 11 6 5	0	0
3	B	1	Total C O 11 6 5	0	0
3	B	1	Total C O 10 6 4	0	0
3	B	1	Total C O 11 6 5	0	0

- Molecule 4 is 2-(ACETYLAMINO)-2-DEOXY-A-D-GLUCOPYRANOSE (three-letter code: NDG) (formula: C<sub>8</sub>H<sub>15</sub>NO<sub>6</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
4	A	1	15	8	1	6	0	0
4	B	1	15	8	1	6	0	1

- Molecule 5 is N-ACETYL-D-GLUCOSAMINE (three-letter code: NAG) (formula:  $C_8H_{15}NO_6$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
5	B	1	15	8	1	6	0	1

- Molecule 6 is water.

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
6	A	236	Total 236	O 236	0	0
6	B	187	Total 187	O 187	0	0

### 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: FUCOSE-SPECIFIC LECTIN FLEA

Chain A:  94% 5%



- Molecule 1: FUCOSE-SPECIFIC LECTIN FLEA

Chain B:  92% 7%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	47.41Å 88.35Å 79.78Å 90.00° 102.99° 90.00°	Depositor
Resolution (Å)	77.74 – 2.00 38.41 – 2.00	Depositor EDS
% Data completeness (in resolution range)	88.1 (77.74-2.00) 88.2 (38.41-2.00)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.40 (at 2.00Å)	Xtrriage
Refinement program	REFMAC 5.6.0117	Depositor
R, $R_{free}$	0.172 , 0.212 0.175 , 0.210	Depositor DCC
$R_{free}$ test set	1911 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	15.9	Xtrriage
Anisotropy	0.457	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 42.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	5482	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	16.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.85% 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: NA, NAG, NDG, 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	0.73	3/2514 (0.1%)	0.75	2/3427 (0.1%)
1	B	0.74	3/2511 (0.1%)	0.75	2/3423 (0.1%)
All	All	0.73	6/5025 (0.1%)	0.75	4/6850 (0.1%)

All (6) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	241	TRP	CD2-CE2	5.46	1.48	1.41
1	A	245	TRP	CD2-CE2	5.42	1.47	1.41
1	A	271	TRP	CD2-CE2	5.35	1.47	1.41
1	A	299	TRP	CD2-CE2	5.28	1.47	1.41
1	B	179	TRP	CD2-CE2	5.14	1.47	1.41
1	B	141	TRP	CD2-CE2	5.07	1.47	1.41

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	39	LEU	CA-CB-CG	6.68	130.66	115.30
1	A	30	ASP	CB-CG-OD1	5.25	123.02	118.30
1	A	39	LEU	CA-CB-CG	5.15	127.14	115.30
1	B	12	ARG	NE-CZ-NH2	-5.14	117.73	120.30

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	2443	0	2329	6	0
1	B	2440	0	2327	9	0
2	A	1	0	0	0	0
3	A	65	0	70	0	0
3	B	65	0	70	0	0
4	A	15	0	14	0	0
4	B	15	0	14	0	0
5	B	15	0	14	0	0
6	A	236	0	0	2	0
6	B	187	0	0	1	0
All	All	5482	0	4838	15	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 2.

All (15) 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:105:GLN:HG3	6:B:2067:HOH:O	1.99	0.62
1:B:49:GLU:CD	1:B:49:GLU:H	2.10	0.55
1:A:105:GLN:HG3	6:A:2108:HOH:O	2.08	0.52
1:B:34:SER:OG	1:B:55:ASN:ND2	2.42	0.52
1:B:80:GLY:O	1:B:105:GLN:NE2	2.43	0.51
1:B:34:SER:HA	1:B:55:ASN:HD22	1.77	0.50
1:A:254:LYS:HE3	6:A:2195:HOH:O	2.14	0.48
1:A:220:PHE:CZ	1:A:279:PRO:HG3	2.49	0.48
1:B:114:ALA:CB	1:B:127:ILE:HD12	2.45	0.47
1:B:119:GLY:HA3	1:B:167:ARG:HD3	2.00	0.43
1:A:282:ARG:CD	1:A:298:VAL:HG22	2.49	0.43
1:B:216:ALA:HB1	1:B:268:ILE:HG22	2.01	0.43
1:B:242:GLN:O	1:B:254:LYS:HA	2.20	0.42
1:A:39:LEU:HD12	1:A:39:LEU:C	2.39	0.41
1:A:34:SER:OG	1:A:55:ASN:ND2	2.54	0.40

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	312/315 (99%)	304 (97%)	7 (2%)	1 (0%)	43	39
1	B	312/315 (99%)	304 (97%)	7 (2%)	1 (0%)	43	39
All	All	624/630 (99%)	608 (97%)	14 (2%)	2 (0%)	43	39

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	46	ASN
1	A	46	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	249/250 (100%)	244 (98%)	5 (2%)	58	61
1	B	248/250 (99%)	243 (98%)	5 (2%)	58	61
All	All	497/500 (99%)	487 (98%)	10 (2%)	58	61

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	39	LEU
1	A	147	LYS
1	A	195	HIS
1	A	254	LYS

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	280	ASP
1	B	39	LEU
1	B	147	LYS
1	B	258	THR
1	B	280	ASP
1	B	306	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (6) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	46	ASN
1	A	55	ASN
1	B	46	ASN
1	B	55	ASN
1	B	105	GLN
1	B	189	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](#)

Of 16 ligands modelled in this entry, 1 is monoatomic - leaving 15 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
3	FUC	A	910	-	11,11,11	0.74	0	15,16,16	0.85	0
3	FUC	A	920	4	9,10,11	0.59	0	13,14,16	0.77	0
4	NDG	A	921	3	15,15,15	0.63	0	21,21,21	0.91	0
3	FUC	A	930	-	11,11,11	0.95	0	15,16,16	1.13	1 (6%)
3	FUC	A	940	-	11,11,11	0.83	0	15,16,16	0.96	0
3	FUC	A	950	-	11,11,11	0.77	0	15,16,16	0.70	0
3	FUC	A	960	-	11,11,11	0.67	0	15,16,16	0.84	1 (6%)
3	FUC	B	910	-	11,11,11	0.69	0	15,16,16	0.60	0
3	FUC	B	920	-	11,11,11	0.75	0	15,16,16	1.00	0
3	FUC	B	930	-	11,11,11	0.85	0	15,16,16	1.27	1 (6%)
3	FUC	B	940	-	11,11,11	0.63	0	15,16,16	1.09	2 (13%)
3	FUC	B	950	5,4	9,10,11	0.65	0	13,14,16	1.12	0
5	NAG	B	951[A]	3	15,15,15	0.67	0	21,21,21	1.38	1 (4%)
4	NDG	B	952[B]	3	15,15,15	0.74	0	21,21,21	1.37	1 (4%)
3	FUC	B	960	-	11,11,11	0.82	0	15,16,16	1.04	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
3	FUC	A	910	-	-	0/0/20/20	0/1/1/1
3	FUC	A	920	4	-	0/0/17/20	0/1/1/1
4	NDG	A	921	3	-	0/6/26/26	0/1/1/1
3	FUC	A	930	-	-	0/0/20/20	0/1/1/1
3	FUC	A	940	-	-	0/0/20/20	0/1/1/1
3	FUC	A	950	-	-	0/0/20/20	0/1/1/1
3	FUC	A	960	-	-	0/0/20/20	0/1/1/1
3	FUC	B	910	-	-	0/0/20/20	0/1/1/1
3	FUC	B	920	-	-	0/0/20/20	0/1/1/1
3	FUC	B	930	-	-	0/0/20/20	0/1/1/1
3	FUC	B	940	-	-	0/0/20/20	0/1/1/1
3	FUC	B	950	5,4	-	0/0/17/20	0/1/1/1
5	NAG	B	951[A]	3	-	0/6/26/26	0/1/1/1
4	NDG	B	952[B]	3	-	0/6/26/26	0/1/1/1
3	FUC	B	960	-	-	0/0/20/20	0/1/1/1

There are no bond length outliers.

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	B	951[A]	NAG	O5-C1-C2	-4.73	104.76	109.52
4	B	952[B]	NDG	O-C1-C2	-4.73	104.76	109.52
3	B	940	FUC	C6-C5-C4	-2.54	108.69	113.06
3	B	940	FUC	O5-C5-C4	2.09	113.01	109.62
3	B	960	FUC	C1-C2-C3	2.21	114.78	110.36
3	A	930	FUC	O2-C2-C1	2.32	114.58	109.14
3	A	960	FUC	C1-C2-C3	2.42	115.21	110.36
3	B	930	FUC	C1-C2-C3	3.06	116.48	110.36

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 [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	314/315 (99%)	-0.56	0 100 100	9, 14, 24, 32	0
1	B	314/315 (99%)	-0.55	0 100 100	8, 14, 25, 31	0
All	All	628/630 (99%)	-0.56	0 100 100	8, 14, 24, 32	0

There are no RSRZ outliers to report.

### 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. 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	B-factors(Å <sup>2</sup> )	Q<0.9
3	FUC	B	910	11/11	0.85	0.18	22,28,33,35	0
3	FUC	A	910	11/11	0.87	0.19	20,24,27,30	0
3	FUC	A	930	11/11	0.87	0.14	20,22,26,27	0
3	FUC	B	960	11/11	0.88	0.18	21,25,28,29	0
3	FUC	B	930	11/11	0.90	0.18	16,19,23,24	0
3	FUC	A	940	11/11	0.91	0.14	22,23,27,29	0
3	FUC	B	940	11/11	0.91	0.13	16,20,22,24	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
4	NDG	A	921	15/15	0.92	0.20	22,29,34,35	0
4	NDG	B	952[B]	15/15	0.92	0.12	20,25,28,28	15
3	FUC	A	960	11/11	0.92	0.14	16,18,24,25	0
5	NAG	B	951[A]	15/15	0.93	0.13	20,25,28,28	15
3	FUC	A	950	11/11	0.93	0.11	18,20,22,22	0
3	FUC	B	920	11/11	0.94	0.09	14,16,17,18	0
3	FUC	B	950	10/11	0.96	0.07	13,14,15,16	0
3	FUC	A	920	10/11	0.97	0.07	14,15,15,16	0
2	NA	A	700	1/1	0.99	0.16	11,11,11,11	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.