



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 19, 2016 – 10:59 PM GMT

PDB ID : 5F9A
Title : Blood group antigen binding adhesin BabA of Helicobacter pylori strain P436
in complex with blood group H Lewis b hexasaccharide
Authors : Moonens, K.; Gideonsson, P.; Subedi, S.; Romao, E.; Oscarson, S.; Muylder-
mans, S.; Boren, T.; Remaut, H.
Deposited on : 2015-12-09
Resolution : 2.44 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
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

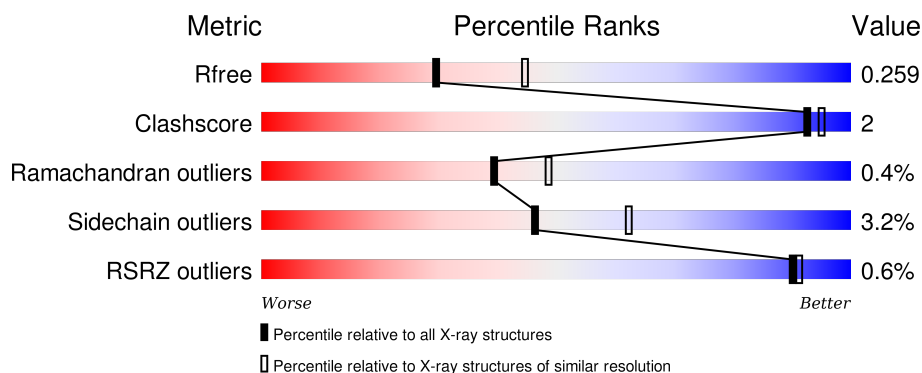
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION



The reported resolution of this entry is 2.44 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	1003 (2.46-2.42)
Clashscore	102246	1071 (2.46-2.42)
Ramachandran outliers	100387	1065 (2.46-2.42)
Sidechain outliers	100360	1065 (2.46-2.42)
RSRZ outliers	91569	1005 (2.46-2.42)

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 ≥ 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	463	 83% 6% • 10%
2	C	120	 87% 8% 6%

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 4084 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 Adhesin binding fucosylated histo-blood group antigen,Adhesin,Adhesin binding fucosylated histo-blood group antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	418	Total	C	N	O	S	0	1	0
			3108	1921	538	636	13			

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	3	ALA	-	expression tag	UNP O52269
A	4	SER	-	expression tag	UNP O52269
A	5	TRP	-	expression tag	UNP O52269
A	6	SER	-	expression tag	UNP O52269
A	7	HIS	-	expression tag	UNP O52269
A	8	PRO	-	expression tag	UNP O52269
A	9	GLN	-	expression tag	UNP O52269
A	10	PHE	-	expression tag	UNP O52269
A	11	GLU	-	expression tag	UNP O52269
A	12	LYS	-	expression tag	UNP O52269
A	13	SER	-	expression tag	UNP O52269
A	14	GLY	-	expression tag	UNP O52269
A	15	GLY	-	expression tag	UNP O52269
A	16	GLY	-	expression tag	UNP O52269
A	17	GLY	-	expression tag	UNP O52269
A	18	GLY	-	expression tag	UNP O52269
A	19	LEU	-	expression tag	UNP O52269
A	20	VAL	-	expression tag	UNP O52269
A	21	PRO	-	expression tag	UNP O52269
A	22	ARG	-	expression tag	UNP O52269
A	23	GLY	-	expression tag	UNP O52269
A	24	SER	-	expression tag	UNP O52269
A	458	GLY	-	expression tag	UNP O52269
A	459	SER	-	expression tag	UNP O52269
A	460	HIS	-	expression tag	UNP O52269
A	461	HIS	-	expression tag	UNP O52269

Continued on next page...

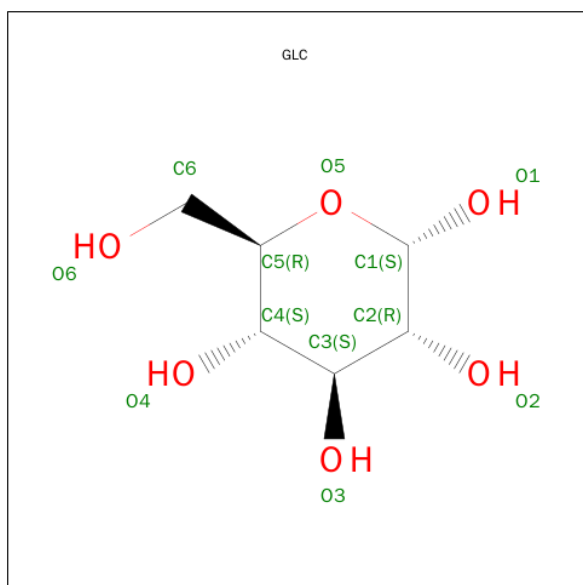
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	462	HIS	-	expression tag	UNP O52269
A	463	HIS	-	expression tag	UNP O52269
A	464	HIS	-	expression tag	UNP O52269
A	465	HIS	-	expression tag	UNP O52269

- Molecule 2 is a protein called Nanobody Nb-ER19.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	113	Total	C	N	O	S	0	0	0
			863	540	155	163	5			

- Molecule 3 is ALPHA-D-GLUCOSE (three-letter code: GLC) (formula: C₆H₁₂O₆).



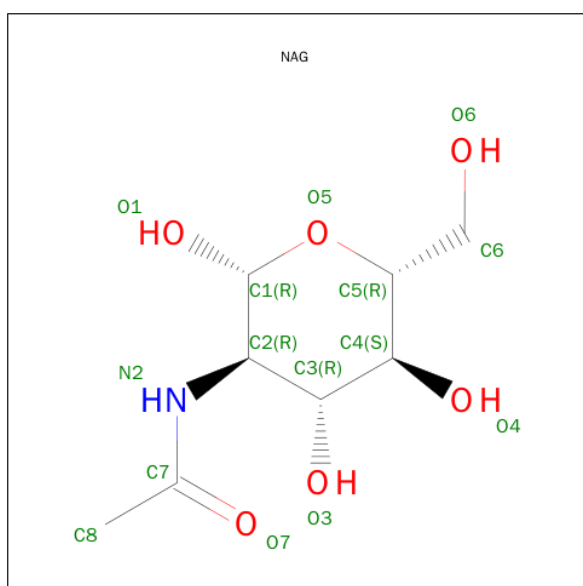
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			12	6	6		

- Molecule 4 is BETA-D-GALACTOSE (three-letter code: GAL) (formula: C₆H₁₂O₆).



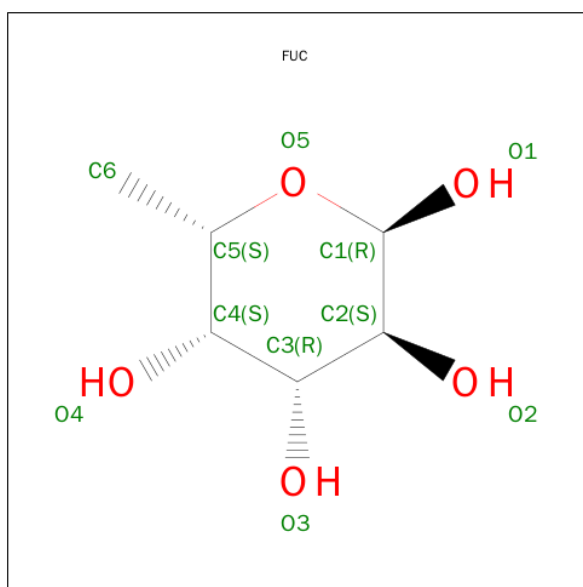
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			11	6	5		
4	A	1	Total	C	O	0	0
			11	6	5		

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



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

- Molecule 6 is ALPHA-L-FUCOSE (three-letter code: FUC) (formula: $C_6H_{12}O_5$).

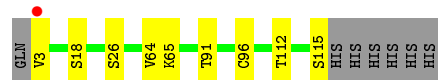


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	A	1	Total	C	O	0	0
			10	6	4		
6	A	1	Total	C	O	0	0
			10	6	4		

- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	32	Total	O	0	0
			32	32		
7	C	13	Total	O	0	0
			13	13		

- Molecule 1: Adhesin binding fucosylated histo-blood group antigen, Adhesin, Adhesin binding fucosylated histo-blood group antigen



4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α , β , γ	124.05Å 136.17Å 124.63Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	51.54 – 2.44 51.54 – 2.44	Depositor EDS
% Data completeness (in resolution range)	99.3 (51.54-2.44) 99.3 (51.54-2.44)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.78 (at 2.45Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, R_{free}	0.205 , 0.257 0.208 , 0.259	Depositor DCC
R_{free} test set	1996 reflections (5.35%)	DCC
Wilson B-factor (Å ²)	51.7	Xtriage
Anisotropy	0.058	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 21.7	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 39274 reflections	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	4084	wwPDB-VP
Average B, all atoms (Å ²)	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.34% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: GLC, GAL, 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	0.83	0/3163	0.87	5/4304 (0.1%)
2	C	0.84	1/882 (0.1%)	0.89	0/1195
All	All	0.84	1/4045 (0.0%)	0.88	5/5499 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	18	SER	CB-OG	-5.84	1.34	1.42

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	76	ASP	CB-CG-OD1	6.04	123.73	118.30
1	A	139	ARG	NE-CZ-NH1	5.74	123.17	120.30
1	A	39	LEU	CA-CB-CG	-5.47	102.72	115.30
1	A	315	LEU	CA-CB-CG	5.34	127.59	115.30
1	A	76	ASP	CB-CG-OD2	-5.32	113.51	118.30

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	3108	0	3016	12	0
2	C	863	0	845	4	0
3	A	12	0	11	0	0
4	A	22	0	18	0	0
5	A	14	0	11	0	0
6	A	20	0	20	0	0
7	A	32	0	0	0	0
7	C	13	0	0	0	0
All	All	4084	0	3921	16	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.

The worst 5 of 16 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:C:3:VAL:N	2:C:26:SER:HG	1.52	1.07
2:C:91:THR:HG23	2:C:112:THR:HA	1.83	0.61
2:C:3:VAL:N	2:C:26:SER:OG	2.30	0.58
1:A:305:ILE:HA	1:A:308:MET:HE3	1.85	0.57
1:A:394:ASN:O	1:A:395:LYS:HB2	2.06	0.54

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	415/463 (90%)	399 (96%)	14 (3%)	2 (0%)	34	41
2	C	111/120 (92%)	108 (97%)	3 (3%)	0	100	100
All	All	526/583 (90%)	507 (96%)	17 (3%)	2 (0%)	39	49

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	459	SER
1	A	205	GLY

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	341/374 (91%)	329 (96%)	12 (4%)	43	60
2	C	92/99 (93%)	90 (98%)	2 (2%)	60	74
All	All	433/473 (92%)	419 (97%)	14 (3%)	46	62

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	209	ARG
1	A	230	SER
1	A	378	THR
1	A	207	ASP
1	A	317	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
2	C	77	ASN

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

6 non-standard protein/DNA/RNA residues 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
3	GLC	A	1001	4	12,12,12	0.78	0	17,17,17	2.54	8 (47%)
4	GAL	A	1002	3,5	11,11,12	0.76	0	15,15,17	1.49	3 (20%)
5	NAG	A	1003	4,6	14,14,15	0.91	0	15,19,21	2.74	9 (60%)
4	GAL	A	1004	5,6	11,11,12	1.30	1 (9%)	15,15,17	1.11	1 (6%)
6	FUC	A	1005	4	10,10,11	1.00	1 (10%)	13,14,16	1.33	1 (7%)
6	FUC	A	1006	5	10,10,11	0.87	0	13,14,16	2.45	8 (61%)

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	GLC	A	1001	4	-	0/2/22/22	0/1/1/1
4	GAL	A	1002	3,5	-	0/2/19/22	0/1/1/1
5	NAG	A	1003	4,6	-	0/6/23/26	0/1/1/1
4	GAL	A	1004	5,6	-	0/2/19/22	0/1/1/1
6	FUC	A	1005	4	-	0/0/17/20	0/1/1/1
6	FUC	A	1006	5	-	0/0/17/20	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	1004	GAL	O5-C1	-3.24	1.38	1.43
6	A	1005	FUC	O5-C1	-2.03	1.40	1.43

The worst 5 of 30 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	1003	NAG	O3-C3-C2	-5.58	97.43	109.37
5	A	1003	NAG	O4-C4-C5	-4.73	96.77	109.23
6	A	1006	FUC	C1-C2-C3	-4.03	104.67	109.55
6	A	1005	FUC	O5-C1-C2	-3.66	105.05	110.89
6	A	1006	FUC	C2-C3-C4	-2.73	106.28	111.05

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

6 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$
3	GLC	A	1001	4	12,12,12	0.78	0	17,17,17	2.54	8 (47%)
4	GAL	A	1002	3,5	11,11,12	0.76	0	15,15,17	1.49	3 (20%)
5	NAG	A	1003	4,6	14,14,15	0.91	0	15,19,21	2.74	9 (60%)
4	GAL	A	1004	5,6	11,11,12	1.30	1 (9%)	15,15,17	1.11	1 (6%)
6	FUC	A	1005	4	10,10,11	1.00	1 (10%)	13,14,16	1.33	1 (7%)
6	FUC	A	1006	5	10,10,11	0.87	0	13,14,16	2.45	8 (61%)

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	GLC	A	1001	4	-	0/2/22/22	0/1/1/1
4	GAL	A	1002	3,5	-	0/2/19/22	0/1/1/1
5	NAG	A	1003	4,6	-	0/6/23/26	0/1/1/1
4	GAL	A	1004	5,6	-	0/2/19/22	0/1/1/1
6	FUC	A	1005	4	-	0/0/17/20	0/1/1/1

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
6	FUC	A	1006	5	-	0/0/17/20	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	1004	GAL	O5-C1	-3.24	1.38	1.43
6	A	1005	FUC	O5-C1	-2.03	1.40	1.43

The worst 5 of 30 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	1003	NAG	O3-C3-C2	-5.58	97.43	109.37
5	A	1003	NAG	O4-C4-C5	-4.73	96.77	109.23
6	A	1006	FUC	C1-C2-C3	-4.03	104.67	109.55
6	A	1005	FUC	O5-C1-C2	-3.66	105.05	110.89
6	A	1006	FUC	C2-C3-C4	-2.73	106.28	111.05

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, 95th 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 ²)	Q<0.9
1	A	418/463 (90%)	-0.41	2 (0%) 91 92	38, 52, 90, 120	0
2	C	113/120 (94%)	-0.49	1 (0%) 85 87	43, 56, 72, 117	0
All	All	531/583 (91%)	-0.43	3 (0%) 90 91	38, 53, 88, 120	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	3	VAL	6.4
1	A	405	GLY	2.5
1	A	335	ASN	2.5

6.2 Non-standard residues in protein, DNA, RNA chains [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, 95th 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 ²)	Q<0.9
6	FUC	A	1005	10/11	0.98	0.10	-0.46	40,48,51,53	0
4	GAL	A	1004	11/12	0.97	0.12	-	61,64,75,85	0
6	FUC	A	1006	10/11	0.96	0.15	-	69,80,85,92	0
5	NAG	A	1003	14/15	0.97	0.09	-	60,64,72,73	0
3	GLC	A	1001	12/12	0.92	0.24	-	81,91,97,100	0
4	GAL	A	1002	11/12	0.97	0.17	-	68,73,96,104	0

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, 95th 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(\AA^2)	Q<0.9
6	FUC	A	1005	10/11	0.98	0.10	-0.46	40,48,51,53	0
4	GAL	A	1004	11/12	0.97	0.12	-	61,64,75,85	0
6	FUC	A	1006	10/11	0.96	0.15	-	69,80,85,92	0
5	NAG	A	1003	14/15	0.97	0.09	-	60,64,72,73	0
3	GLC	A	1001	12/12	0.92	0.24	-	81,91,97,100	0
4	GAL	A	1002	11/12	0.97	0.17	-	68,73,96,104	0

6.5 Other polymers [i](#)

There are no such residues in this entry.