



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 12:50 am GMT

PDB ID : 4JUJ  
Title : Crystal structure of 1918 pandemic influenza virus hemagglutinin mutnat D225G complexed with human receptor analogue LSTc  
Authors : Zhang, W.; Shi, Y.; Qi, J.; Gao, F.; Li, Q.; Fan, Z.; Yan, J.; Gao, G.F.  
Deposited on : 2013-03-25  
Resolution : 3.01 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 : trunk28620  
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) : recalc28949

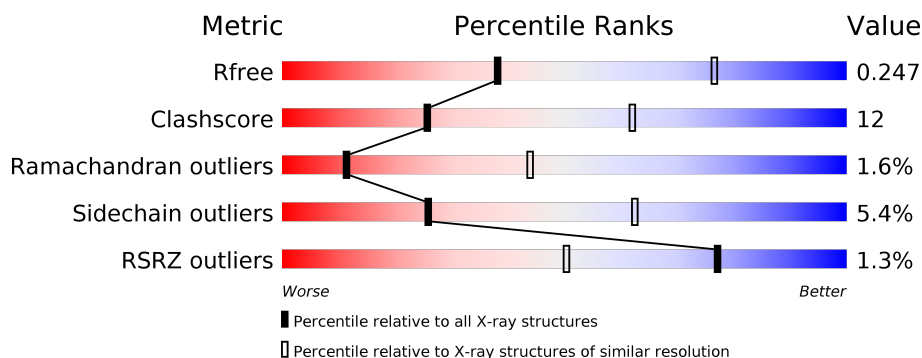
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.01 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))
$R_{free}$	100719	1924 (3.04-3.00)
Clashscore	112137	2279 (3.04-3.00)
Ramachandran outliers	110173	2207 (3.04-3.00)
Sidechain outliers	110143	2210 (3.04-3.00)
RSRZ outliers	101464	1948 (3.04-3.00)

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	324	<div> <div>76%</div> <div>23%</div> <div>.</div> </div>
1	C	324	<div> <div>%</div> <div>77%</div> <div>22%</div> <div>.</div> </div>
1	E	324	<div> <div>%</div> <div>76%</div> <div>23%</div> <div>.</div> </div>
2	B	170	<div> <div>%</div> <div>72%</div> <div>24%</div> <div>.</div> </div>
2	D	170	<div> <div>3%</div> <div>71%</div> <div>20%</div> <div>5%</div> <div>.</div> </div>
2	F	170	<div> <div>4%</div> <div>70%</div> <div>22%</div> <div>.</div> <div>.</div> <div>.</div> </div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 11784 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 Hemagglutinin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	324	Total	C	N	O	S	0	0	0
			2512	1584	434	483	11			
1	C	324	Total	C	N	O	S	0	0	0
			2512	1584	434	483	11			
1	E	324	Total	C	N	O	S	0	0	0
			2512	1584	434	483	11			

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	225	GLY	ASP	ENGINEERED MUTATION	UNP Q9WFX3
A	326	ALA	-	EXPRESSION TAG	UNP Q9WFX3
A	327	ARG	-	EXPRESSION TAG	UNP Q9WFX3
C	225	GLY	ASP	ENGINEERED MUTATION	UNP Q9WFX3
C	326	ALA	-	EXPRESSION TAG	UNP Q9WFX3
C	327	ARG	-	EXPRESSION TAG	UNP Q9WFX3
E	225	GLY	ASP	ENGINEERED MUTATION	UNP Q9WFX3
E	326	ALA	-	EXPRESSION TAG	UNP Q9WFX3
E	327	ARG	-	EXPRESSION TAG	UNP Q9WFX3

- Molecule 2 is a protein called Hemagglutinin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	170	Total	C	N	O	S	0	0	0
			1368	855	236	271	6			
2	D	164	Total	C	N	O	S	0	0	0
			1317	825	225	261	6			
2	F	165	Total	C	N	O	S	0	0	0
			1326	830	226	264	6			

- Molecule 3 is a polymer of unknown type called SUGAR (2-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	2	Total	C	N	O	0	0
			28	16	2	10		
3	C	2	Total	C	N	O	0	0
			28	16	2	10		
3	E	2	Total	C	N	O	0	0
			28	16	2	10		

- Molecule 4 is a polymer of unknown type called SUGAR (3-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	3	Total	C	N	O	0	0
			45	25	2	18		
4	C	3	Total	C	N	O	0	0
			45	25	2	18		

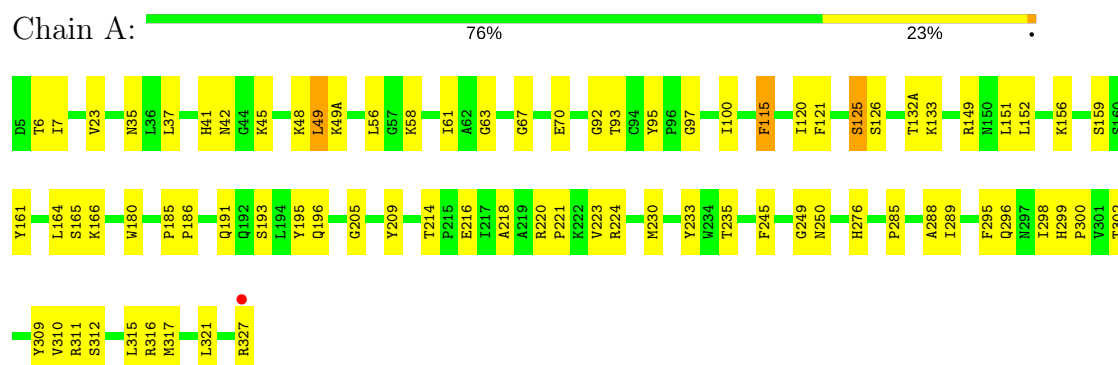
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	21	Total	O	0	0
			21	21		
5	B	6	Total	O	0	0
			6	6		
5	C	7	Total	O	0	0
			7	7		
5	D	2	Total	O	0	0
			2	2		
5	E	23	Total	O	0	0
			23	23		
5	F	4	Total	O	0	0
			4	4		

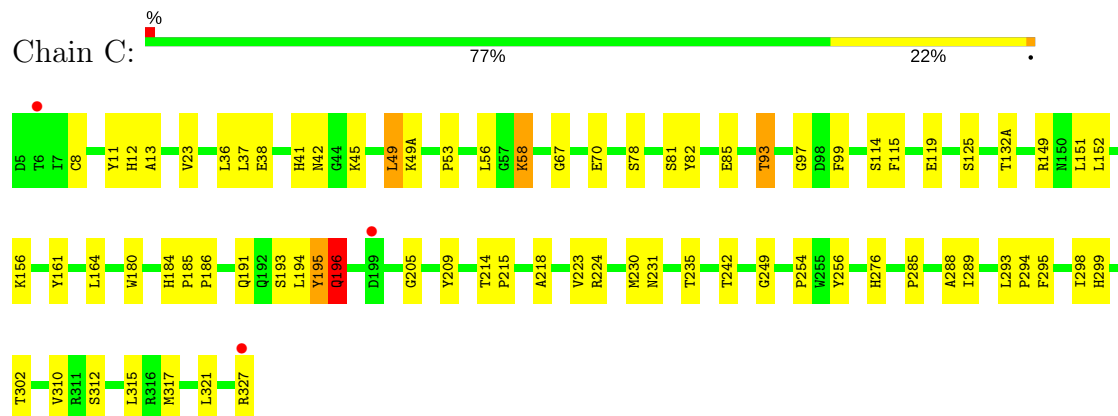
### 3 Residue-property plots [i](#)

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.

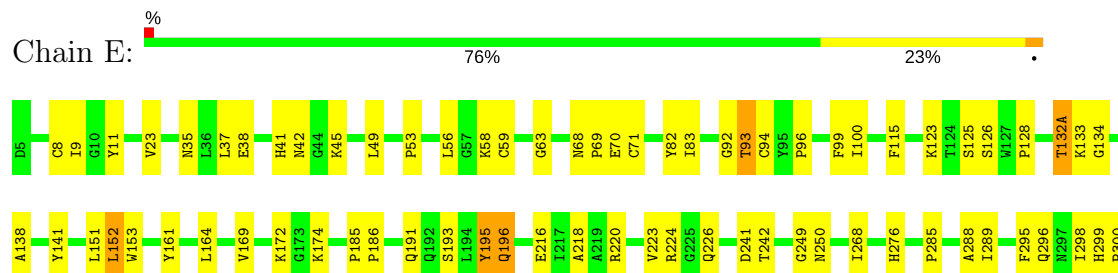
#### • Molecule 1: Hemagglutinin



#### • Molecule 1: Hemagglutinin



#### • Molecule 1: Hemagglutinin





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	71.91Å 242.15Å 71.75Å 90.00° 119.66° 90.00°	Depositor
Resolution (Å)	38.25 – 3.01 43.48 – 3.01	Depositor EDS
% Data completeness (in resolution range)	88.7 (38.25-3.01) 96.3 (43.48-3.01)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.38 (at 3.01Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.5_2)	Depositor
R, $R_{free}$	0.202 , 0.247 0.202 , 0.247	Depositor DCC
$R_{free}$ test set	2042 reflections (5.06%)	DCC
Wilson B-factor (Å <sup>2</sup> )	56.8	Xtriage
Anisotropy	0.483	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 43.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.000 for l,k,-h-l 0.000 for -h-l,k,h 0.017 for -h-l,-k,l 0.019 for h,-k,-h-l 0.015 for l,-k,h	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	11784	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	69.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.67% 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: SIA, GAL, 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	0.31	1/2576 (0.0%)	0.43	0/3508
1	C	0.23	0/2576	0.46	2/3508 (0.1%)
1	E	0.24	0/2576	0.45	2/3508 (0.1%)
2	B	0.28	0/1395	0.43	0/1878
2	D	0.27	0/1344	0.44	0/1811
2	F	0.23	0/1353	0.49	1/1823 (0.1%)
All	All	0.26	1/11820 (0.0%)	0.45	5/16036 (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
2	D	0	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	195	TYR	CE1-CZ	-5.29	1.31	1.38

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	F	664	GLU	N-CA-CB	10.56	129.61	110.60
1	C	195	TYR	CB-CA-C	8.68	127.76	110.40
1	C	196	GLN	CB-CA-C	-8.43	93.53	110.40
1	E	195	TYR	CB-CA-C	8.08	126.57	110.40
1	E	196	GLN	CB-CA-C	-6.39	97.62	110.40



There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	D	662	TYR	Peptide

## 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	2512	0	2440	59	0
1	C	2512	0	2440	56	0
1	E	2512	0	2440	51	0
2	B	1368	0	1285	43	0
2	D	1317	0	1231	42	0
2	F	1326	0	1237	41	0
3	A	28	0	25	1	0
3	C	28	0	25	1	0
3	E	28	0	25	1	0
4	A	45	0	38	2	0
4	C	45	0	38	3	0
5	A	21	0	0	0	0
5	B	6	0	0	0	0
5	C	7	0	0	0	0
5	D	2	0	0	0	0
5	E	23	0	0	0	0
5	F	4	0	0	0	0
All	All	11784	0	11224	273	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 12.

The worst 5 of 273 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:C:196:GLN:OE1	1:C:196:GLN:O	1.71	1.08
2:D:661:LYS:HD3	2:D:662:TYR:N	1.74	1.02
2:F:662:TYR:O	2:F:663:SER:HB3	1.62	0.97

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:F:663:SER:O	2:F:664:GLU:HG2	1.72	0.88
1:E:310:VAL:HG12	1:E:312:SER:H	1.36	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	322/324 (99%)	300 (93%)	21 (6%)	1 (0%)	44	80
1	C	322/324 (99%)	300 (93%)	21 (6%)	1 (0%)	44	80
1	E	322/324 (99%)	303 (94%)	18 (6%)	1 (0%)	44	80
2	B	168/170 (99%)	150 (89%)	13 (8%)	5 (3%)	5	26
2	D	162/170 (95%)	141 (87%)	13 (8%)	8 (5%)	2	14
2	F	163/170 (96%)	139 (85%)	17 (10%)	7 (4%)	3	17
All	All	1459/1482 (98%)	1333 (91%)	103 (7%)	23 (2%)	11	44

5 of 23 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	F	663	SER
2	F	664	GLU
2	B	561	THR
2	B	635	ASN
2	D	643	LYS

### 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	278/278 (100%)	266 (96%)	12 (4%)	33	71
1	C	278/278 (100%)	266 (96%)	12 (4%)	33	71
1	E	278/278 (100%)	267 (96%)	11 (4%)	36	73
2	B	145/145 (100%)	136 (94%)	9 (6%)	21	56
2	D	139/145 (96%)	128 (92%)	11 (8%)	14	45
2	F	140/145 (97%)	127 (91%)	13 (9%)	10	36
All	All	1258/1269 (99%)	1190 (95%)	68 (5%)	26	62

5 of 68 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	196	GLN
2	D	575	ARG
2	F	580	LEU
1	C	302	THR
2	D	526	HIS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 37 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	250	ASN
2	D	629	ASN
2	F	625	GLN
2	D	530	GLN
2	D	572	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

## 5.5 Carbohydrates ⓘ

12 carbohydrates 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	NAG	A	801	1,3	14,14,15	0.48	0	15,19,21	0.72	0
3	NAG	A	802	3	14,14,15	0.51	0	15,19,21	0.65	0
4	SIA	A	803	4	17,20,21	0.39	0	19,28,31	2.45	2 (10%)
4	GAL	A	804	4	11,11,12	0.62	0	13,15,17	0.81	0
4	NAG	A	805	4	14,14,15	0.53	0	15,19,21	0.92	0
3	NAG	C	801	1,3	14,14,15	0.48	0	15,19,21	0.90	1 (6%)
3	NAG	C	802	3	14,14,15	0.48	0	15,19,21	0.62	0
4	SIA	C	803	4	17,20,21	0.33	0	19,28,31	2.69	3 (15%)
4	GAL	C	804	4	11,11,12	0.64	0	13,15,17	0.86	1 (7%)
4	NAG	C	805	4	14,14,15	0.56	0	15,19,21	0.76	0
3	NAG	E	801	1,3	14,14,15	0.56	0	15,19,21	0.66	0
3	NAG	E	802	3	14,14,15	0.54	0	15,19,21	0.79	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
3	NAG	A	801	1,3	-	0/6/23/26	0/1/1/1
3	NAG	A	802	3	-	0/6/23/26	0/1/1/1
4	SIA	A	803	4	-	0/14/34/38	0/1/1/1
4	GAL	A	804	4	-	0/2/19/22	0/1/1/1
4	NAG	A	805	4	-	0/6/23/26	0/1/1/1
3	NAG	C	801	1,3	-	0/6/23/26	0/1/1/1
3	NAG	C	802	3	-	0/6/23/26	0/1/1/1
4	SIA	C	803	4	-	0/14/34/38	0/1/1/1
4	GAL	C	804	4	-	0/2/19/22	0/1/1/1
4	NAG	C	805	4	-	0/6/23/26	0/1/1/1
3	NAG	E	801	1,3	-	0/6/23/26	0/1/1/1
3	NAG	E	802	3	-	0/6/23/26	0/1/1/1

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	C	803	SIA	O6-C2-C3	-10.47	90.97	109.82
4	A	803	SIA	O6-C2-C3	-9.45	92.81	109.82
4	C	803	SIA	C4-C5-N5	2.08	114.68	110.40
3	C	801	NAG	C1-O5-C5	2.28	115.31	112.17
4	C	804	GAL	C1-C2-C3	2.35	112.62	109.65

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

7 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	801	NAG	1	0
4	A	803	SIA	1	0
4	A	805	NAG	1	0
3	C	801	NAG	1	0
4	C	803	SIA	1	0
4	C	805	NAG	2	0
3	E	801	NAG	1	0

## 5.6 Ligand geometry [i](#)

There are no ligands in this entry.

## 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	324/324 (100%)	-0.29	1 (0%)	93 82	29, 54, 85, 187	0
1	C	324/324 (100%)	-0.12	3 (0%)	84 61	40, 69, 101, 176	0
1	E	324/324 (100%)	-0.36	2 (0%)	89 71	30, 56, 91, 168	0
2	B	170/170 (100%)	-0.12	1 (0%)	89 71	42, 65, 94, 134	0
2	D	164/170 (96%)	0.16	5 (3%)	51 23	43, 90, 144, 174	0
2	F	165/170 (97%)	0.16	7 (4%)	37 15	40, 81, 141, 161	0
All	All	1471/1482 (99%)	-0.15	19 (1%)	77 50	29, 64, 118, 187	0

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	327	ARG	6.5
1	E	327	ARG	5.0
1	A	327	ARG	3.8
2	D	533	GLY	3.2
2	F	531	GLY	2.9

### 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](#)

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	SIA	C	803	20/21	0.93	0.17	-0.13	62,78,84,85	0
4	SIA	A	803	20/21	0.96	0.17	-0.42	38,55,67,68	0
3	NAG	A	801	14/15	0.96	0.14	-0.95	43,50,59,60	0
3	NAG	E	801	14/15	0.97	0.14	-1.02	50,54,60,62	0
3	NAG	C	801	14/15	0.96	0.12	-3.22	64,71,75,81	0
3	NAG	C	802	14/15	0.91	0.15	-	70,79,96,96	0
4	GAL	C	804	11/12	0.88	0.19	-	84,94,105,108	0
3	NAG	E	802	14/15	0.92	0.18	-	70,87,94,95	0
3	NAG	A	802	14/15	0.91	0.20	-	63,70,85,87	0
4	NAG	C	805	14/15	0.81	0.29	-	107,114,121,122	0
4	NAG	A	805	14/15	0.83	0.32	-	101,118,124,128	0
4	GAL	A	804	11/12	0.90	0.21	-	75,95,108,109	0

## 6.4 Ligands [i](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.