



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 15, 2017 – 12:51 am GMT

PDB ID : 4CQY  
Title : H5 (tyTy) Del133/Ile155Thr Mutant Haemagglutinin in Complex with Avian Receptor Analogue LSTa  
Authors : Xiong, X.; Xiao, H.; Martin, S.R.; Coombs, P.J.; Liu, J.; Collins, P.J.; Vachieri, S.G.; Walker, P.A.; Lin, Y.P.; McCauley, J.W.; Gamblin, S.J.; Skehel, J.J.  
Deposited on : 2014-02-21  
Resolution : 2.05 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	:	1.7.2 (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

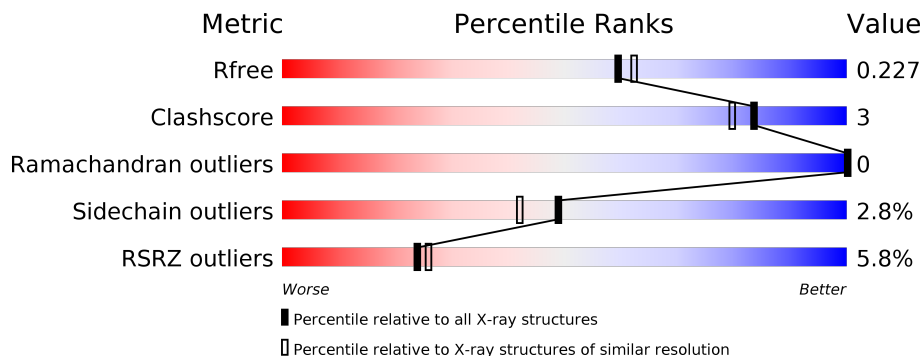
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.05 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	1316 (2.04-2.04)
Clashscore	112137	1394 (2.04-2.04)
Ramachandran outliers	110173	1383 (2.04-2.04)
Sidechain outliers	110143	1383 (2.04-2.04)
RSRZ outliers	101464	1319 (2.04-2.04)

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	327	<div> <div>3%</div> <div>92%</div> <div>7%</div> </div>
1	C	327	<div> <div>2%</div> <div>92%</div> <div>6%</div> </div>
1	E	327	<div> <div>%</div> <div>90%</div> <div>8%</div> </div>
2	B	166	<div> <div>7%</div> <div>86%</div> <div>11%</div> </div>
2	D	166	<div> <div>17%</div> <div>85%</div> <div>13%</div> </div>
2	F	166	<div> <div>15%</div> <div>86%</div> <div>11%</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
3	NAG	A	1324	X	-	-	-
3	NAG	C	1323	-	-	-	X
3	NAG	E	1322	X	-	-	-
5	PO4	C	1327	-	-	-	X
5	PO4	E	1327	-	-	-	X

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 13226 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 HAEMAGGLUTININ HA1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	324	Total	C	N	O	S	0	0	0
			2572	1618	449	491	14			
1	C	322	Total	C	N	O	S	0	0	0
			2549	1606	443	486	14			
1	E	322	Total	C	N	O	S	0	0	0
			2552	1607	443	488	14			

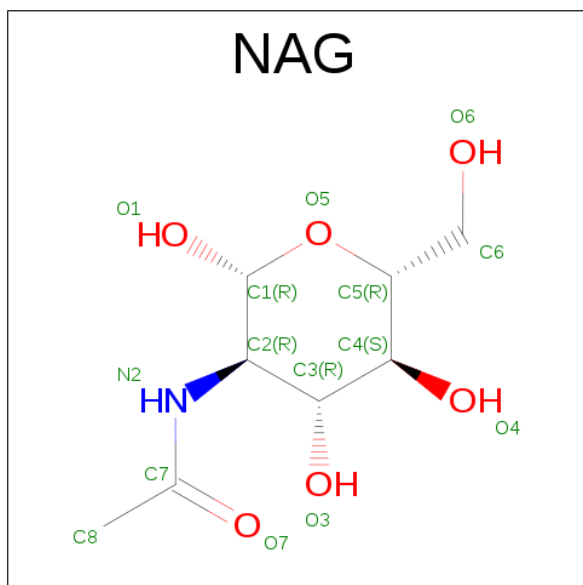
There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	ASP	-	EXPRESSION TAG	UNP Q207Z6
A	0	PRO	-	EXPRESSION TAG	UNP Q207Z6
A	.	-	ALA	DELETION	UNP Q207Z6
A	150	THR	ILE	ENGINEERED MUTATION	UNP Q207Z6
A	322	ARG	GLY	CONFLICT	UNP Q207Z6
A	324	THR	ARG	CONFLICT	UNP Q207Z6
C	-1	ASP	-	EXPRESSION TAG	UNP Q207Z6
C	0	PRO	-	EXPRESSION TAG	UNP Q207Z6
C	.	-	ALA	DELETION	UNP Q207Z6
C	150	THR	ILE	ENGINEERED MUTATION	UNP Q207Z6
C	322	ARG	GLY	CONFLICT	UNP Q207Z6
C	324	THR	ARG	CONFLICT	UNP Q207Z6
E	-1	ASP	-	EXPRESSION TAG	UNP Q207Z6
E	0	PRO	-	EXPRESSION TAG	UNP Q207Z6
E	.	-	ALA	DELETION	UNP Q207Z6
E	150	THR	ILE	ENGINEERED MUTATION	UNP Q207Z6
E	322	ARG	GLY	CONFLICT	UNP Q207Z6
E	324	THR	ARG	CONFLICT	UNP Q207Z6

- Molecule 2 is a protein called HAEMAGGLUTININ HA2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	163	Total	C	N	O	S	0	0	0
			1320	822	229	261	8			
2	D	163	Total	C	N	O	S	0	0	0
			1320	822	229	261	8			
2	F	164	Total	C	N	O	S	0	0	0
			1325	825	230	262	8			

- Molecule 3 is SUGAR (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	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	C	1	Total	C	N	O	0	0
			14	8	1	5		
3	C	1	Total	C	N	O	0	0
			14	8	1	5		
3	C	1	Total	C	N	O	0	0
			14	8	1	5		
3	E	1	Total	C	N	O	0	0
			14	8	1	5		
3	E	1	Total	C	N	O	0	0
			14	8	1	5		

*Continued on next page...*

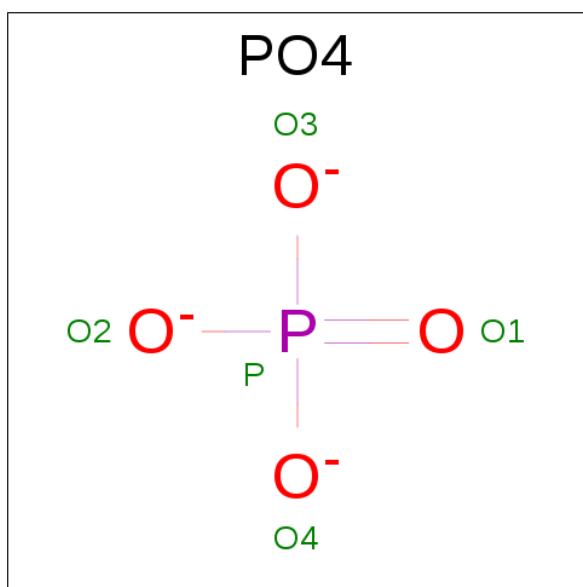
Continued from previous page...

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

- 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
			46	25	2	19		
4	C	3	Total	C	N	O	0	0
			46	25	2	19		
4	E	3	Total	C	N	O	0	0
			46	25	2	19		

- Molecule 5 is PHOSPHATE ION (three-letter code: PO4) (formula: O<sub>4</sub>P).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	C	1	Total	O	P	0	0
			5	4	1		
5	E	1	Total	O	P	0	0
			5	4	1		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	296	Total	O	0	0
			296	296		

Continued on next page...

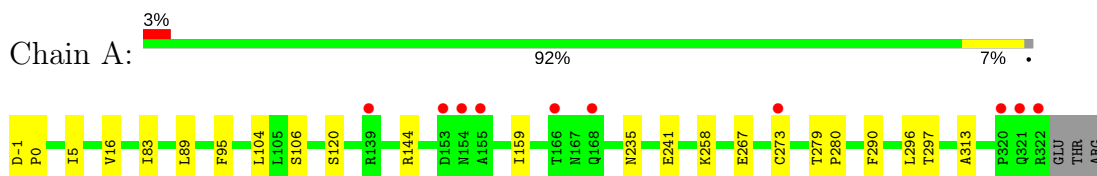
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	B	126	Total 126	O 126	0	0
6	C	366	Total 366	O 366	0	0
6	D	87	Total 87	O 87	0	0
6	E	372	Total 372	O 372	0	0
6	F	67	Total 67	O 67	0	0

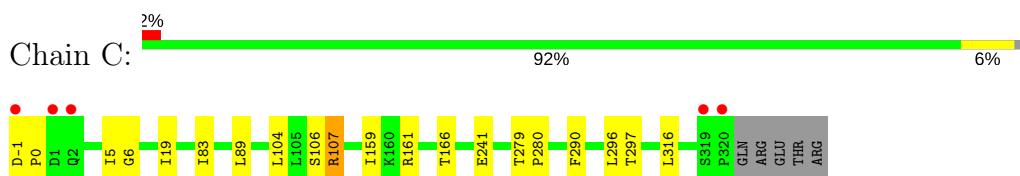
### 3 Residue-property plots [i](#)

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.

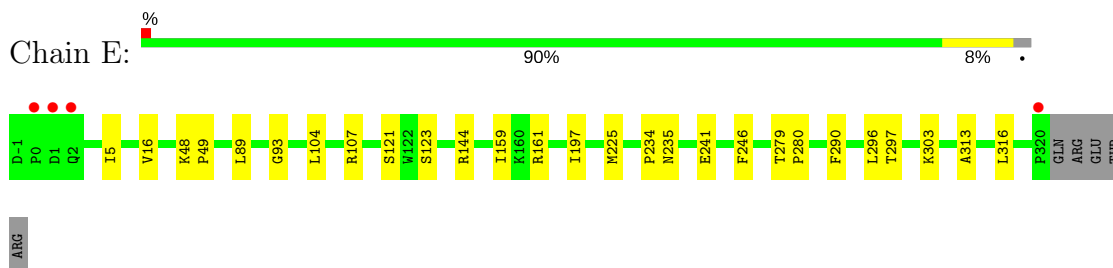
#### • Molecule 1: HAEMAGGLUTININ HA1



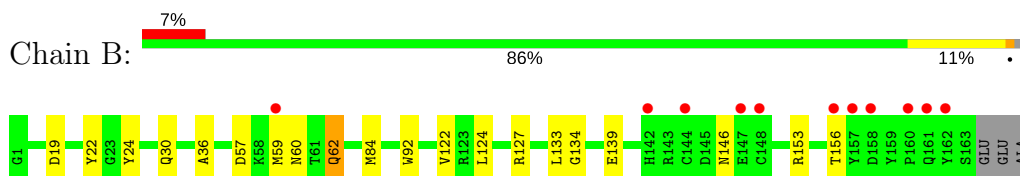
#### • Molecule 1: HAEMAGGLUTININ HA1



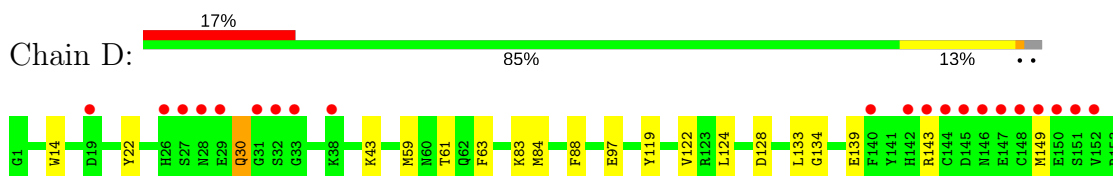
#### • Molecule 1: HAEMAGGLUTININ HA1



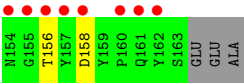
#### • Molecule 2: HAEMAGGLUTININ HA2



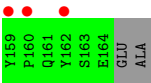
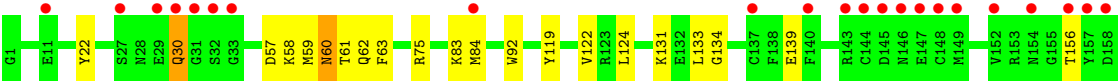
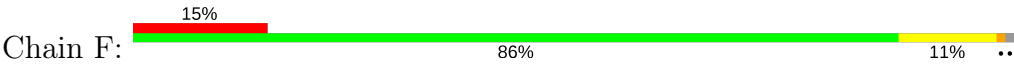
#### • Molecule 2: HAEMAGGLUTININ HA2







● Molecule 2: HAEMAGGLUTININ HA2



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	89.00Å 116.35Å 101.84Å 90.00° 91.69° 90.00°	Depositor
Resolution (Å)	48.74 – 2.05 48.69 – 2.05	Depositor EDS
% Data completeness (in resolution range)	96.6 (48.74-2.05) 96.6 (48.69-2.05)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.07 (at 2.05Å)	Xtriage
Refinement program	REFMAC 5.7.0032	Depositor
R, $R_{free}$	0.193 , 0.227 0.193 , 0.227	Depositor DCC
$R_{free}$ test set	6342 reflections (5.32%)	DCC
Wilson B-factor (Å <sup>2</sup> )	29.1	Xtriage
Anisotropy	0.635	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 43.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.024 for h,-k,-l	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	13226	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.15% 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: PO4, GAL, NAG, SIA

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	0.32	0/2634	0.53	0/3580
1	C	0.34	0/2611	0.54	0/3550
1	E	0.37	0/2614	0.56	0/3554
2	B	0.34	0/1347	0.49	0/1813
2	D	0.32	0/1347	0.50	0/1813
2	F	0.31	0/1352	0.48	0/1820
All	All	0.34	0/11905	0.52	0/16130

There are no bond length outliers.

There are no bond angle outliers.

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	2572	0	2504	10	0
1	C	2549	0	2481	14	0
1	E	2552	0	2483	14	0
2	B	1320	0	1224	13	0
2	D	1320	0	1224	14	0
2	F	1325	0	1226	18	0
3	A	42	0	39	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	C	42	0	39	0	0
3	E	42	0	39	1	0
4	A	46	0	40	0	0
4	C	46	0	40	0	0
4	E	46	0	40	0	0
5	C	5	0	0	1	0
5	E	5	0	0	1	0
6	A	296	0	0	2	0
6	B	126	0	0	1	0
6	C	366	0	0	3	0
6	D	87	0	0	0	0
6	E	372	0	0	1	0
6	F	67	0	0	0	0
All	All	13226	0	11379	65	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 3.

All (65) 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:E:161:ARG:NH2	5:E:1327:PO4:O4	2.12	0.82
1:E:279:THR:HG22	1:E:297:THR:HG22	1.65	0.77
1:A:106:SER:HB3	6:A:2125:HOH:O	1.85	0.77
1:A:5:ILE:HD11	2:B:122:VAL:HG21	1.72	0.71
2:B:19:ASP:HB3	2:B:36:ALA:HB2	1.75	0.68
1:E:290:PHE:HZ	2:F:59:MET:HG3	1.62	0.65
2:B:30:GLN:HE22	2:B:146:ASN:H	1.45	0.62
1:A:83:ILE:HD12	1:A:267:GLU:OE2	1.99	0.62
2:B:57:ASP:O	2:B:60:ASN:HB2	1.99	0.61
1:E:5:ILE:HD11	2:F:122:VAL:HG21	1.83	0.59
2:D:30:GLN:H	2:D:30:GLN:HE21	1.51	0.58
2:F:133:LEU:HD11	2:F:139:GLU:HB2	1.85	0.58
1:C:279:THR:HG22	1:C:297:THR:HG22	1.84	0.57
1:A:280:PRO:HD3	1:A:296:LEU:O	2.05	0.57
1:A:279:THR:HG22	1:A:297:THR:HG22	1.88	0.56
1:C:161:ARG:HH22	5:C:1327:PO4:P	2.30	0.55
1:A:290:PHE:HZ	2:B:59:MET:HG3	1.72	0.54
1:A:-1:ASP:N	1:A:0:PRO:CD	2.73	0.52
1:C:-1:ASP:N	6:C:2001:HOH:O	2.43	0.51
1:E:16:VAL:HG21	1:E:313:ALA:HB2	1.92	0.51

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:5:ILE:HD11	2:D:122:VAL:HG21	1.93	0.50
2:B:127:ARG:HH22	2:F:131:LYS:HD3	1.75	0.50
2:F:62:GLN:HG3	2:F:92:TRP:CD2	2.46	0.49
2:B:84:MET:HG3	2:D:84:MET:HG3	1.95	0.49
1:A:258:LYS:HD2	6:A:2181:HOH:O	2.13	0.48
1:C:280:PRO:HD3	1:C:296:LEU:O	2.13	0.48
2:D:97:GLU:HB3	2:F:58:LYS:HE3	1.95	0.48
1:C:5:ILE:HG13	2:D:119:TYR:HA	1.95	0.48
2:B:62:GLN:HG2	2:B:92:TRP:CD2	2.48	0.47
2:D:134:GLY:HA2	2:F:124:LEU:HD22	1.95	0.47
1:C:290:PHE:HZ	2:D:59:MET:HG3	1.80	0.47
1:C:0:PRO:HA	2:D:143:ARG:HH12	1.80	0.47
6:B:2061:HOH:O	2:F:83:LYS:HD2	2.13	0.47
1:E:234:PRO:O	1:E:235:ASN:HB2	2.13	0.47
2:D:133:LEU:HD11	2:D:139:GLU:HB2	1.97	0.46
1:A:159:ILE:O	1:A:241:GLU:HA	2.16	0.46
1:A:16:VAL:HG21	1:A:313:ALA:HB2	1.98	0.46
1:C:159:ILE:O	1:C:241:GLU:HA	2.16	0.45
1:C:-1:ASP:N	1:C:0:PRO:HD2	2.32	0.45
1:E:159:ILE:O	1:E:241:GLU:HA	2.16	0.45
2:D:84:MET:HE3	2:D:88:PHE:HD2	1.80	0.45
1:E:48:LYS:HG3	1:E:49:PRO:HD2	1.99	0.45
2:B:127:ARG:NH2	2:F:131:LYS:HD3	2.32	0.45
2:F:63:PHE:CZ	2:F:84:MET:CE	3.00	0.45
2:B:134:GLY:HA2	2:D:124:LEU:HD22	1.99	0.44
2:F:30:GLN:HE21	2:F:30:GLN:H	1.65	0.44
2:F:57:ASP:O	2:F:60:ASN:HB2	2.17	0.44
1:E:280:PRO:HD3	1:E:296:LEU:O	2.18	0.44
1:E:5:ILE:HG13	2:F:119:TYR:HA	1.99	0.43
3:E:1322:NAG:H81	6:E:2356:HOH:O	2.19	0.43
2:F:84:MET:HE2	2:F:84:MET:C	2.38	0.43
2:B:133:LEU:HD11	2:B:139:GLU:HB2	2.00	0.42
2:F:83:LYS:HA	2:F:83:LYS:HD3	1.91	0.42
1:E:303:LYS:HE3	2:F:60:ASN:O	2.19	0.42
1:E:93:GLY:HA3	1:E:225:MET:O	2.20	0.42
2:D:63:PHE:CZ	2:D:84:MET:HE1	2.54	0.42
1:C:104:LEU:HD13	6:C:2156:HOH:O	2.20	0.41
2:B:124:LEU:HD22	2:F:134:GLY:HA2	2.01	0.41
1:C:106:SER:HB2	1:C:107:ARG:HE	1.85	0.41
1:C:166:THR:HG22	6:C:2233:HOH:O	2.21	0.41
1:E:303:LYS:HD2	2:F:62:GLN:HB3	2.02	0.41

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:24:TYR:CD1	2:B:153:ARG:HD3	2.56	0.40
2:D:83:LYS:HD3	2:D:83:LYS:HA	1.95	0.40
1:E:197:ILE:HD11	1:E:246:PHE:HA	2.04	0.40
1:C:6:GLY:HA3	2:D:14:TRP:CZ3	2.57	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	322/327 (98%)	312 (97%)	10 (3%)	0	100	100
1	C	320/327 (98%)	313 (98%)	7 (2%)	0	100	100
1	E	320/327 (98%)	311 (97%)	9 (3%)	0	100	100
2	B	161/166 (97%)	159 (99%)	2 (1%)	0	100	100
2	D	161/166 (97%)	157 (98%)	4 (2%)	0	100	100
2	F	162/166 (98%)	158 (98%)	4 (2%)	0	100	100
All	All	1446/1479 (98%)	1410 (98%)	36 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 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	291/294 (99%)	284 (98%)	7 (2%)	54	48
1	C	288/294 (98%)	283 (98%)	5 (2%)	66	63
1	E	289/294 (98%)	282 (98%)	7 (2%)	54	48
2	B	139/141 (99%)	136 (98%)	3 (2%)	57	51
2	D	139/141 (99%)	131 (94%)	8 (6%)	23	13
2	F	139/141 (99%)	133 (96%)	6 (4%)	33	25
All	All	1285/1305 (98%)	1249 (97%)	36 (3%)	49	42

All (36) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	89	LEU
1	A	95	PHE
1	A	104	LEU
1	A	120	SER
1	A	144	ARG
1	A	235	ASN
1	A	273	CYS
2	B	22	TYR
2	B	62	GLN
2	B	156	THR
1	C	19	ILE
1	C	83	ILE
1	C	89	LEU
1	C	107	ARG
1	C	316	LEU
2	D	22	TYR
2	D	30	GLN
2	D	43	LYS
2	D	61	THR
2	D	128	ASP
2	D	149	MET
2	D	156	THR
2	D	158	ASP
1	E	89	LEU
1	E	104	LEU
1	E	107	ARG
1	E	121	SER
1	E	123	SER
1	E	144	ARG
1	E	316	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
2	F	22	TYR
2	F	30	GLN
2	F	60	ASN
2	F	61	THR
2	F	75	ARG
2	F	156	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	115	GLN
2	B	30	GLN
1	C	235	ASN
2	D	30	GLN
2	F	30	GLN

### 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 ⓘ

9 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	$\# Z  > 2$
4	SIA	A	1326	4	17,20,21	0.36	0	19,28,31	0.89	1 (5%)
4	GAL	A	1327	4	11,11,12	0.41	0	13,15,17	0.98	1 (7%)



Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
4	NAG	A	1328	4	15,15,15	0.41	0	21,21,21	1.16	1 (4%)
4	SIA	C	1324	4	17,20,21	0.64	0	19,28,31	0.85	0
4	GAL	C	1325	4	11,11,12	0.40	0	13,15,17	0.96	0
4	NAG	C	1326	4	15,15,15	0.45	0	21,21,21	1.46	2 (9%)
4	SIA	E	1324	4	17,20,21	0.53	0	19,28,31	0.91	1 (5%)
4	GAL	E	1325	4	11,11,12	0.37	0	13,15,17	0.73	0
4	NAG	E	1326	4	15,15,15	0.42	0	21,21,21	1.51	3 (14%)

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	SIA	A	1326	4	-	0/14/34/38	0/1/1/1
4	GAL	A	1327	4	-	0/2/19/22	0/1/1/1
4	NAG	A	1328	4	-	0/6/26/26	0/1/1/1
4	SIA	C	1324	4	-	0/14/34/38	0/1/1/1
4	GAL	C	1325	4	-	0/2/19/22	0/1/1/1
4	NAG	C	1326	4	-	0/6/26/26	0/1/1/1
4	SIA	E	1324	4	-	0/14/34/38	0/1/1/1
4	GAL	E	1325	4	-	0/2/19/22	0/1/1/1
4	NAG	E	1326	4	-	0/6/26/26	0/1/1/1

There are no bond length outliers.

All (9) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	C	1326	NAG	C1-C2-N2	-5.10	104.82	110.73
4	E	1326	NAG	C1-C2-N2	-5.03	104.90	110.73
4	A	1328	NAG	C1-C2-N2	-4.09	106.00	110.73
4	E	1324	SIA	C6-C5-N5	-2.90	105.91	111.00
4	A	1326	SIA	C4-C5-N5	-2.72	104.78	110.40
4	E	1326	NAG	O5-C1-C2	-2.07	107.44	109.52
4	E	1326	NAG	C3-C4-C5	2.55	114.70	110.22
4	C	1326	NAG	C3-C4-C5	2.63	114.84	110.22
4	A	1327	GAL	C1-C2-C3	3.15	113.64	109.65

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.6 Ligand geometry

11 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	NAG	A	1323	1	14,14,15	0.46	0	15,19,21	1.27	2 (13%)
3	NAG	A	1324	1	14,14,15	0.59	0	15,19,21	1.42	3 (20%)
3	NAG	A	1325	1	14,14,15	0.43	0	15,19,21	1.23	1 (6%)
3	NAG	C	1321	1	14,14,15	0.45	0	15,19,21	1.31	1 (6%)
3	NAG	C	1322	1	14,14,15	0.43	0	15,19,21	1.32	2 (13%)
3	NAG	C	1323	1	14,14,15	0.40	0	15,19,21	1.04	1 (6%)
5	PO4	C	1327	-	4,4,4	0.68	0	6,6,6	0.46	0
3	NAG	E	1321	1	14,14,15	0.52	0	15,19,21	1.29	2 (13%)
3	NAG	E	1322	1	14,14,15	0.56	0	15,19,21	1.26	3 (20%)
3	NAG	E	1323	1	14,14,15	0.49	0	15,19,21	1.12	2 (13%)
5	PO4	E	1327	-	4,4,4	0.83	0	6,6,6	0.48	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	1323	1	-	0/6/23/26	0/1/1/1
3	NAG	A	1324	1	1/1/5/7	0/6/23/26	0/1/1/1
3	NAG	A	1325	1	-	0/6/23/26	0/1/1/1
3	NAG	C	1321	1	-	0/6/23/26	0/1/1/1
3	NAG	C	1322	1	-	0/6/23/26	0/1/1/1
3	NAG	C	1323	1	-	0/6/23/26	0/1/1/1
5	PO4	C	1327	-	-	0/0/0/0	0/0/0/0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	NAG	E	1321	1	-	0/6/23/26	0/1/1/1
3	NAG	E	1322	1	1/1/5/7	0/6/23/26	0/1/1/1
3	NAG	E	1323	1	-	0/6/23/26	0/1/1/1
5	PO4	E	1327	-	-	0/0/0/0	0/0/0/0

There are no bond length outliers.

All (17) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	E	1322	NAG	C1-C2-N2	-2.34	106.50	110.49
3	A	1323	NAG	C6-C5-C4	-2.05	108.21	113.00
3	E	1321	NAG	C8-C7-N2	2.03	119.78	116.11
3	E	1323	NAG	C1-C2-N2	2.04	113.97	110.49
3	C	1322	NAG	C8-C7-N2	2.17	120.03	116.11
3	A	1324	NAG	C8-C7-N2	2.19	120.06	116.11
3	E	1322	NAG	C4-C3-C2	2.22	114.27	111.02
3	C	1323	NAG	C1-O5-C5	2.32	115.36	112.17
3	E	1323	NAG	C1-O5-C5	2.45	115.55	112.17
3	A	1324	NAG	C2-N2-C7	2.62	126.77	122.94
3	A	1324	NAG	C1-O5-C5	2.64	115.81	112.17
3	E	1322	NAG	C2-N2-C7	2.73	126.92	122.94
3	A	1323	NAG	C1-O5-C5	2.75	115.96	112.17
3	C	1321	NAG	C1-O5-C5	2.92	116.19	112.17
3	E	1321	NAG	C1-O5-C5	3.27	116.68	112.17
3	C	1322	NAG	C1-O5-C5	3.33	116.75	112.17
3	A	1325	NAG	C1-O5-C5	3.90	117.53	112.17

All (2) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
3	A	1324	NAG	C1
3	E	1322	NAG	C1

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	C	1327	PO4	1	0
3	E	1322	NAG	1	0
5	E	1327	PO4	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	324/327 (99%)	-0.11	10 (3%) 49 55	24, 40, 62, 97	0
1	C	322/327 (98%)	-0.20	5 (1%) 72 75	21, 32, 50, 82	0
1	E	322/327 (98%)	-0.18	4 (1%) 79 81	19, 30, 45, 89	0
2	B	163/166 (98%)	0.29	11 (6%) 19 20	19, 36, 67, 84	0
2	D	163/166 (98%)	0.96	29 (17%) 2 1	21, 47, 113, 137	0
2	F	164/166 (98%)	0.84	25 (15%) 2 2	23, 47, 140, 196	0
All	All	1458/1479 (98%)	0.13	84 (5%) 24 26	19, 35, 82, 196	0

All (84) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	156	THR	8.3
2	D	29	GLU	7.1
2	F	31	GLY	6.9
2	F	144	CYS	6.7
2	D	143	ARG	6.2
2	D	157	TYR	5.8
2	D	152	VAL	5.8
2	D	162	TYR	5.7
2	D	144	CYS	5.7
2	D	148	CYS	5.6
2	D	147	GLU	4.9
2	F	162	TYR	4.8
1	A	322	ARG	4.6
2	D	33	GLY	4.5
2	D	150	GLU	4.5
2	D	32	SER	4.3
2	D	158	ASP	4.1
1	C	320	PRO	4.0
2	F	147	GLU	3.9

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	D	161	GLN	3.8
2	D	160	PRO	3.7
2	F	156	THR	3.6
2	F	143	ARG	3.6
2	D	145	ASP	3.6
1	A	155	ALA	3.6
2	F	29	GLU	3.6
2	F	140	PHE	3.5
2	D	38	LYS	3.4
2	F	137	CYS	3.4
2	D	142	HIS	3.4
2	D	26	HIS	3.4
2	F	145	ASP	3.1
1	C	1	ASP	3.0
1	A	320	PRO	2.9
2	B	162	TYR	2.9
1	E	0	PRO	2.8
2	D	31	GLY	2.8
2	D	149	MET	2.8
2	D	146	ASN	2.8
2	F	152	VAL	2.8
2	F	32	SER	2.7
2	B	157	TYR	2.7
2	F	149	MET	2.7
2	D	27	SER	2.6
1	A	273	CYS	2.6
1	A	321	GLN	2.6
2	F	159	TYR	2.6
1	E	1	ASP	2.6
2	B	161	GLN	2.6
2	F	148	CYS	2.6
2	D	154	ASN	2.5
2	F	27	SER	2.5
2	F	146	ASN	2.5
2	D	151	SER	2.4
2	F	11	GLU	2.4
1	E	2	GLN	2.3
2	B	147	GLU	2.3
2	D	28	ASN	2.3
1	A	154	ASN	2.3
2	F	160	PRO	2.3
2	B	59	MET	2.3

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	C	-1	ASP	2.3
2	F	33	GLY	2.3
2	F	157	TYR	2.2
1	A	139	ARG	2.2
1	C	2	GLN	2.2
2	B	160	PRO	2.2
2	B	156	THR	2.2
1	A	168	GLN	2.2
2	B	144	CYS	2.2
2	F	84	MET	2.2
1	A	166	THR	2.2
2	F	158	ASP	2.2
2	D	155	GLY	2.1
2	F	154	ASN	2.1
2	B	158	ASP	2.1
1	C	319	SER	2.1
2	D	140	PHE	2.1
2	B	148	CYS	2.1
1	E	320	PRO	2.1
2	F	30	GLN	2.1
2	B	142	HIS	2.0
1	A	153	ASP	2.0
2	D	19	ASP	2.0

## 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(Å <sup>2</sup> )	Q<0.9
4	SIA	C	1324	20/21	0.96	0.10	0.25	23,25,33,34	0
4	SIA	A	1326	20/21	0.91	0.11	0.10	35,42,48,50	0
4	GAL	E	1325	11/12	0.96	0.09	-0.31	31,38,42,43	0

*Continued on next page...*

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
4	SIA	E	1324	20/21	0.96	0.09	-1.05	25,26,29,30	0
4	GAL	C	1325	11/12	0.95	0.09	-	32,34,37,40	0
4	NAG	E	1326	15/15	0.91	0.24	-	47,59,62,64	0
4	NAG	A	1328	15/15	0.86	0.33	-	64,73,75,76	0
4	NAG	C	1326	15/15	0.84	0.34	-	46,61,63,68	0
4	GAL	A	1327	11/12	0.91	0.12	-	47,51,56,57	0

## 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
5	PO4	E	1327	5/5	0.78	0.55	43.47	56,67,70,70	0
5	PO4	C	1327	5/5	0.82	0.52	17.00	74,78,80,82	0
3	NAG	C	1323	14/15	0.92	0.17	2.32	36,39,43,48	0
3	NAG	A	1325	14/15	0.89	0.20	0.94	37,44,48,50	0
3	NAG	E	1323	14/15	0.93	0.14	0.72	28,32,38,40	0
3	NAG	A	1323	14/15	0.94	0.10	-0.48	30,36,41,41	0
3	NAG	A	1324	14/15	0.85	0.23	-	48,59,66,66	0
3	NAG	C	1321	14/15	0.85	0.23	-	48,55,59,60	0
3	NAG	C	1322	14/15	0.78	0.25	-	50,61,64,65	0
3	NAG	E	1322	14/15	0.81	0.28	-	51,62,66,66	0
3	NAG	E	1321	14/15	0.88	0.18	-	40,49,52,54	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.