



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 1, 2016 – 03:06 PM GMT

PDB ID : 4BH0  
Title : H5 (tyTy) Influenza Virus Haemagglutinin in Complex with Human Receptor Analogue 6'-SLN  
Authors : Xiong, X.; Coombs, P.J.; Martin, S.R.; Liu, J.; Xiao, H.; McCauley, J.W.; Locher, K.; Walker, P.A.; Collins, P.J.; Kawaoka, Y.; Skehel, J.J.; Gamblin, S.J.  
Deposited on : 2013-03-29  
Resolution : 2.36 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 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
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) : trunk26865

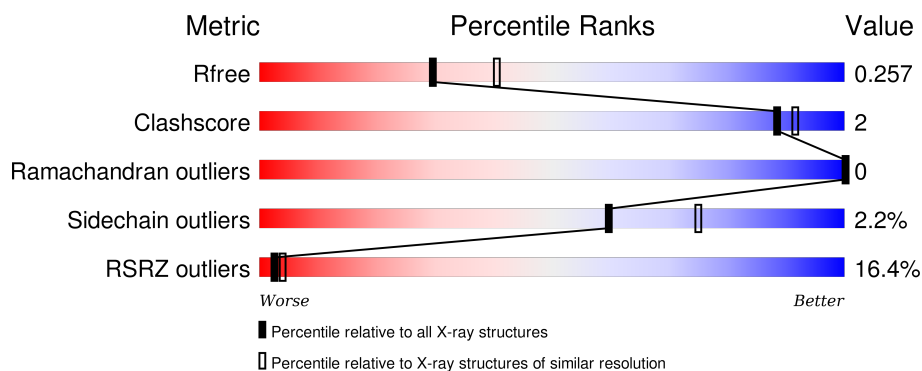
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.36 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	1352 (2.38-2.34)
Clashscore	102246	1456 (2.38-2.34)
Ramachandran outliers	100387	1435 (2.38-2.34)
Sidechain outliers	100360	1436 (2.38-2.34)
RSRZ outliers	91569	1358 (2.38-2.34)

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>5%</div> <div>90%</div> <div>7% ..</div> </div>
1	C	327	<div> <div>5%</div> <div>91%</div> <div>7% .</div> </div>
1	E	327	<div> <div>3%</div> <div>90%</div> <div>8% .</div> </div>
2	B	166	<div> <div>44%</div> <div>83%</div> <div>. . 12%</div> </div>
2	D	166	<div> <div>46%</div> <div>84%</div> <div>5% . 11%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	F	166	

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
5	PO4	A	1323	-	-	-	X
5	PO4	A	1325	-	-	-	X
5	PO4	C	1324	-	-	-	X
5	PO4	E	1324	-	-	-	X

## 2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 11922 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	319	Total	C	N	O	S	0	0	0
			2526	1592	439	481	14			
1	C	319	Total	C	N	O	S	0	0	0
			2519	1589	439	477	14			
1	E	320	Total	C	N	O	S	0	0	0
			2523	1591	440	478	14			

There are 15 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	PRO	-	EXPRESSION TAG	UNP Q207Z6
A	323	ARG	-	EXPRESSION TAG	UNP Q207Z6
A	324	GLU	-	EXPRESSION TAG	UNP Q207Z6
A	325	THR	-	EXPRESSION TAG	UNP Q207Z6
A	326	ARG	-	EXPRESSION TAG	UNP Q207Z6
C	0	PRO	-	EXPRESSION TAG	UNP Q207Z6
C	323	ARG	-	EXPRESSION TAG	UNP Q207Z6
C	324	GLU	-	EXPRESSION TAG	UNP Q207Z6
C	325	THR	-	EXPRESSION TAG	UNP Q207Z6
C	326	ARG	-	EXPRESSION TAG	UNP Q207Z6
E	0	PRO	-	EXPRESSION TAG	UNP Q207Z6
E	323	ARG	-	EXPRESSION TAG	UNP Q207Z6
E	324	GLU	-	EXPRESSION TAG	UNP Q207Z6
E	325	THR	-	EXPRESSION TAG	UNP Q207Z6
E	326	ARG	-	EXPRESSION TAG	UNP Q207Z6

- Molecule 2 is a protein called HEMAGGLUTININ.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	146	Total	C	N	O	S	0	0	0
			1131	698	199	226	8			

*Continued on next page...*

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	148	Total	C	N	O	S	0	0	0
			1159	718	203	230	8			
2	F	148	Total	C	N	O	S	0	0	0
			1161	716	207	230	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	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		

- Molecule 4 is a polymer of unknown type called SUGAR (2-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	2	Total	C	N	O	0	0
			32	17	1	14		

- Molecule 5 is PHOSPHATE ION (three-letter code: PO4) (formula:  $O_4P$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	O	P	0	0
			5	4	1		
5	A	1	Total	O	P	0	0
			5	4	1		
5	A	1	Total	O	P	0	0
			5	4	1		
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 a polymer of unknown type called SUGAR (3-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
6	C	3	Total	C	N	O	0	0
			46	25	2	19		
6	E	3	Total	C	N	O	0	0
			46	25	2	19		

- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	180	Total	O	0	0
			180	180		
7	B	53	Total	O	0	0
			53	53		

*Continued on next page...*

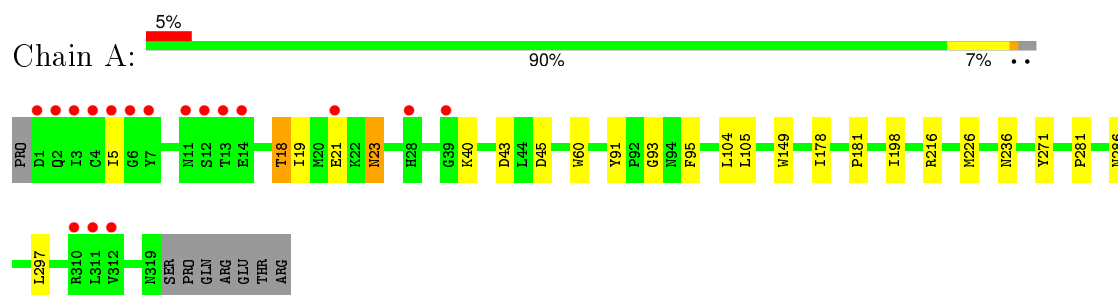
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	C	208	Total 208	O 208	0	0
7	D	44	Total 44	O 44	0	0
7	E	176	Total 176	O 176	0	0
7	F	51	Total 51	O 51	0	0

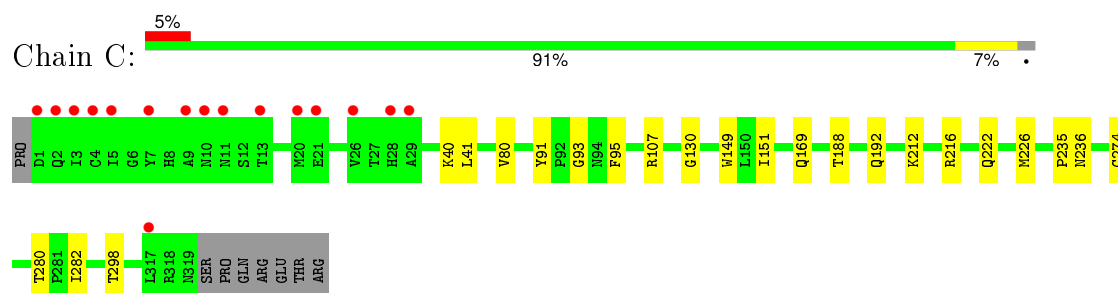
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

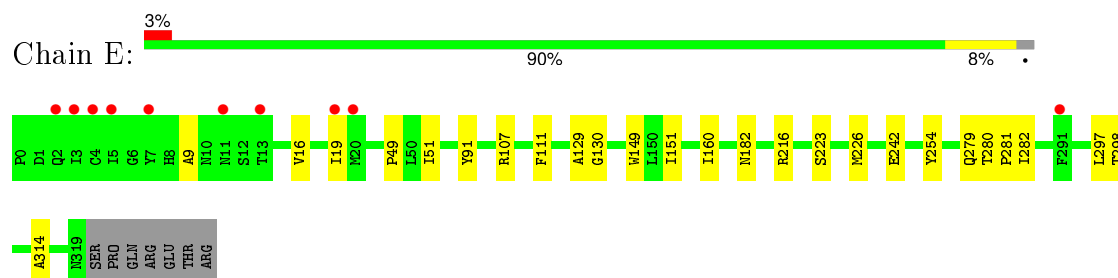
#### • Molecule 1: HEMAGGLUTININ



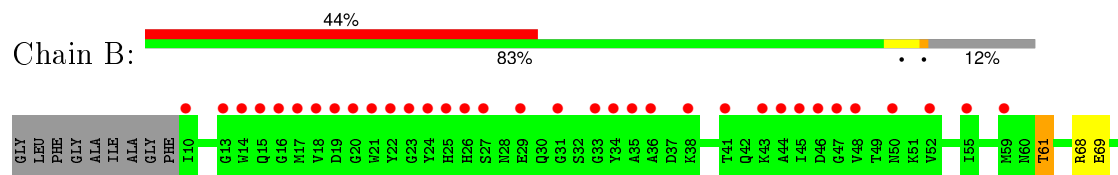
#### • Molecule 1: HEMAGGLUTININ



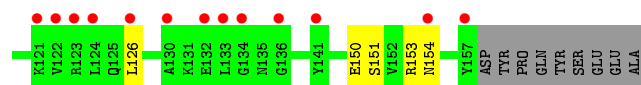
#### • Molecule 1: HEMAGGLUTININ



#### • Molecule 2: HEMAGGLUTININ







## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	70.35Å 228.30Å 71.27Å 90.00° 114.01° 90.00°	Depositor
Resolution (Å)	56.61 – 2.36 56.55 – 2.36	Depositor EDS
% Data completeness (in resolution range)	97.0 (56.61-2.36) 93.1 (56.55-2.36)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.88 (at 2.37Å)	Xtriage
Refinement program	REFMAC 5.7.0032	Depositor
R, $R_{free}$	0.211 , 0.258 0.213 , 0.257	Depositor DCC
$R_{free}$ test set	4094 reflections (5.29%)	DCC
Wilson B-factor (Å <sup>2</sup> )	40.9	Xtriage
Anisotropy	0.633	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	(Not available) , (Not available)	EDS
Estimated twinning fraction	0.067 for l,-k,h	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.42$ , $\langle L^2 \rangle = 0.24$	Xtriage
Outliers	0 of 81446 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	11922	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	63.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.53% 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, 0.375 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: 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	0.37	0/2586	0.57	0/3515
1	C	0.36	0/2579	0.58	0/3506
1	E	0.37	0/2584	0.58	2/3513 (0.1%)
2	B	0.31	0/1152	0.40	0/1556
2	D	0.30	0/1181	0.43	0/1594
2	F	0.32	0/1182	0.45	0/1594
All	All	0.35	0/11264	0.53	2/15278 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	216	ARG	NE-CZ-NH2	-6.15	117.22	120.30
1	E	216	ARG	NE-CZ-NH1	5.59	123.10	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	2526	0	2460	15	0
1	C	2519	0	2454	12	0
1	E	2523	0	2452	16	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	1131	0	993	7	0
2	D	1159	0	1030	3	0
2	F	1161	0	1041	5	0
3	A	14	0	13	0	0
3	C	14	0	13	0	0
3	E	14	0	13	0	0
4	A	32	0	28	1	0
5	A	15	0	0	1	0
5	C	5	0	0	0	0
5	E	5	0	0	0	0
6	C	46	0	40	3	0
6	E	46	0	40	0	0
7	A	180	0	0	3	0
7	B	53	0	0	2	0
7	C	208	0	0	2	0
7	D	44	0	0	0	0
7	E	176	0	0	3	0
7	F	51	0	0	0	0
All	All	11922	0	10577	54	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 (54) 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:129:ALA:HB3	7:E:2089:HOH:O	1.76	0.84
1:C:280:THR:HG22	1:C:298:THR:HG22	1.68	0.75
1:A:286:ASN:ND2	7:A:2157:HOH:O	2.21	0.65
5:A:1325:PO4:O1	7:A:2180:HOH:O	2.14	0.64
1:C:169:GLN:NE2	7:C:2137:HOH:O	2.08	0.61
1:A:5:ILE:HD11	2:B:122:VAL:HG21	1.91	0.52
1:C:212:LYS:O	1:C:216:ARG:NH2	2.42	0.52
1:C:91:TYR:CD1	1:C:226:MET:HG2	2.45	0.52
2:B:68:ARG:NH1	7:B:2021:HOH:O	2.26	0.51
1:A:93:GLY:HA3	1:A:226:MET:O	2.10	0.51
1:E:19:ILE:HD11	2:F:102:MET:HG2	1.91	0.51
1:A:281:PRO:HD3	1:A:297:LEU:O	2.10	0.50
1:E:281:PRO:HD3	1:E:297:LEU:O	2.10	0.50
1:E:130:GLY:HA3	1:E:149:TRP:HB3	1.94	0.50
1:C:93:GLY:HA3	1:C:226:MET:O	2.14	0.48

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:D:72:ASN:HB2	2:D:75:ARG:NH2	2.29	0.48
1:C:188:THR:HG23	7:C:2146:HOH:O	2.13	0.48
1:E:151:ILE:HG12	7:E:2089:HOH:O	2.15	0.47
1:E:129:ALA:CB	7:E:2089:HOH:O	2.50	0.47
1:C:41:LEU:HD13	1:C:80:VAL:HG21	1.96	0.46
1:E:280:THR:HG22	1:E:298:THR:HG22	1.98	0.46
1:E:111:PHE:CD1	1:E:254:TYR:HB3	2.51	0.45
1:A:149:TRP:CZ2	4:A:1321:SIA:H112	2.52	0.45
1:A:91:TYR:CD1	1:A:226:MET:HG2	2.52	0.45
2:B:61:THR:O	2:B:61:THR:OG1	2.32	0.45
1:E:16:VAL:HG21	1:E:314:ALA:HB2	1.99	0.44
2:F:24:TYR:CD1	2:F:153:ARG:HG2	2.52	0.44
6:C:1323:NAG:H82	6:C:1323:NAG:H1	1.99	0.44
1:C:235:PRO:O	1:C:236:ASN:HB2	2.17	0.44
1:E:279:GLN:NE2	1:E:280:THR:O	2.51	0.44
1:E:49:PRO:HG2	1:E:51:ILE:HD11	2.00	0.44
1:C:222:GLN:HE22	6:C:1322:GAL:H5	1.83	0.44
1:C:151:ILE:HG21	6:C:1321:SIA:H111	1.99	0.44
2:F:150:GLU:O	2:F:154:ASN:N	2.49	0.43
1:E:9:ALA:HB2	2:F:13:GLY:HA3	2.00	0.43
1:A:60:TRP:HZ3	1:A:105:LEU:HD21	1.82	0.43
1:A:178:ILE:HD12	1:A:198:ILE:HD12	2.01	0.43
1:A:18:THR:HG22	1:A:21:GLU:HB2	1.99	0.43
1:A:19:ILE:HD11	2:B:102:MET:CG	2.49	0.42
1:A:19:ILE:HD11	2:B:102:MET:HG2	2.00	0.42
2:D:28:ASN:HD21	2:D:145:ASP:HA	1.84	0.42
1:C:188:THR:HG22	1:C:192:GLN:O	2.20	0.42
1:E:91:TYR:CD1	1:E:226:MET:HG2	2.55	0.42
2:B:150:GLU:HG2	7:B:2052:HOH:O	2.20	0.41
1:E:160:ILE:O	1:E:242:GLU:HA	2.19	0.41
1:E:182:ASN:OD1	1:E:223:SER:HB3	2.19	0.41
2:D:150:GLU:O	2:D:154:ASN:N	2.53	0.41
1:A:43:ASP:OD1	1:A:271:TYR:OH	2.30	0.41
1:A:181:PRO:O	1:A:216:ARG:NH2	2.53	0.41
1:C:130:GLY:HA3	1:C:149:TRP:HB3	2.03	0.41
1:E:19:ILE:HD11	2:F:102:MET:CG	2.50	0.41
1:A:104:LEU:HD11	7:A:2143:HOH:O	2.20	0.41
1:A:23:ASN:N	1:A:23:ASN:HD22	2.19	0.41
2:B:150:GLU:O	2:B:154:ASN:N	2.50	0.40

There are no symmetry-related clashes.

## 5.3 Torsion angles ⓘ

### 5.3.1 Protein backbone ⓘ

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	317/327 (97%)	303 (96%)	14 (4%)	0	100	100
1	C	317/327 (97%)	301 (95%)	16 (5%)	0	100	100
1	E	318/327 (97%)	307 (96%)	11 (4%)	0	100	100
2	B	144/166 (87%)	139 (96%)	5 (4%)	0	100	100
2	D	146/166 (88%)	139 (95%)	7 (5%)	0	100	100
2	F	146/166 (88%)	145 (99%)	1 (1%)	0	100	100
All	All	1388/1479 (94%)	1334 (96%)	54 (4%)	0	100	100

There are no Ramachandran outliers to report.

### 5.3.2 Protein sidechains ⓘ

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	284/293 (97%)	278 (98%)	6 (2%)	61	76
1	C	282/293 (96%)	277 (98%)	5 (2%)	66	81
1	E	282/293 (96%)	280 (99%)	2 (1%)	88	95
2	B	112/141 (79%)	109 (97%)	3 (3%)	52	67
2	D	116/141 (82%)	112 (97%)	4 (3%)	44	57
2	F	117/141 (83%)	111 (95%)	6 (5%)	29	36
All	All	1193/1302 (92%)	1167 (98%)	26 (2%)	60	75

All (26) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	18	THR
1	A	23	ASN
1	A	40	LYS
1	A	45	ASP
1	A	95	PHE
1	A	236	ASN
2	B	61	THR
2	B	69	GLU
2	B	86	ASP
1	C	40	LYS
1	C	95	PHE
1	C	107	ARG
1	C	274	CYS
1	C	282	ILE
2	D	61	THR
2	D	72	ASN
2	D	84	MET
2	D	124	LEU
1	E	107	ARG
1	E	282	ILE
2	F	50	ASN
2	F	61	THR
2	F	75	ARG
2	F	112	ASP
2	F	126	LEU
2	F	151	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (13) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	23	ASN
1	A	279	GLN
2	B	50	ASN
2	B	125	GLN
1	C	2	GLN
1	C	240	ASN
2	D	28	ASN
2	D	50	ASN
2	D	72	ASN
1	E	2	GLN
1	E	23	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	E	169	GLN
2	F	50	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 ⓘ

8 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	1321	4	16,20,21	0.39	0	18,28,31	1.26	3 (16%)
4	GAL	A	1322	4	12,12,12	0.54	0	17,17,17	1.11	2 (11%)
6	SIA	C	1321	6	16,20,21	0.37	0	18,28,31	1.14	1 (5%)
6	GAL	C	1322	6	11,11,12	0.67	0	14,15,17	1.22	1 (7%)
6	NAG	C	1323	6	15,15,15	0.64	0	17,21,21	1.52	2 (11%)
6	SIA	E	1321	6	16,20,21	0.36	0	18,28,31	1.01	1 (5%)
6	GAL	E	1322	6	11,11,12	0.52	0	14,15,17	1.19	1 (7%)
6	NAG	E	1323	6	15,15,15	0.51	0	17,21,21	1.14	1 (5%)

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	1321	4	-	0/14/34/38	0/1/1/1
4	GAL	A	1322	4	-	0/2/22/22	0/1/1/1
6	SIA	C	1321	6	-	0/14/34/38	0/1/1/1
6	GAL	C	1322	6	-	0/2/19/22	0/1/1/1
6	NAG	C	1323	6	-	0/6/26/26	0/1/1/1
6	SIA	E	1321	6	-	0/14/34/38	0/1/1/1
6	GAL	E	1322	6	-	0/2/19/22	0/1/1/1
6	NAG	E	1323	6	-	0/6/26/26	0/1/1/1

There are no bond length outliers.

All (12) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	C	1321	SIA	C6-C5-N5	-3.51	104.95	111.07
4	A	1321	SIA	C6-C5-N5	-3.47	105.02	111.07
6	E	1321	SIA	C6-C5-N5	-2.58	106.58	111.07
4	A	1321	SIA	C7-C6-C5	-2.15	111.06	114.32
6	E	1322	GAL	C1-O5-C5	-2.05	109.64	112.25
4	A	1322	GAL	C1-C2-C3	2.11	113.56	110.43
4	A	1322	GAL	O5-C5-C6	2.32	112.22	106.36
6	E	1323	NAG	C1-O5-C5	2.53	118.14	113.47
4	A	1321	SIA	O6-C6-C5	2.59	112.73	108.48
6	C	1323	NAG	C8-C7-N2	2.91	121.67	116.11
6	C	1322	GAL	C1-C2-C3	3.79	114.02	109.54
6	C	1323	NAG	C2-N2-C7	4.03	133.46	123.10

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1321	SIA	1	0
6	C	1321	SIA	1	0
6	C	1322	GAL	1	0
6	C	1323	NAG	1	0

## 5.6 Ligand geometry ⓘ

8 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	1320	1	14,14,15	0.48	0	15,19,21	1.54	2 (13%)
5	PO4	A	1323	-	4,4,4	0.44	0	6,6,6	0.28	0
5	PO4	A	1324	-	4,4,4	0.39	0	6,6,6	0.28	0
5	PO4	A	1325	-	4,4,4	0.41	0	6,6,6	0.29	0
3	NAG	C	1320	1	14,14,15	0.43	0	15,19,21	1.26	1 (6%)
5	PO4	C	1324	-	4,4,4	0.39	0	6,6,6	0.27	0
3	NAG	E	1320	1	14,14,15	0.38	0	15,19,21	1.28	2 (13%)
5	PO4	E	1324	-	4,4,4	0.35	0	6,6,6	0.27	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	1320	1	-	0/6/23/26	0/1/1/1
5	PO4	A	1323	-	-	0/0/0/0	0/0/0/0
5	PO4	A	1324	-	-	0/0/0/0	0/0/0/0
5	PO4	A	1325	-	-	0/0/0/0	0/0/0/0
3	NAG	C	1320	1	-	0/6/23/26	0/1/1/1
5	PO4	C	1324	-	-	0/0/0/0	0/0/0/0
3	NAG	E	1320	1	-	0/6/23/26	0/1/1/1
5	PO4	E	1324	-	-	0/0/0/0	0/0/0/0

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	1320	NAG	C3-C2-N2	-2.43	104.74	110.56
3	E	1320	NAG	C3-C2-N2	-2.26	105.14	110.56
3	C	1320	NAG	C1-O5-C5	3.51	116.70	112.25
3	E	1320	NAG	C1-O5-C5	3.74	116.99	112.25
3	A	1320	NAG	C1-O5-C5	4.51	117.97	112.25

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	1325	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	319/327 (97%)	0.41	17 (5%)	30 45	22, 42, 105, 181	0
1	C	319/327 (97%)	0.36	16 (5%)	32 47	23, 41, 95, 174	0
1	E	320/327 (97%)	0.31	10 (3%)	52 65	19, 43, 86, 102	0
2	B	146/166 (87%)	2.76	73 (50%)	0 0	22, 118, 168, 186	0
2	D	148/166 (89%)	2.62	77 (52%)	0 0	24, 116, 172, 201	0
2	F	148/166 (89%)	1.40	37 (25%)	1 1	24, 89, 120, 127	0
All	All	1400/1479 (94%)	0.96	230 (16%)	2 4	19, 49, 164, 201	0

All (230) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	20	GLY	16.2
2	D	152	VAL	12.4
1	C	2	GLN	11.8
2	B	124	LEU	11.6
2	D	21	TRP	11.6
2	B	138	PHE	11.1
2	B	128	ASP	11.0
1	C	1	ASP	10.7
2	B	130	ALA	10.3
2	B	134	GLY	10.2
2	D	138	PHE	10.2
2	D	140	PHE	10.0
2	D	156	THR	9.7
2	D	33	GLY	8.9
1	A	2	GLN	8.8
1	A	11	ASN	8.7
2	F	21	TRP	8.3
2	B	26	HIS	8.2
2	B	121	LYS	8.1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	B	21	TRP	7.7
1	A	5	ILE	7.6
2	B	36	ALA	7.5
2	B	140	PHE	7.5
2	B	137	CYS	7.3
2	B	22	TYR	7.3
1	C	5	ILE	7.2
2	B	144	CYS	7.2
2	B	133	LEU	7.0
2	D	153	ARG	6.9
2	B	23	GLY	6.8
2	B	148	CYS	6.8
2	B	24	TYR	6.7
2	B	16	GLY	6.5
2	B	19	ASP	6.4
2	D	22	TYR	6.4
2	D	14	TRP	6.4
2	F	126	LEU	6.3
2	D	154	ASN	6.3
1	C	3	ILE	6.3
1	A	3	ILE	6.3
2	D	16	GLY	6.3
2	F	118	LEU	6.1
2	D	127	ARG	6.1
2	B	129	ASN	6.1
2	F	141	TYR	5.9
2	F	44	ALA	5.8
2	F	130	ALA	5.7
2	D	126	LEU	5.6
2	B	27	SER	5.6
2	B	17	MET	5.6
2	D	155	GLY	5.6
2	D	125	GLN	5.6
2	D	24	TYR	5.5
2	F	124	LEU	5.3
2	F	43	LYS	5.3
2	B	31	GLY	5.3
2	D	139	GLU	5.3
2	B	151	SER	5.3
2	D	131	LYS	5.2
2	D	124	LEU	5.1
2	B	25	HIS	5.0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	B	136	GLY	5.0
2	B	119	TYR	4.9
2	B	18	VAL	4.9
2	B	118	LEU	4.9
1	A	12	SER	4.8
1	C	4	CYS	4.8
2	F	122	VAL	4.8
2	B	127	ARG	4.7
2	D	27	SER	4.7
2	D	44	ALA	4.7
2	B	154	ASN	4.7
2	D	31	GLY	4.7
2	B	110	PHE	4.7
2	D	23	GLY	4.7
1	C	317	LEU	4.6
1	A	4	CYS	4.6
2	B	132	GLU	4.6
2	B	122	VAL	4.6
2	D	144	CYS	4.5
2	B	44	ALA	4.5
2	F	121	LYS	4.5
2	D	141	TYR	4.5
2	D	133	LEU	4.5
2	B	10	ILE	4.4
1	A	13	THR	4.4
2	D	118	LEU	4.4
2	D	43	LYS	4.3
2	B	141	TYR	4.3
2	B	20	GLY	4.3
2	F	48	VAL	4.3
2	F	40	SER	4.3
2	B	43	LYS	4.2
2	D	35	ALA	4.2
2	D	52	VAL	4.2
2	B	48	VAL	4.2
2	B	149	MET	4.2
2	D	17	MET	4.2
2	D	25	HIS	4.1
2	D	15	GLN	4.1
2	D	42	GLN	4.1
2	D	137	CYS	4.0
2	D	136	GLY	4.0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	B	150	GLU	4.0
2	F	132	GLU	4.0
2	B	46	ASP	4.0
2	D	26	HIS	4.0
2	B	142	HIS	4.0
2	B	38	LYS	3.9
1	A	7	TYR	3.9
2	D	130	ALA	3.9
2	B	52	VAL	3.9
2	B	45	ILE	3.9
2	D	34	TYR	3.9
2	F	19	ASP	3.8
2	B	139	GLU	3.8
2	F	53	ASN	3.8
1	E	5	ILE	3.8
2	B	33	GLY	3.8
2	B	117	ASN	3.8
1	C	7	TYR	3.8
2	F	12	GLY	3.7
2	B	125	GLN	3.7
2	B	59	MET	3.7
2	F	98	LEU	3.6
1	C	10	ASN	3.6
2	F	54	SER	3.6
2	D	19	ASP	3.6
2	D	132	GLU	3.5
2	F	46	ASP	3.5
2	D	61	THR	3.5
2	D	123	ARG	3.5
2	F	16	GLY	3.5
2	B	55	ILE	3.5
2	D	143	ARG	3.4
2	F	133	LEU	3.4
2	D	149	MET	3.4
2	B	147	GLU	3.4
1	A	1	ASP	3.4
2	D	48	VAL	3.3
1	C	11	ASN	3.3
2	F	36	ALA	3.3
2	D	157	TYR	3.3
1	C	28	HIS	3.3
1	A	39	GLY	3.3

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	B	113	SER	3.3
2	F	52	VAL	3.3
2	D	142	HIS	3.3
1	A	14	GLU	3.2
2	F	157	TYR	3.2
2	B	126	LEU	3.2
2	D	116	LYS	3.2
2	D	55	ILE	3.2
2	D	60	ASN	3.2
2	D	147	GLU	3.2
2	B	47	GLY	3.1
2	D	13	GLY	3.1
2	F	55	ILE	3.1
2	B	145	ASP	3.1
1	E	20	MET	3.0
2	D	117	ASN	3.0
2	D	102	MET	3.0
2	D	120	ASP	3.0
1	E	4	CYS	3.0
2	F	22	TYR	2.9
2	D	32	SER	2.9
2	D	134	GLY	2.9
1	E	3	ILE	2.9
1	E	13	THR	2.9
2	F	154	ASN	2.9
2	F	123	ARG	2.9
2	B	50	ASN	2.9
2	B	35	ALA	2.8
2	D	46	ASP	2.8
2	B	34	TYR	2.8
2	B	146	ASN	2.8
2	D	128	ASP	2.8
2	D	37	ASP	2.8
2	D	145	ASP	2.8
2	D	121	LYS	2.8
2	B	102	MET	2.7
1	E	19	ILE	2.7
2	D	45	ILE	2.7
1	E	291	PHE	2.7
2	D	119	TYR	2.7
2	D	151	SER	2.6
1	C	29	ALA	2.6

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	B	29	GLU	2.6
1	A	310	ARG	2.6
2	B	106	ARG	2.6
2	B	114	ASN	2.6
1	A	6	GLY	2.6
2	D	39	GLU	2.5
2	B	131	LYS	2.5
2	F	117	ASN	2.5
2	B	135	ASN	2.5
2	F	50	ASN	2.5
2	B	120	ASP	2.4
1	A	28	HIS	2.4
2	B	14	TRP	2.4
1	A	311	LEU	2.3
1	E	7	TYR	2.3
2	F	136	GLY	2.3
2	F	102	MET	2.3
2	D	38	LYS	2.3
2	F	116	LYS	2.3
2	D	40	SER	2.3
2	B	41	THR	2.2
1	C	9	ALA	2.2
2	B	15	GLN	2.2
2	D	41	THR	2.2
2	D	36	ALA	2.2
1	A	21	GLU	2.2
2	D	10	ILE	2.2
1	C	26	VAL	2.2
1	C	20	MET	2.2
1	C	13	THR	2.2
2	B	13	GLY	2.1
2	D	111	HIS	2.1
2	D	11	GLU	2.1
2	F	59	MET	2.1
2	F	38	LYS	2.1
1	E	11	ASN	2.1
1	C	21	GLU	2.1
1	E	2	GLN	2.1
2	F	134	GLY	2.1
2	D	115	VAL	2.1
1	A	312	VAL	2.0
2	D	30	GLN	2.0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	F	106	ARG	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	A	1321	20/21	0.88	0.17	1.14	51,58,63,63	0
6	SIA	C	1321	20/21	0.96	0.15	0.37	34,36,42,44	0
6	SIA	E	1321	20/21	0.98	0.13	-0.57	24,28,31,32	0
4	GAL	A	1322	12/12	0.88	0.34	-	66,75,77,78	0
6	GAL	E	1322	11/12	0.89	0.11	-	33,48,53,55	0
6	NAG	E	1323	15/15	0.82	0.28	-	55,66,69,70	0
6	NAG	C	1323	15/15	0.83	0.25	-	68,75,76,76	0
6	GAL	C	1322	11/12	0.84	0.16	-	44,56,61,62	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(Å <sup>2</sup> )	Q<0.9
5	PO4	E	1324	5/5	0.93	0.33	10.67	67,70,72,73	0
5	PO4	A	1323	5/5	0.94	0.26	6.55	71,73,75,76	0
5	PO4	A	1325	5/5	0.95	0.26	3.66	65,67,70,70	0
5	PO4	C	1324	5/5	0.95	0.19	2.51	65,66,69,72	0
5	PO4	A	1324	5/5	0.91	0.20	1.08	60,61,66,68	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
3	NAG	E	1320	14/15	0.91	0.11	-1.08	39,42,51,52	0
3	NAG	C	1320	14/15	0.93	0.11	-4.30	37,41,44,48	0
3	NAG	A	1320	14/15	0.94	0.10	-4.48	31,35,41,43	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.