



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 1, 2016 – 06:25 PM GMT

PDB ID : 4LKG  
Title : The structure of hemagglutinin from a avian-origin H7N9 influenza virus (A/Shanghai/1/2013) in complex with avian receptor analog 3'SLNLN  
Authors : Shi, Y.; Zhang, W.; Wang, F.; Qi, J.; Song, H.; Wu, Y.; Gao, F.; Zhang, Y.; Fan, Z.; Gong, W.; Wang, D.; Shu, Y.; Wang, Y.; Yan, J.; Gao, G.F.  
Deposited on : 2013-07-07  
Resolution : 2.99 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

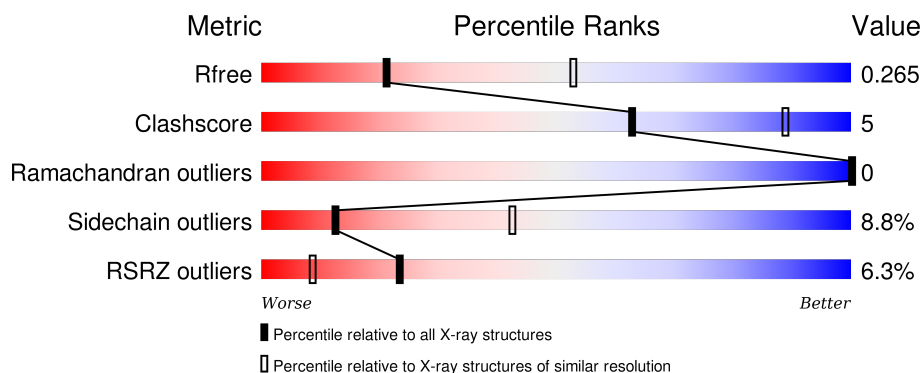
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.99 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	1578 (3.00-3.00)
Clashscore	102246	1912 (3.00-3.00)
Ramachandran outliers	100387	1853 (3.00-3.00)
Sidechain outliers	100360	1856 (3.00-3.00)
RSRZ outliers	91569	1592 (3.00-3.00)

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	316	<div> <div>80%</div> <div>19%</div> <div>.</div> </div>
1	C	316	<div> <div>3%</div> <div>79%</div> <div>20%</div> <div>.</div> </div>
1	E	316	<div> <div>6%</div> <div>79%</div> <div>20%</div> <div>.</div> </div>
2	B	170	<div> <div>8%</div> <div>85%</div> <div>14%</div> <div>.</div> </div>
2	D	170	<div> <div>8%</div> <div>85%</div> <div>15%</div> <div>.</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	F	170	

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	803	-	-	X	-

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 11531 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	316	Total	C	N	O	S	0	0	0
			2415	1497	435	468	15			
1	C	316	Total	C	N	O	S	0	0	0
			2415	1497	435	468	15			
1	E	316	Total	C	N	O	S	0	0	0
			2415	1497	435	468	15			

- 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
			1378	851	240	280	7			
2	D	170	Total	C	N	O	S	0	0	0
			1378	851	240	280	7			
2	F	170	Total	C	N	O	S	0	0	0
			1378	851	240	280	7			

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

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

- Molecule 4 is SUGAR (N-ACETYL-D-GLUCOSAMINE) (three-letter code: NAG) (formula:  $C_8H_{15}NO_6$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	B	1	Total	C	N	O	0	0
			14	8	1	5		
4	D	1	Total	C	N	O	0	0
			14	8	1	5		
4	F	1	Total	C	N	O	0	0
			14	8	1	5		

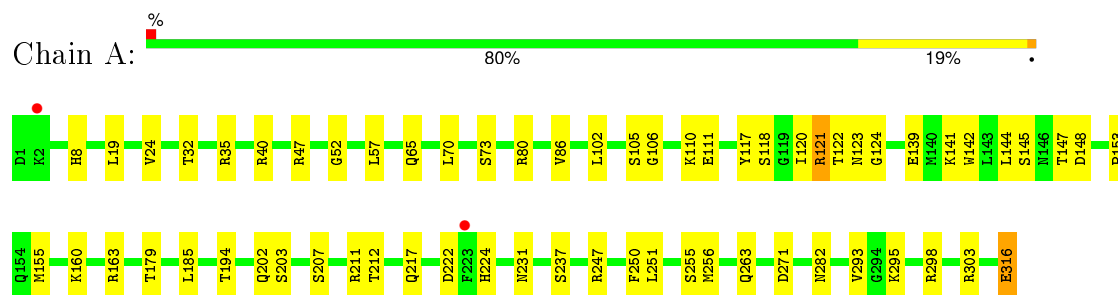
- Molecule 5 is a polymer of unknown type called SUGAR (2-MER).

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

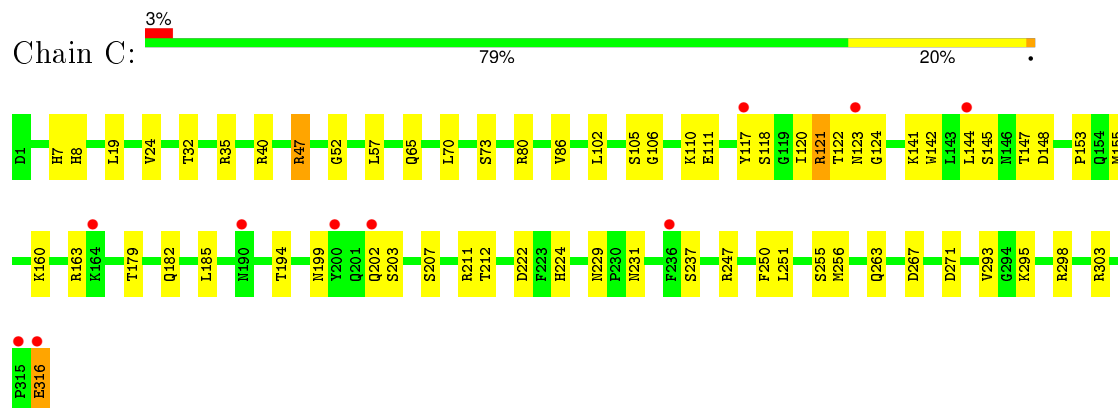
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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 ( $\text{RSRZ} > 2$ ). 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.

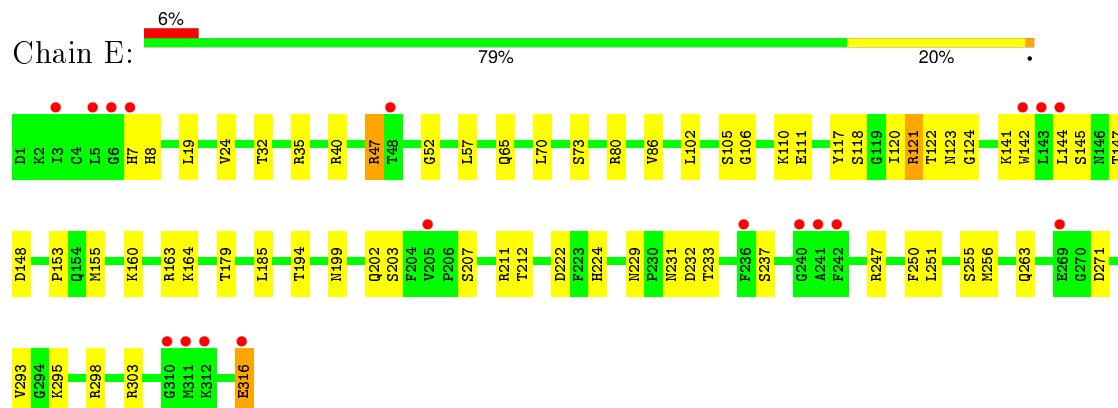
- Molecule 1: hemagglutinin



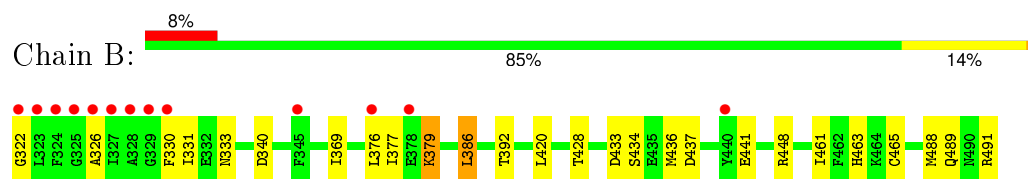
- Molecule 1: hemagglutinin



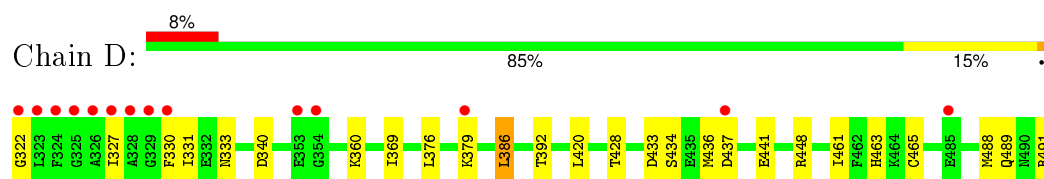
- Molecule 1: hemagglutinin



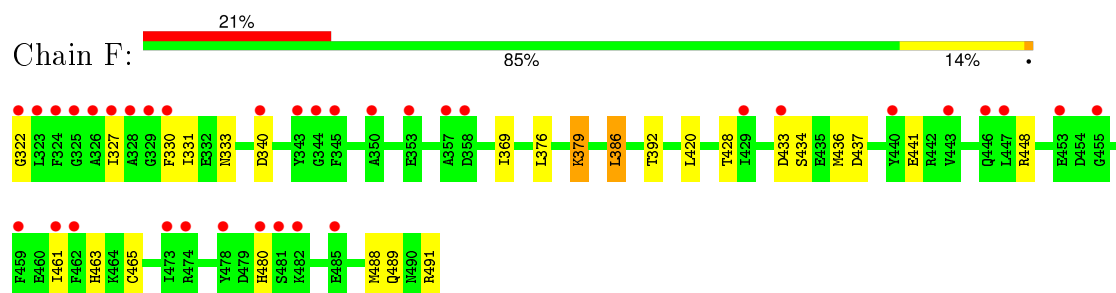
## ● Molecule 2: hemagglutinin



## ● Molecule 2: hemagglutinin



## ● Molecule 2: hemagglutinin



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 3	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	110.76Å 110.76Å 133.23Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	42.59 – 2.99 42.59 – 2.99	Depositor EDS
% Data completeness (in resolution range)	99.7 (42.59-2.99) 99.7 (42.59-2.99)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.21 (at 3.01Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.3_928)	Depositor
R, $R_{free}$	0.246 , 0.278 0.233 , 0.265	Depositor DCC
$R_{free}$ test set	1836 reflections (5.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	72.7	Xtriage
Anisotropy	0.168	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.23 , 4.2	EDS
Estimated twinning fraction	0.029 for -h,-k,l 0.109 for h,-h-k,-l 0.037 for -k,-h,-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.45$ , $\langle L^2 \rangle = 0.27$	Xtriage
Outliers	0 of 36748 reflections	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	11531	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	100.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 15.49% 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: 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	0.27	0/2460	0.46	0/3323
1	C	0.26	0/2460	0.46	0/3323
1	E	0.26	0/2460	0.46	0/3323
2	B	0.25	0/1402	0.41	0/1889
2	D	0.25	0/1402	0.42	0/1889
2	F	0.28	0/1402	0.43	0/1889
All	All	0.26	0/11586	0.45	0/15636

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	2415	0	2366	25	1
1	C	2415	0	2366	35	0
1	E	2415	0	2366	26	1
2	B	1378	0	1278	12	0
2	D	1378	0	1278	13	0
2	F	1378	0	1278	13	0
3	A	46	0	40	14	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	B	14	0	13	0	0
4	D	14	0	13	0	0
4	F	14	0	13	0	0
5	C	32	0	28	1	0
5	E	32	0	28	1	0
All	All	11531	0	11067	119	1

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 5.

All (119) 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)
3:A:803:NAG:H82	1:C:267:ASP:HB3	1.32	1.12
3:A:803:NAG:C8	1:C:267:ASP:HB3	1.96	0.96
1:A:217:GLN:HE22	3:A:802:GAL:H3	1.33	0.93
1:C:40:ARG:HG2	1:C:263:GLN:HG3	1.53	0.88
3:A:803:NAG:C7	1:C:267:ASP:HB3	2.08	0.84
3:A:803:NAG:H82	1:C:267:ASP:CB	2.13	0.77
2:F:448:ARG:HD3	2:F:480:HIS:CD2	2.26	0.70
1:E:144:LEU:HD11	1:E:185:LEU:HD22	1.75	0.69
1:A:121:ARG:NH1	1:A:145:SER:O	2.26	0.68
1:A:144:LEU:HD11	1:A:185:LEU:HD22	1.76	0.68
1:E:121:ARG:HB3	1:E:144:LEU:HB3	1.75	0.68
1:A:121:ARG:HB3	1:A:144:LEU:HB3	1.74	0.68
1:C:121:ARG:HB3	1:C:144:LEU:HB3	1.74	0.68
1:E:121:ARG:NH1	1:E:145:SER:O	2.26	0.68
3:A:803:NAG:H82	1:C:267:ASP:C	2.15	0.67
1:C:144:LEU:HD11	1:C:185:LEU:HD22	1.76	0.66
1:C:121:ARG:NH1	1:C:145:SER:O	2.27	0.66
1:E:70:LEU:O	1:E:110:LYS:NZ	2.30	0.64
1:C:70:LEU:O	1:C:110:LYS:NZ	2.31	0.64
1:A:70:LEU:O	1:A:110:LYS:NZ	2.32	0.63
5:C:801:SIA:O1B	5:C:801:SIA:H6	1.99	0.63
1:E:316:GLU:HA	2:F:333:ASN:HD22	1.65	0.61
3:A:802:GAL:O2	3:A:803:NAG:H62	2.02	0.59
2:F:448:ARG:HD3	2:F:480:HIS:NE2	2.17	0.59
1:C:316:GLU:HA	2:D:333:ASN:HD22	1.68	0.59
1:A:316:GLU:HA	2:B:333:ASN:HD22	1.68	0.58
5:E:801:SIA:H6	5:E:801:SIA:O1B	2.03	0.58
1:A:222:ASP:HB3	1:A:224:HIS:CE1	2.40	0.57

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:E:222:ASP:HB3	1:E:224:HIS:CE1	2.40	0.57
1:C:222:ASP:HB3	1:C:224:HIS:CE1	2.40	0.56
1:A:40:ARG:HG2	1:A:263:GLN:HG3	1.87	0.56
1:E:40:ARG:HG2	1:E:263:GLN:HG3	1.87	0.56
1:C:111:GLU:OE2	1:C:163:ARG:NH1	2.40	0.55
1:C:160:LYS:HE3	1:C:231:ASN:HA	1.89	0.55
1:E:111:GLU:OE2	1:E:163:ARG:NH1	2.38	0.55
1:A:111:GLU:OE2	1:A:163:ARG:NH1	2.39	0.54
3:A:803:NAG:N2	1:C:267:ASP:HB3	2.21	0.54
3:A:803:NAG:H82	1:C:267:ASP:O	2.09	0.53
1:E:160:LYS:HE3	1:E:231:ASN:HA	1.90	0.53
1:A:160:LYS:HE3	1:A:231:ASN:HA	1.90	0.53
2:D:437:ASP:O	2:D:441:GLU:HG2	2.10	0.52
2:F:437:ASP:O	2:F:441:GLU:HG2	2.10	0.52
1:C:207:SER:O	1:C:211:ARG:NH2	2.42	0.51
2:B:437:ASP:O	2:B:441:GLU:HG2	2.09	0.51
3:A:801:SIA:O1B	3:A:801:SIA:H6	2.09	0.51
1:E:293:VAL:HG11	2:F:386:LEU:HD13	1.93	0.50
2:F:461:ILE:HG22	2:F:463:HIS:H	1.77	0.49
1:C:293:VAL:HG11	2:D:386:LEU:HD13	1.95	0.49
2:F:376:LEU:HD22	2:F:420:LEU:HD21	1.95	0.48
1:A:217:GLN:HE22	3:A:802:GAL:C3	2.14	0.48
1:A:207:SER:O	1:A:211:ARG:NH2	2.45	0.48
2:D:376:LEU:HD22	2:D:420:LEU:HD21	1.96	0.48
1:A:153:PRO:O	1:A:155:MET:HG3	2.14	0.48
2:B:376:LEU:HD22	2:B:420:LEU:HD21	1.95	0.48
1:C:153:PRO:O	1:C:155:MET:HG3	2.14	0.48
1:A:47:ARG:NH1	1:A:73:SER:HB3	2.29	0.47
1:C:163:ARG:HD3	1:C:250:PHE:CE2	2.50	0.47
2:D:461:ILE:HG22	2:D:463:HIS:H	1.78	0.47
1:E:124:GLY:HA3	1:E:142:TRP:HB3	1.97	0.47
2:B:461:ILE:HG22	2:B:463:HIS:H	1.78	0.47
1:E:207:SER:O	1:E:211:ARG:NH2	2.44	0.47
1:E:153:PRO:O	1:E:155:MET:HG3	2.15	0.46
1:E:35:ARG:NH1	1:E:303:ARG:O	2.47	0.46
1:A:163:ARG:HD3	1:A:250:PHE:CE2	2.50	0.46
1:A:35:ARG:NH1	1:A:303:ARG:O	2.48	0.46
1:E:106:GLY:HA2	1:E:255:SER:HB3	1.98	0.46
1:E:163:ARG:HD3	1:E:250:PHE:CE2	2.50	0.46
2:D:327:ILE:N	2:D:433:ASP:OD1	2.42	0.46
1:E:47:ARG:NH1	1:E:73:SER:HB3	2.30	0.45

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:106:GLY:HA2	1:C:255:SER:HB3	1.99	0.45
2:F:327:ILE:N	2:F:433:ASP:OD1	2.42	0.45
1:A:106:GLY:HA2	1:A:255:SER:HB3	1.98	0.45
1:C:35:ARG:NH1	1:C:303:ARG:O	2.49	0.45
1:E:7:HIS:CD2	2:F:327:ILE:HG12	2.51	0.45
1:E:163:ARG:HD3	1:E:250:PHE:CZ	2.52	0.45
2:D:330:PHE:CD1	2:D:331:ILE:HG13	2.52	0.45
2:D:448:ARG:HE	2:D:448:ARG:HB3	1.61	0.45
1:C:47:ARG:NH1	1:C:73:SER:HB3	2.32	0.45
2:F:330:PHE:CD1	2:F:331:ILE:HG13	2.52	0.45
1:A:293:VAL:HG11	2:B:386:LEU:HD13	1.97	0.45
1:C:117:TYR:HB3	1:C:120:ILE:HD11	1.99	0.44
3:A:803:NAG:HN2	1:C:267:ASP:CB	2.30	0.44
1:A:124:GLY:HA3	1:A:142:TRP:HB3	1.99	0.44
1:C:124:GLY:HA3	1:C:142:TRP:HB3	1.98	0.44
2:B:322:GLY:HA3	2:B:433:ASP:OD2	2.18	0.44
2:D:322:GLY:HA3	2:D:433:ASP:OD2	2.18	0.43
1:A:105:SER:HB2	1:A:251:LEU:HD22	2.00	0.43
1:A:117:TYR:HB3	1:A:120:ILE:HD11	2.00	0.43
2:F:379:LYS:HB2	2:F:379:LYS:HE3	1.82	0.43
1:A:163:ARG:HD3	1:A:250:PHE:CZ	2.53	0.43
2:B:330:PHE:CD1	2:B:331:ILE:HG13	2.52	0.43
1:E:117:TYR:HB3	1:E:120:ILE:HD11	2.00	0.43
1:C:105:SER:HB2	1:C:251:LEU:HD22	2.01	0.43
3:A:803:NAG:HN2	1:C:267:ASP:HB3	1.84	0.43
2:D:369:ILE:HD11	2:D:428:THR:HG23	2.01	0.43
2:B:448:ARG:HB3	2:B:448:ARG:HE	1.61	0.42
2:F:322:GLY:HA3	2:F:433:ASP:OD2	2.19	0.42
1:E:52:GLY:HA2	1:E:80:ARG:HG3	2.00	0.42
1:C:52:GLY:HA2	1:C:80:ARG:HG3	2.01	0.42
1:C:163:ARG:HD3	1:C:250:PHE:CZ	2.53	0.42
1:E:199:ASN:OD1	1:E:229:ASN:ND2	2.51	0.42
1:E:105:SER:HB2	1:E:251:LEU:HD22	2.02	0.42
1:A:282:ASN:HB3	2:B:377:ILE:HG23	2.02	0.42
2:F:369:ILE:HD11	2:F:428:THR:HG23	2.02	0.41
1:C:199:ASN:OD1	1:C:229:ASN:ND2	2.49	0.41
1:C:57:LEU:HD23	1:C:102:LEU:HD12	2.03	0.41
3:A:803:NAG:H82	1:C:267:ASP:CA	2.51	0.41
2:B:369:ILE:HD11	2:B:428:THR:HG23	2.01	0.41
1:C:35:ARG:HD3	1:C:303:ARG:HG2	2.03	0.41
1:A:52:GLY:HA2	1:A:80:ARG:HG3	2.02	0.41

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:E:232:ASP:OD1	1:E:233:THR:N	2.50	0.41
2:B:379:LYS:HE3	2:B:379:LYS:HB2	1.82	0.41
1:E:35:ARG:HD3	1:E:303:ARG:HG2	2.02	0.41
1:E:57:LEU:HD23	1:E:102:LEU:HD12	2.03	0.41
1:A:57:LEU:HD23	1:A:102:LEU:HD12	2.03	0.41
2:D:360:LYS:HB2	2:D:360:LYS:HE3	1.86	0.41
2:D:331:ILE:HG13	2:D:331:ILE:H	1.74	0.40
2:B:326:ALA:HB3	2:B:433:ASP:OD1	2.22	0.40
1:C:7:HIS:CD2	2:D:327:ILE:HG12	2.57	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:139:GLU:OE1	1:E:164:LYS:NZ[2_545]	2.11	0.09

## 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	314/316 (99%)	292 (93%)	22 (7%)	0	100	100
1	C	314/316 (99%)	291 (93%)	23 (7%)	0	100	100
1	E	314/316 (99%)	291 (93%)	23 (7%)	0	100	100
2	B	168/170 (99%)	159 (95%)	9 (5%)	0	100	100
2	D	168/170 (99%)	159 (95%)	9 (5%)	0	100	100
2	F	168/170 (99%)	159 (95%)	9 (5%)	0	100	100
All	All	1446/1458 (99%)	1351 (93%)	95 (7%)	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	265/265 (100%)	240 (91%)	25 (9%)	11	39
1	C	265/265 (100%)	238 (90%)	27 (10%)	9	33
1	E	265/265 (100%)	239 (90%)	26 (10%)	10	36
2	B	145/145 (100%)	135 (93%)	10 (7%)	19	56
2	D	145/145 (100%)	135 (93%)	10 (7%)	19	56
2	F	145/145 (100%)	135 (93%)	10 (7%)	19	56
All	All	1230/1230 (100%)	1122 (91%)	108 (9%)	12	42

All (108) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	8	HIS
1	A	19	LEU
1	A	24	VAL
1	A	32	THR
1	A	65	GLN
1	A	86	VAL
1	A	118	SER
1	A	121	ARG
1	A	122	THR
1	A	123	ASN
1	A	141	LYS
1	A	147	THR
1	A	148	ASP
1	A	179	THR
1	A	194	THR
1	A	202	GLN
1	A	203	SER
1	A	212	THR
1	A	237	SER
1	A	247	ARG
1	A	256	MET
1	A	271	ASP

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	295	LYS
1	A	298	ARG
1	A	316	GLU
2	B	340	ASP
2	B	379	LYS
2	B	386	LEU
2	B	392	THR
2	B	434	SER
2	B	436	MET
2	B	465	CYS
2	B	488	MET
2	B	489	GLN
2	B	491	ARG
1	C	8	HIS
1	C	19	LEU
1	C	24	VAL
1	C	32	THR
1	C	47	ARG
1	C	65	GLN
1	C	86	VAL
1	C	118	SER
1	C	121	ARG
1	C	122	THR
1	C	123	ASN
1	C	141	LYS
1	C	147	THR
1	C	148	ASP
1	C	179	THR
1	C	182	GLN
1	C	194	THR
1	C	202	GLN
1	C	203	SER
1	C	212	THR
1	C	237	SER
1	C	247	ARG
1	C	256	MET
1	C	271	ASP
1	C	295	LYS
1	C	298	ARG
1	C	316	GLU
2	D	340	ASP
2	D	379	LYS

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
2	D	386	LEU
2	D	392	THR
2	D	434	SER
2	D	436	MET
2	D	465	CYS
2	D	488	MET
2	D	489	GLN
2	D	491	ARG
1	E	8	HIS
1	E	19	LEU
1	E	24	VAL
1	E	32	THR
1	E	47	ARG
1	E	65	GLN
1	E	86	VAL
1	E	118	SER
1	E	121	ARG
1	E	122	THR
1	E	123	ASN
1	E	141	LYS
1	E	147	THR
1	E	148	ASP
1	E	179	THR
1	E	194	THR
1	E	202	GLN
1	E	203	SER
1	E	212	THR
1	E	237	SER
1	E	247	ARG
1	E	256	MET
1	E	271	ASP
1	E	295	LYS
1	E	298	ARG
1	E	316	GLU
2	F	340	ASP
2	F	379	LYS
2	F	386	LEU
2	F	392	THR
2	F	434	SER
2	F	436	MET
2	F	465	CYS
2	F	488	MET

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
2	F	489	GLN
2	F	491	ARG

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	217	GLN
1	A	224	HIS
2	B	383	GLN
1	C	224	HIS
1	E	224	HIS

### 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 ⓘ

7 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	SIA	A	801	3	16,20,21	0.25	0	18,28,31	0.68	1 (5%)
3	GAL	A	802	3	11,11,12	0.63	0	14,15,17	1.02	0
3	NAG	A	803	3	15,15,15	0.50	0	17,21,21	1.34	1 (5%)
5	SIA	C	801	5	16,20,21	0.26	0	18,28,31	0.68	1 (5%)
5	GAL	C	802	5	12,12,12	0.48	0	17,17,17	0.53	0
5	SIA	E	801	5	16,20,21	0.25	0	18,28,31	0.68	1 (5%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
5	GAL	E	802	5	12,12,12	0.43	0	17,17,17	0.51	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	SIA	A	801	3	-	0/14/34/38	0/1/1/1
3	GAL	A	802	3	-	0/2/19/22	0/1/1/1
3	NAG	A	803	3	-	0/6/26/26	0/1/1/1
5	SIA	C	801	5	-	0/14/34/38	0/1/1/1
5	GAL	C	802	5	-	0/2/22/22	0/1/1/1
5	SIA	E	801	5	-	0/14/34/38	0/1/1/1
5	GAL	E	802	5	-	0/2/22/22	0/1/1/1

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	801	SIA	C7-C6-C5	-2.19	111.00	114.32
5	C	801	SIA	C7-C6-C5	-2.19	111.00	114.32
5	E	801	SIA	C7-C6-C5	-2.17	111.03	114.32
3	A	803	NAG	C4-C3-C2	4.21	116.26	110.43

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5 monomers are involved in 16 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	801	SIA	1	0
3	A	802	GAL	3	0
3	A	803	NAG	11	0
5	C	801	SIA	1	0
5	E	801	SIA	1	0

## 5.6 Ligand geometry

3 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	$\# Z  > 2$
4	NAG	B	601	2	14,14,15	0.48	0	15,19,21	1.04	1 (6%)
4	NAG	D	601	2	14,14,15	0.51	0	15,19,21	0.63	0
4	NAG	F	601	2	14,14,15	0.46	0	15,19,21	0.92	1 (6%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	NAG	B	601	2	-	0/6/23/26	0/1/1/1
4	NAG	D	601	2	-	0/6/23/26	0/1/1/1
4	NAG	F	601	2	-	0/6/23/26	0/1/1/1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
4	F	601	NAG	C1-O5-C5	2.64	115.59	112.25
4	B	601	NAG	C1-O5-C5	3.35	116.50	112.25

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 ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues

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	316/316 (100%)	-0.04	2 (0%) 90 73	43, 73, 111, 140	0
1	C	316/316 (100%)	0.15	10 (3%) 51 23	46, 91, 145, 232	0
1	E	316/316 (100%)	0.33	18 (5%) 27 10	59, 101, 169, 236	0
2	B	170/170 (100%)	0.32	13 (7%) 17 6	39, 80, 148, 277	0
2	D	170/170 (100%)	0.33	14 (8%) 14 5	41, 99, 156, 229	0
2	F	170/170 (100%)	1.34	35 (20%) 1 1	55, 158, 221, 306	0
All	All	1458/1458 (100%)	0.33	92 (6%) 23 9	39, 91, 179, 306	0

All (92) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	327	ILE	19.6
2	F	326	ALA	17.5
2	F	322	GLY	12.7
2	B	322	GLY	12.2
2	B	323	LEU	11.9
2	F	330	PHE	10.0
2	F	328	ALA	9.6
2	B	324	PHE	9.3
2	F	329	GLY	8.1
2	F	323	LEU	7.8
1	E	5	LEU	7.0
2	F	482	LYS	6.9
2	B	328	ALA	6.9
2	B	330	PHE	6.5
2	F	343	TYR	6.2
1	E	143	LEU	6.0
2	B	326	ALA	5.8
2	F	481	SER	5.8
2	D	322	GLY	5.8

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	F	459	PHE	5.7
2	F	325	GLY	5.6
1	C	316	GLU	5.6
1	E	310	GLY	5.6
1	E	142	TRP	5.5
2	D	327	ILE	5.3
2	D	324	PHE	5.0
2	F	353	GLU	5.0
2	D	323	LEU	5.0
2	F	485	GLU	5.0
2	F	344	GLY	4.9
2	F	324	PHE	4.8
1	C	190	ASN	4.7
2	D	328	ALA	4.7
2	F	443	VAL	4.6
1	E	240	GLY	4.6
2	D	353	GLU	4.3
2	F	461	ILE	4.2
2	F	447	LEU	4.1
1	C	164	LYS	3.9
1	E	3	ILE	3.7
2	F	473	ILE	3.7
1	C	200	TYR	3.6
2	F	357	ALA	3.6
2	B	327	ILE	3.5
1	E	316	GLU	3.5
2	D	379	LYS	3.5
2	B	325	GLY	3.5
2	D	325	GLY	3.4
2	F	440	TYR	3.4
1	E	242	PHE	3.4
1	C	236	PHE	3.2
1	E	6	GLY	3.1
2	F	455	GLY	3.1
2	F	480	HIS	3.1
2	D	329	GLY	3.1
1	E	7	HIS	3.1
1	E	311	MET	3.0
2	B	329	GLY	3.0
1	A	2	LYS	2.9
1	C	202	GLN	2.9
2	F	345	PHE	2.9

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	F	478	TYR	2.9
2	D	326	ALA	2.8
2	D	354	GLY	2.7
1	E	269	GLU	2.6
2	F	358	ASP	2.6
2	F	350	ALA	2.6
2	D	485	GLU	2.6
1	E	312	LYS	2.6
2	D	330	PHE	2.5
1	E	241	ALA	2.5
1	E	236	PHE	2.5
1	E	205	VAL	2.5
2	B	345	PHE	2.5
2	B	376	LEU	2.4
1	C	315	PRO	2.4
2	F	340	ASP	2.4
1	A	223	PHE	2.4
1	E	144	LEU	2.3
2	D	437	ASP	2.3
2	F	446	GLN	2.3
1	C	117	TYR	2.3
2	F	474	ARG	2.3
2	F	433	ASP	2.2
2	B	378	GLU	2.2
2	B	440	TYR	2.2
2	F	453	GLU	2.2
1	E	48	THR	2.2
2	F	462	PHE	2.2
1	C	123	ASN	2.0
2	F	429	ILE	2.0
1	C	144	LEU	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
3	NAG	A	803	15/15	0.75	0.24	1.30	133,154,172,175	0
5	SIA	E	801	20/21	0.90	0.26	0.37	130,136,145,147	0
3	GAL	A	802	11/12	0.93	0.21	0.36	84,89,95,106	0
3	SIA	A	801	20/21	0.91	0.19	0.05	90,106,115,115	0
5	SIA	C	801	20/21	0.91	0.15	-0.43	92,102,113,115	0
5	GAL	C	802	12/12	0.81	0.20	-	125,148,178,196	0
5	GAL	E	802	12/12	0.76	0.25	-	123,142,151,156	0

## 6.4 Ligands

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	NAG	D	601	14/15	0.92	0.13	-0.83	83,90,107,114	0
4	NAG	F	601	14/15	0.88	0.19	-	96,120,140,140	0
4	NAG	B	601	14/15	0.86	0.22	-	70,84,88,89	0

## 6.5 Other polymers

There are no such residues in this entry.