



# Full wwPDB X-ray Structure Validation Report

Feb 27, 2014 – 03:57 PM GMT

PDB ID : 4F23  
Title : Influenza A virus hemagglutinin H16 HA0 structure with an alpha-helix con-  
formation in the cleavage site: a potential drug target  
Authors : Lu, X.; Shi, Y.; Gao, F.; Xiao, H.; Qi, J.; Gao, G.F.  
Deposited on : 2012-05-07  
Resolution : 1.70 Å(reported)

This is a full wwPDB 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/ValidationPDFNotes.html>

---

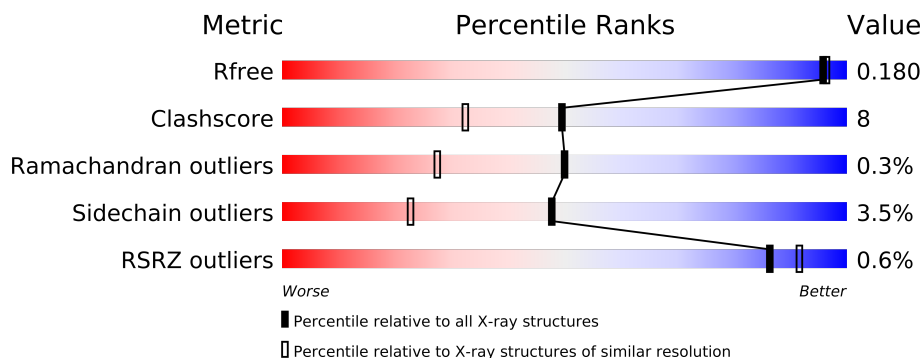
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.15 2013  
Xtriage (Phenix) : dev-1323  
EDS : stable22639  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable22683

# 1 Overall quality at a glance

The reported resolution of this entry is 1.70 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))
$R_{free}$	66092	2456 (1.70-1.70)
Clashscore	79885	2929 (1.70-1.70)
Ramachandran outliers	78287	2878 (1.70-1.70)
Sidechain outliers	78261	2878 (1.70-1.70)
RSRZ outliers	66119	2456 (1.70-1.70)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	515	
1	B	515	
1	C	515	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
3	NAG	A	603	-	X
3	NAG	B	603	-	X
3	NAG	C	603	-	X

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 14190 atoms, of which 0 are hydrogen and 0 are deuterium.

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	509	Total	C	N	O	S	0	3	0
			4040	2522	719	781	18			
1	B	509	Total	C	N	O	S	0	1	0
			4025	2513	716	778	18			
1	C	509	Total	C	N	O	S	0	0	0
			4019	2510	715	776	18			

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-5	HIS	-	EXPRESSION TAG	UNP Q5DL24
A	-4	HIS	-	EXPRESSION TAG	UNP Q5DL24
A	-3	HIS	-	EXPRESSION TAG	UNP Q5DL24
A	-2	HIS	-	EXPRESSION TAG	UNP Q5DL24
A	-1	HIS	-	EXPRESSION TAG	UNP Q5DL24
A	0	HIS	-	EXPRESSION TAG	UNP Q5DL24
A	1	ALA	-	SEE REMARK 999	UNP Q5DL24
A	2	ASP	-	SEE REMARK 999	UNP Q5DL24
A	3	GLY	-	SEE REMARK 999	UNP Q5DL24
A	4	ILE	-	SEE REMARK 999	UNP Q5DL24
A	5	GLN	-	SEE REMARK 999	UNP Q5DL24
A	505	ARG	-	SEE REMARK 999	UNP Q5DL24
A	506	LEU	-	SEE REMARK 999	UNP Q5DL24
A	507	VAL	-	SEE REMARK 999	UNP Q5DL24
A	508	PRO	-	SEE REMARK 999	UNP Q5DL24
A	509	ARG	-	SEE REMARK 999	UNP Q5DL24
B	-5	HIS	-	EXPRESSION TAG	UNP Q5DL24
B	-4	HIS	-	EXPRESSION TAG	UNP Q5DL24
B	-3	HIS	-	EXPRESSION TAG	UNP Q5DL24
B	-2	HIS	-	EXPRESSION TAG	UNP Q5DL24
B	-1	HIS	-	EXPRESSION TAG	UNP Q5DL24
B	0	HIS	-	EXPRESSION TAG	UNP Q5DL24
B	1	ALA	-	SEE REMARK 999	UNP Q5DL24

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	2	ASP	-	SEE REMARK 999	UNP Q5DL24
B	3	GLY	-	SEE REMARK 999	UNP Q5DL24
B	4	ILE	-	SEE REMARK 999	UNP Q5DL24
B	5	GLN	-	SEE REMARK 999	UNP Q5DL24
B	505	ARG	-	SEE REMARK 999	UNP Q5DL24
B	506	LEU	-	SEE REMARK 999	UNP Q5DL24
B	507	VAL	-	SEE REMARK 999	UNP Q5DL24
B	508	PRO	-	SEE REMARK 999	UNP Q5DL24
B	509	ARG	-	SEE REMARK 999	UNP Q5DL24
C	-5	HIS	-	EXPRESSION TAG	UNP Q5DL24
C	-4	HIS	-	EXPRESSION TAG	UNP Q5DL24
C	-3	HIS	-	EXPRESSION TAG	UNP Q5DL24
C	-2	HIS	-	EXPRESSION TAG	UNP Q5DL24
C	-1	HIS	-	EXPRESSION TAG	UNP Q5DL24
C	0	HIS	-	EXPRESSION TAG	UNP Q5DL24
C	1	ALA	-	SEE REMARK 999	UNP Q5DL24
C	2	ASP	-	SEE REMARK 999	UNP Q5DL24
C	3	GLY	-	SEE REMARK 999	UNP Q5DL24
C	4	ILE	-	SEE REMARK 999	UNP Q5DL24
C	5	GLN	-	SEE REMARK 999	UNP Q5DL24
C	505	ARG	-	SEE REMARK 999	UNP Q5DL24
C	506	LEU	-	SEE REMARK 999	UNP Q5DL24
C	507	VAL	-	SEE REMARK 999	UNP Q5DL24
C	508	PRO	-	SEE REMARK 999	UNP Q5DL24
C	509	ARG	-	SEE REMARK 999	UNP Q5DL24

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

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

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-5	HIS	-	EXPRESSION TAG	UNP Q5DL24
A	-4	HIS	-	EXPRESSION TAG	UNP Q5DL24
A	-3	HIS	-	EXPRESSION TAG	UNP Q5DL24

*Continued on next page...*

*Continued from previous page...*

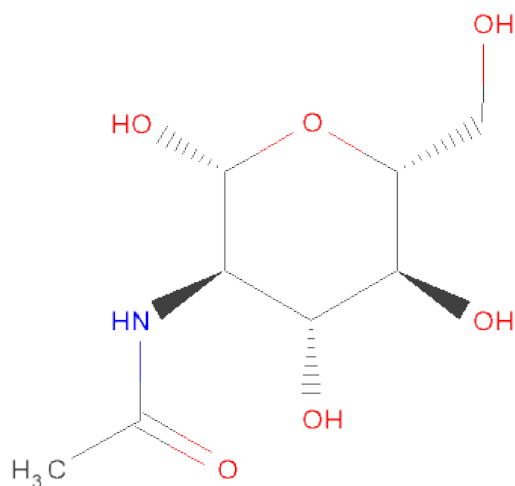
Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	HIS	-	EXPRESSION TAG	UNP Q5DL24
A	-1	HIS	-	EXPRESSION TAG	UNP Q5DL24
A	0	HIS	-	EXPRESSION TAG	UNP Q5DL24
A	1	ALA	-	SEE REMARK 999	UNP Q5DL24
A	2	ASP	-	SEE REMARK 999	UNP Q5DL24
A	3	GLY	-	SEE REMARK 999	UNP Q5DL24
A	4	ILE	-	SEE REMARK 999	UNP Q5DL24
A	5	GLN	-	SEE REMARK 999	UNP Q5DL24
A	505	ARG	-	SEE REMARK 999	UNP Q5DL24
A	506	LEU	-	SEE REMARK 999	UNP Q5DL24
A	507	VAL	-	SEE REMARK 999	UNP Q5DL24
A	508	PRO	-	SEE REMARK 999	UNP Q5DL24
A	509	ARG	-	SEE REMARK 999	UNP Q5DL24
B	-5	HIS	-	EXPRESSION TAG	UNP Q5DL24
B	-4	HIS	-	EXPRESSION TAG	UNP Q5DL24
B	-3	HIS	-	EXPRESSION TAG	UNP Q5DL24
B	-2	HIS	-	EXPRESSION TAG	UNP Q5DL24
B	-1	HIS	-	EXPRESSION TAG	UNP Q5DL24
B	0	HIS	-	EXPRESSION TAG	UNP Q5DL24
B	1	ALA	-	SEE REMARK 999	UNP Q5DL24
B	2	ASP	-	SEE REMARK 999	UNP Q5DL24
B	3	GLY	-	SEE REMARK 999	UNP Q5DL24
B	4	ILE	-	SEE REMARK 999	UNP Q5DL24
B	5	GLN	-	SEE REMARK 999	UNP Q5DL24
B	505	ARG	-	SEE REMARK 999	UNP Q5DL24
B	506	LEU	-	SEE REMARK 999	UNP Q5DL24
B	507	VAL	-	SEE REMARK 999	UNP Q5DL24
B	508	PRO	-	SEE REMARK 999	UNP Q5DL24
B	509	ARG	-	SEE REMARK 999	UNP Q5DL24
C	-5	HIS	-	EXPRESSION TAG	UNP Q5DL24
C	-4	HIS	-	EXPRESSION TAG	UNP Q5DL24
C	-3	HIS	-	EXPRESSION TAG	UNP Q5DL24
C	-2	HIS	-	EXPRESSION TAG	UNP Q5DL24
C	-1	HIS	-	EXPRESSION TAG	UNP Q5DL24
C	0	HIS	-	EXPRESSION TAG	UNP Q5DL24
C	1	ALA	-	SEE REMARK 999	UNP Q5DL24
C	2	ASP	-	SEE REMARK 999	UNP Q5DL24
C	3	GLY	-	SEE REMARK 999	UNP Q5DL24
C	4	ILE	-	SEE REMARK 999	UNP Q5DL24
C	5	GLN	-	SEE REMARK 999	UNP Q5DL24
C	505	ARG	-	SEE REMARK 999	UNP Q5DL24
C	506	LEU	-	SEE REMARK 999	UNP Q5DL24

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	507	VAL	-	SEE REMARK 999	UNP Q5DL24
C	508	PRO	-	SEE REMARK 999	UNP Q5DL24
C	509	ARG	-	SEE REMARK 999	UNP Q5DL24

- 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	B	1	Total	C	N	O	0	0
			14	8	1	5		
3	C	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 4 is water.

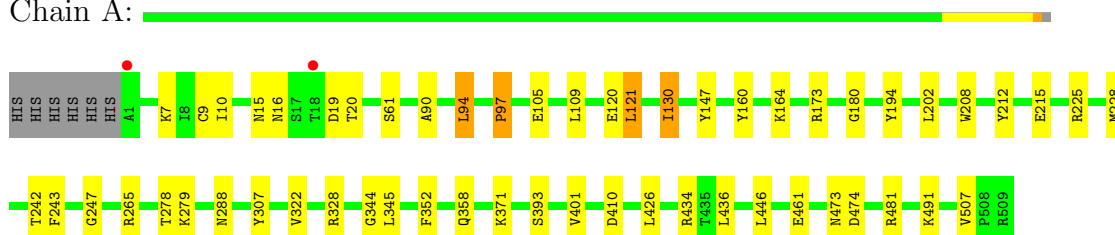
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	676	Total	O	0	0
			676	676		
4	B	650	Total	O	0	0
			650	650		
4	C	654	Total	O	0	0
			654	654		

### 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.

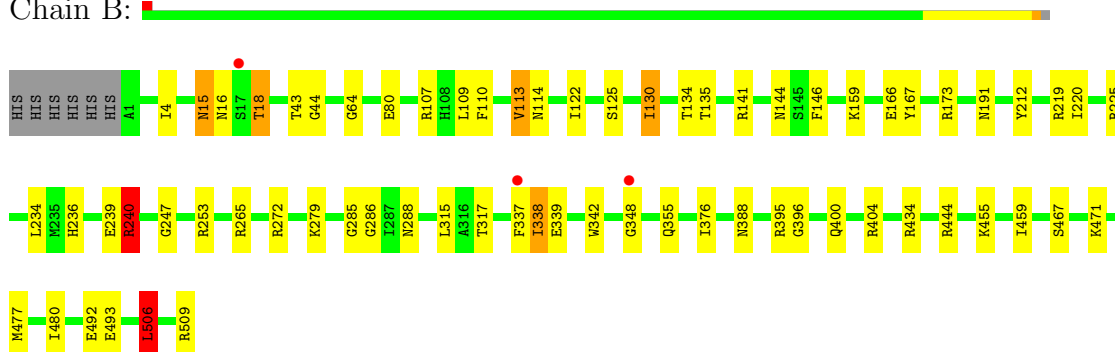
#### • Molecule 1: Hemagglutinin

Chain A:



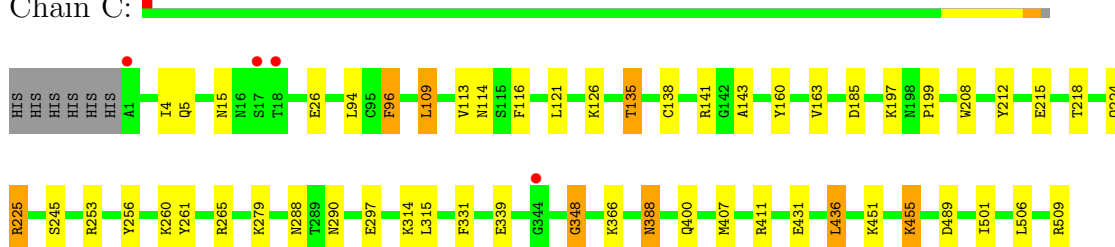
#### • Molecule 1: Hemagglutinin

Chain B:



#### • Molecule 1: Hemagglutinin

Chain C:



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	69.33Å 240.94Å 69.67Å 90.00° 119.79° 90.00°	Depositor
Resolution (Å)	25.00 – 1.70 24.15 – 1.71	Depositor EDS
% Data completeness (in resolution range)	97.5 (25.00-1.70) 97.6 (24.15-1.71)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.09 (at 1.71Å)	Xtriage
Refinement program	REFMAC 5.5.0102	Depositor
R, $R_{free}$	0.160 , 0.192 0.154 , 0.180	Depositor DCC
$R_{free}$ test set	10596 reflections (5.30%)	DCC
Wilson B-factor (Å <sup>2</sup> )	16.2	Xtriage
Anisotropy	0.759	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 34.8	EDS
Estimated twinning fraction	0.659 for H, K, L 0.209 for -H-L, K, H 0.132 for L, K, -H-L 0.217 for -h-l,k,h 0.217 for l,k,-h-l 0.097 for h,-k,-h-l 0.095 for -h-l,-k,l 0.090 for l,-k,h	Xtriage
Reported twinning fraction	0.659 for H, K, L 0.209 for -H-L, K, H 0.132 for L, K, -H-L	Depositor
L-test for twinning	$\langle  L  \rangle = 0.41$ , $\langle L^2 \rangle = 0.23$	Xtriage
Outliers	0 of 210913 reflections	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	14190	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	21.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.91% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.



## 5 Model quality ⓘ

### 5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section:  
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	1.27	11/4123 (0.3%)	1.03	12/5583 (0.2%)
1	B	1.28	11/4108 (0.3%)	1.04	13/5565 (0.2%)
1	C	1.30	15/4102 (0.4%)	1.00	6/5556 (0.1%)
All	All	1.29	37/12333 (0.3%)	1.02	31/16704 (0.2%)

All (37) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	431	GLU	CB-CG	8.63	1.68	1.52
1	C	348	GLY	N-CA	-7.75	1.34	1.46
1	C	256	TYR	CE1-CZ	-7.22	1.29	1.38
1	A	208	TRP	CE3-CZ3	7.21	1.50	1.38
1	B	493	GLU	CB-CG	6.67	1.64	1.52
1	A	194	TYR	CG-CD2	6.58	1.47	1.39
1	C	143	ALA	CA-CB	-6.36	1.39	1.52
1	B	146	PHE	CE1-CZ	6.10	1.49	1.37
1	B	492	GLU	CD-OE2	6.03	1.32	1.25
1	B	212	TYR	CE1-CZ	6.03	1.46	1.38
1	C	431	GLU	CD-OE2	5.93	1.32	1.25
1	C	138	CYS	CB-SG	5.91	1.92	1.82
1	B	396	GLY	C-O	5.89	1.33	1.23
1	C	208	TRP	CB-CG	5.86	1.60	1.50
1	B	247	GLY	N-CA	5.85	1.54	1.46
1	C	96	PHE	CD2-CE2	-5.80	1.27	1.39
1	B	166	GLU	CB-CG	-5.79	1.41	1.52
1	B	239	GLU	CD-OE1	-5.71	1.19	1.25
1	C	116	PHE	CE1-CZ	5.63	1.48	1.37
1	A	247	GLY	N-CA	5.60	1.54	1.46
1	A	212	TYR	CE2-CZ	-5.58	1.31	1.38
1	B	219	ARG	CZ-NH2	5.43	1.40	1.33
1	A	105	GLU	CD-OE2	-5.42	1.19	1.25

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	331	PHE	CD2-CE2	5.37	1.50	1.39
1	C	297	GLU	CB-CG	5.34	1.62	1.52
1	A	322	VAL	CB-CG2	5.34	1.64	1.52
1	C	215	GLU	CD-OE1	-5.28	1.19	1.25
1	A	461	GLU	CD-OE1	5.25	1.31	1.25
1	A	243	PHE	CG-CD1	5.23	1.46	1.38
1	A	120	GLU	CD-OE2	-5.20	1.20	1.25
1	C	212	TYR	CE2-CZ	-5.18	1.31	1.38
1	A	180	GLY	N-CA	5.12	1.53	1.46
1	A	160	TYR	CE1-CZ	5.08	1.45	1.38
1	B	113	VAL	CB-CG2	-5.07	1.42	1.52
1	C	212	TYR	CE1-CZ	5.05	1.45	1.38
1	B	110	PHE	CG-CD1	5.05	1.46	1.38
1	C	509	ARG	CZ-NH2	5.00	1.39	1.33

All (31) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	509	ARG	NE-CZ-NH1	10.96	125.78	120.30
1	A	265	ARG	NE-CZ-NH2	-9.77	115.41	120.30
1	A	265	ARG	NE-CZ-NH1	9.27	124.94	120.30
1	C	265	ARG	NE-CZ-NH2	-9.08	115.76	120.30
1	B	509	ARG	NE-CZ-NH2	-8.48	116.06	120.30
1	B	240	ARG	NE-CZ-NH2	-8.40	116.10	120.30
1	B	240	ARG	NE-CZ-NH1	7.99	124.29	120.30
1	A	94	LEU	CB-CG-CD1	7.92	124.46	111.00
1	B	130	ILE	CG1-CB-CG2	7.88	128.75	111.40
1	A	481	ARG	NE-CZ-NH2	-7.85	116.38	120.30
1	C	225	ARG	NE-CZ-NH1	7.82	124.21	120.30
1	C	265	ARG	NE-CZ-NH1	7.64	124.12	120.30
1	A	225	ARG	NE-CZ-NH2	-7.41	116.60	120.30
1	C	411	ARG	NE-CZ-NH1	-7.38	116.61	120.30
1	B	434	ARG	NE-CZ-NH1	7.34	123.97	120.30
1	C	436	LEU	CB-CG-CD1	6.76	122.49	111.00
1	B	506	LEU	CB-CG-CD2	6.55	122.13	111.00
1	B	404	ARG	NE-CZ-NH2	-6.19	117.21	120.30
1	A	121	LEU	CB-CG-CD1	6.18	121.51	111.00
1	A	481	ARG	NE-CZ-NH1	6.17	123.39	120.30
1	B	225	ARG	NE-CZ-NH1	6.02	123.31	120.30
1	A	147	TYR	CB-CG-CD2	-5.99	117.41	121.00
1	B	225	ARG	NE-CZ-NH2	-5.91	117.34	120.30
1	C	109	LEU	CB-CG-CD1	5.35	120.10	111.00

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	328	ARG	NE-CZ-NH2	-5.31	117.65	120.30
1	B	395	ARG	NE-CZ-NH1	5.29	122.94	120.30
1	A	410	ASP	CB-CG-OD1	5.27	123.04	118.30
1	A	446	LEU	CB-CG-CD1	-5.25	102.08	111.00
1	A	307	TYR	CB-CG-CD2	-5.16	117.91	121.00
1	B	107	ARG	NE-CZ-NH2	-5.16	117.72	120.30
1	B	167	TYR	CB-CG-CD2	5.13	124.08	121.00

There are no chirality outliers.

There are no planarity outliers.

## 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	4040	0	0	24	0
1	B	4025	0	0	39	0
1	C	4019	0	0	31	0
2	A	28	0	0	0	0
2	B	28	0	0	0	0
2	C	28	0	0	0	0
3	A	14	0	0	0	0
3	B	14	0	0	0	0
3	C	14	0	0	0	0
4	A	676	0	0	23	0
4	B	650	0	0	35	0
4	C	654	0	0	27	0
All	All	14190	0	0	93	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 8.

All (93) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:C:135:THR:OG1	4:C:1268:HOH:O	1.85	0.93

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:B:480:ILE:CD1	4:B:1261:HOH:O	2.22	0.88
1:B:317:THR:O	4:B:1101:HOH:O	1.94	0.85
1:A:352:PHE:CD1	4:A:1175:HOH:O	2.31	0.83
1:C:260:LYS:O	4:C:1278:HOH:O	1.97	0.82
1:B:286:GLY:N	4:B:1062:HOH:O	2.10	0.82
1:B:477:MET:SD	4:B:1261:HOH:O	2.40	0.79
1:A:90:ALA:CB	4:A:1086:HOH:O	2.31	0.79
1:B:467:SER:CB	4:B:872:HOH:O	2.31	0.77
1:C:489:ASP:OD2	4:C:1299:HOH:O	2.06	0.73
1:C:245:SER:O	4:C:1056:HOH:O	2.06	0.73
1:C:5:GLN:NE2	4:C:1338:HOH:O	2.20	0.73
1:C:160:TYR:CE2	4:C:1056:HOH:O	2.40	0.73
1:A:9:CYS:O	4:A:1175:HOH:O	2.06	0.71
1:C:218:THR:CG2	1:C:225:ARG:NH2	2.56	0.68
1:B:459:ILE:CB	4:B:872:HOH:O	2.40	0.68
1:B:342:TRP:NE1	4:B:1151:HOH:O	2.29	0.66
1:C:15:ASN:OD1	4:C:995:HOH:O	2.14	0.65
1:A:202:LEU:CB	4:A:1211:HOH:O	2.43	0.65
1:B:122:ILE:O	4:B:1271:HOH:O	2.15	0.64
1:B:285:GLY:CA	4:B:1062:HOH:O	2.47	0.62
1:C:224:GLN:NE2	4:C:1268:HOH:O	2.35	0.60
1:A:344:GLY:O	4:A:1371:HOH:O	2.16	0.59
1:B:43:THR:OG1	4:B:1062:HOH:O	2.16	0.59
1:B:337:PHE:CE2	1:B:338:ILE:O	2.57	0.58
1:B:125:SER:OG	4:B:1099:HOH:O	2.16	0.58
1:B:144:ASN:ND2	4:B:992:HOH:O	2.37	0.58
1:B:64:GLY:CA	4:B:1113:HOH:O	2.51	0.58
1:C:185:ASP:OD1	1:C:225:ARG:NH1	2.37	0.57
1:B:144:ASN:CB	4:B:992:HOH:O	2.52	0.57
1:B:355:GLN:NE2	4:B:1034:HOH:O	2.36	0.57
1:A:434:ARG:NH2	4:A:781:HOH:O	2.39	0.56
1:C:501:ILE:CD1	4:C:1338:HOH:O	2.53	0.56
1:B:467:SER:O	4:B:872:HOH:O	2.17	0.56
1:B:265:ARG:NH2	4:B:777:HOH:O	2.39	0.55
1:B:141:ARG:NE	4:B:1349:HOH:O	2.39	0.55
1:A:345:LEU:N	4:A:921:HOH:O	2.39	0.54
1:A:242:THR:OG1	4:A:972:HOH:O	2.19	0.54
1:B:80:GLU:CA	4:B:1330:HOH:O	2.56	0.54
1:C:113:VAL:CA	4:C:1278:HOH:O	2.55	0.53
1:B:467:SER:CA	4:B:872:HOH:O	2.56	0.52
4:A:972:HOH:O	1:B:220:ILE:CG1	2.57	0.52
1:A:228:MET:CB	4:A:1292:HOH:O	2.59	0.51

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:215:GLU:CD	4:A:1100:HOH:O	2.49	0.51
1:A:61:SER:N	4:A:1086:HOH:O	2.42	0.51
1:B:114:ASN:C	4:B:1330:HOH:O	2.49	0.50
1:C:114:ASN:ND2	4:C:1323:HOH:O	2.44	0.50
1:C:96:PHE:CE1	4:C:1268:HOH:O	2.64	0.50
1:B:444:ARG:NH1	4:B:1177:HOH:O	2.44	0.50
1:B:253:ARG:CG	4:B:1271:HOH:O	2.60	0.49
1:C:163:VAL:N	4:C:1056:HOH:O	2.45	0.48
1:B:113:VAL:O	4:B:1330:HOH:O	2.20	0.48
1:B:44:GLY:N	4:B:1062:HOH:O	2.47	0.48
1:B:43:THR:N	4:B:1062:HOH:O	2.46	0.47
1:A:10:ILE:CD1	4:A:1175:HOH:O	2.62	0.47
1:A:16:ASN:ND2	4:A:1273:HOH:O	2.46	0.47
1:C:290:ASN:ND2	4:C:1075:HOH:O	2.48	0.47
1:A:7:LYS:NZ	4:A:1280:HOH:O	2.48	0.47
1:C:366:LYS:NZ	4:C:989:HOH:O	2.48	0.47
1:C:451:LYS:NZ	4:C:1210:HOH:O	2.47	0.47
1:C:199:PRO:CA	4:C:1295:HOH:O	2.64	0.46
1:A:279:LYS:NZ	4:A:1294:HOH:O	2.48	0.46
1:C:126:LYS:NZ	4:C:891:HOH:O	2.48	0.45
1:A:491:LYS:NZ	4:A:871:HOH:O	2.49	0.45
1:B:376:ILE:CG1	4:B:1101:HOH:O	2.64	0.45
1:A:473:ASN:ND2	4:A:906:HOH:O	2.49	0.44
1:A:15:ASN:OD1	1:A:16:ASN:N	2.50	0.44
4:B:1051:HOH:O	1:C:407:MET:SD	2.62	0.44
1:C:253:ARG:CG	4:C:1076:HOH:O	2.66	0.44
1:A:97:PRO:O	4:A:1292:HOH:O	2.21	0.43
1:B:240:ARG:NH2	4:B:1204:HOH:O	2.52	0.43
1:B:15:ASN:ND2	1:B:16:ASN:N	2.66	0.43
1:B:191:ASN:OD1	4:B:1209:HOH:O	2.21	0.43
1:A:130:ILE:CA	4:A:1017:HOH:O	2.67	0.43
1:B:459:ILE:N	4:B:872:HOH:O	2.52	0.43
1:C:314:LYS:NZ	4:C:1245:HOH:O	2.52	0.43
1:B:159:LYS:NZ	4:B:763:HOH:O	2.52	0.43
1:C:4:ILE:N	4:C:951:HOH:O	2.51	0.42
1:B:18:THR:N	4:B:925:HOH:O	2.51	0.42
1:C:388:ASN:ND2	4:C:1225:HOH:O	2.51	0.42
1:B:471:LYS:NZ	4:B:977:HOH:O	2.52	0.42
4:A:920:HOH:O	1:B:506:LEU:CD1	2.68	0.42
1:C:455:LYS:NZ	4:C:880:HOH:O	2.53	0.41
1:C:261:TYR:CA	4:C:1278:HOH:O	2.67	0.41
1:A:401:VAL:CG1	1:A:401:VAL:O	2.68	0.41

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:B:236:HIS:NE2	1:C:400:GLN:OE1	2.54	0.41
1:C:96:PHE:CD1	4:C:1268:HOH:O	2.57	0.41
1:B:134:THR:C	4:B:992:HOH:O	2.59	0.41
1:A:164:LYS:NZ	4:A:1296:HOH:O	2.53	0.41
1:A:358:GLN:OE1	1:A:474:ASP:N	2.54	0.41
1:C:163:VAL:CG2	4:C:1056:HOH:O	2.69	0.40
1:A:15:ASN:ND2	4:A:1177:HOH:O	2.54	0.40
1:C:225:ARG:NE	4:C:977:HOH:O	2.53	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	510/515 (99%)	495 (97%)	14 (3%)	1 (0%)	56	33
1	B	508/515 (99%)	488 (96%)	17 (3%)	3 (1%)	33	13
1	C	507/515 (98%)	488 (96%)	18 (4%)	1 (0%)	56	33
All	All	1525/1545 (99%)	1471 (96%)	49 (3%)	5 (0%)	50	27

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	19	ASP
1	B	339	GLU
1	C	348	GLY
1	B	348	GLY
1	B	338	ILE

### 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	442/445 (99%)	428 (97%)	14 (3%)	51	26
1	B	440/445 (99%)	423 (96%)	17 (4%)	43	18
1	C	439/445 (99%)	424 (97%)	15 (3%)	49	23
All	All	1321/1335 (99%)	1275 (96%)	46 (4%)	48	23

All (46) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	20	THR
1	A	94	LEU
1	A	97	PRO
1	A	109	LEU
1	A	121	LEU
1	A	130	ILE
1	A	173	ARG
1	A	278	THR
1	A	288	ASN
1	A	371	LYS
1	A	393	SER
1	A	426	LEU
1	A	436	LEU
1	A	507	VAL
1	B	4	ILE
1	B	15	ASN
1	B	18	THR
1	B	109	LEU
1	B	130	ILE
1	B	135	THR
1	B	173	ARG
1	B	234	LEU
1	B	240	ARG
1	B	272	ARG
1	B	279	LYS
1	B	288	ASN
1	B	315	LEU
1	B	388	ASN
1	B	400	GLN
1	B	455	LYS
1	B	506	LEU
1	C	26	GLU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	C	94	LEU
1	C	109	LEU
1	C	121	LEU
1	C	135	THR
1	C	141	ARG
1	C	197	LYS
1	C	279	LYS
1	C	288	ASN
1	C	315	LEU
1	C	339	GLU
1	C	388	ASN
1	C	436	LEU
1	C	455	LYS
1	C	506	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

There are no RNA chains 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 ⓘ

6 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
2	NAG	A	601	1,2	12,14,15	0.96	1 (8%)	15,19,21	0.84	1 (6%)



Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	NAG	A	602	2	12,14,15	0.99	1 (8%)	15,19,21	1.25	3 (20%)
2	NAG	B	601	1,2	12,14,15	0.80	1 (8%)	15,19,21	1.28	3 (20%)
2	NAG	B	602	2	12,14,15	0.70	0	15,19,21	1.43	2 (13%)
2	NAG	C	601	1,2	12,14,15	1.17	2 (16%)	15,19,21	1.70	1 (6%)
2	NAG	C	602	2	12,14,15	0.72	0	15,19,21	1.94	3 (20%)

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
2	NAG	A	601	1,2	-	0/6/23/26	0/1/1/1
2	NAG	A	602	2	-	0/6/23/26	0/1/1/1
2	NAG	B	601	1,2	-	0/6/23/26	0/1/1/1
2	NAG	B	602	2	-	0/6/23/26	0/1/1/1
2	NAG	C	601	1,2	-	0/6/23/26	0/1/1/1
2	NAG	C	602	2	-	0/6/23/26	0/1/1/1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	601	NAG	O5-C5	-2.42	1.40	1.45
2	C	601	NAG	C2-N2	2.23	1.48	1.46
2	B	601	NAG	O5-C5	-2.23	1.41	1.45
2	C	601	NAG	O5-C5	-2.16	1.41	1.45
2	A	602	NAG	O4-C4	2.09	1.48	1.43

All (13) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	601	NAG	O5-C5-C4	-5.58	103.57	110.65
2	C	602	NAG	O5-C5-C6	5.27	112.52	106.98
2	B	602	NAG	C3-C2-N2	-3.47	106.48	111.76
2	B	602	NAG	C2-N2-C7	3.17	128.41	123.09
2	A	602	NAG	C3-C4-C5	-2.80	105.20	110.20
2	C	602	NAG	C3-C2-N2	-2.46	108.01	111.76
2	C	602	NAG	O5-C5-C4	-2.41	107.59	110.65
2	B	601	NAG	O6-C6-C5	-2.40	103.11	111.36
2	A	602	NAG	C2-N2-C7	-2.27	119.28	123.09
2	B	601	NAG	O7-C7-C8	-2.18	117.79	122.04

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	601	NAG	C8-C7-N2	2.12	120.25	116.11
2	A	601	NAG	C3-C4-C5	-2.07	106.51	110.20
2	A	602	NAG	C6-C5-C4	-2.03	108.08	113.00

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 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
3	NAG	A	603	1	12,14,15	1.08	1 (8%)	15,19,21	1.39	2 (13%)
3	NAG	B	603	1	12,14,15	0.84	1 (8%)	15,19,21	0.96	1 (6%)
3	NAG	C	603	1	12,14,15	1.01	1 (8%)	15,19,21	1.14	2 (13%)

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	603	1	-	0/6/23/26	0/1/1/1
3	NAG	B	603	1	-	0/6/23/26	0/1/1/1
3	NAG	C	603	1	-	0/6/23/26	0/1/1/1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	603	NAG	O5-C5	-2.52	1.40	1.45

Continued on next page...

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	603	NAG	O5-C5	-2.49	1.40	1.45
3	B	603	NAG	O5-C5	-2.05	1.41	1.45

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	603	NAG	C3-C2-N2	-3.17	106.93	111.76
3	A	603	NAG	O3-C3-C4	2.44	115.83	110.35
3	A	603	NAG	O5-C5-C4	-2.44	107.56	110.65
3	B	603	NAG	C2-N2-C7	-2.13	119.51	123.09
3	C	603	NAG	O5-C5-C4	-2.05	108.05	110.65

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 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	509/515 (98%)	-0.09	2 (0%) 90 94	9, 18, 28, 44	0
1	B	509/515 (98%)	-0.10	3 (0%) 86 91	9, 19, 30, 45	0
1	C	509/515 (98%)	-0.11	4 (0%) 83 88	9, 19, 29, 42	0
All	All	1527/1545 (98%)	-0.10	9 (0%) 86 91	9, 19, 29, 45	0

All (9) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1	ALA	5.4
1	A	18	THR	5.2
1	C	1	ALA	3.8
1	B	17	SER	3.1
1	C	17	SER	3.1
1	C	18	THR	2.9
1	C	344	GLY	2.8
1	B	337	PHE	2.6
1	B	348	GLY	2.1

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

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

### 6.3 Carbohydrates ⓘ

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	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	NAG	A	602	14/15	0.19	5.85	27,31,37,39	0
2	NAG	C	601	14/15	0.13	0.85	22,32,35,35	0
2	NAG	A	601	14/15	0.08	-1.38	19,22,27,31	0
2	NAG	B	601	14/15	0.09	-1.43	26,29,33,36	0
2	NAG	B	602	14/15	0.14	-	34,37,40,40	0
2	NAG	C	602	14/15	0.15	-	31,35,41,44	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	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	NAG	B	603	14/15	0.15	7.00	30,34,36,38	0
3	NAG	C	603	14/15	0.18	3.20	32,36,39,41	0
3	NAG	A	603	14/15	0.14	2.57	26,29,33,37	0

## 6.5 Other polymers

There are no such residues in this entry.