



# Full wwPDB X-ray Structure Validation Report

Feb 27, 2014 – 02:48 PM GMT

PDB ID : 4JF7  
Title : Structure of the parainfluenza virus 5 (PIV5) hemagglutinin-neuraminidase(HN) ectodomain  
Authors : Welch, B.D.; Yuan, P.; Bose, S.; Kors, C.A.; Lamb, R.A.; Jardetzky, T.S.  
Deposited on : 2013-02-27  
Resolution : 2.50 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>

---

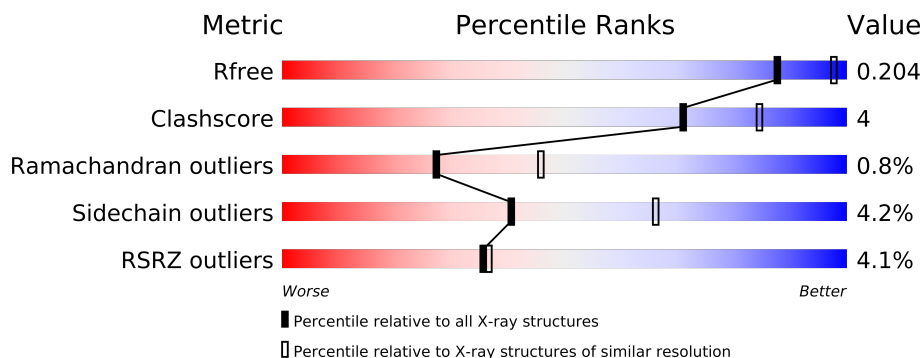
The following versions of software and data (see [references](#)) were used in the production of this report:

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 2.50 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	2784 (2.50-2.50)
Clashscore	79885	3562 (2.50-2.50)
Ramachandran outliers	78287	3480 (2.50-2.50)
Sidechain outliers	78261	3482 (2.50-2.50)
RSRZ outliers	66119	2785 (2.50-2.50)

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	513	
1	B	513	
1	C	513	
1	D	513	

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

Mol	Type	Chain	Res	Geometry	Electron density
2	CA	B	601	-	X
6	SO4	C	602	-	X

## 2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 16400 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-neuraminidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	D	492	Total	C	N	O	S	0	0	0
			3816	2431	630	728	27			
1	A	501	Total	C	N	O	S	0	0	0
			3888	2477	642	741	28			
1	B	501	Total	C	N	O	S	0	0	0
			3888	2474	643	743	28			
1	C	493	Total	C	N	O	S	0	0	0
			3822	2436	631	727	28			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	53	SER	-	EXPRESSION TAG	UNP P04850
D	54	PRO	-	EXPRESSION TAG	UNP P04850
D	55	SER	ASN	CONFLICT	UNP P04850
A	53	SER	-	EXPRESSION TAG	UNP P04850
A	54	PRO	-	EXPRESSION TAG	UNP P04850
A	55	SER	ASN	CONFLICT	UNP P04850
B	53	SER	-	EXPRESSION TAG	UNP P04850
B	54	PRO	-	EXPRESSION TAG	UNP P04850
B	55	SER	ASN	CONFLICT	UNP P04850
C	53	SER	-	EXPRESSION TAG	UNP P04850
C	54	PRO	-	EXPRESSION TAG	UNP P04850
C	55	SER	ASN	CONFLICT	UNP P04850

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total	Ca	0	0
			1	1		
2	A	1	Total	Ca	0	0
			1	1		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	D	1	Total	Ca	0	0
			1	1		
2	C	1	Total	Ca	0	0
			1	1		

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

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	D	3	Total	C	N	O	0	0
			39	22	2	15		

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	53	SER	-	EXPRESSION TAG	UNP P04850
D	54	PRO	-	EXPRESSION TAG	UNP P04850
D	55	SER	ASN	CONFLICT	UNP P04850

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

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

There are 12 discrepancies between the modelled and reference sequences:

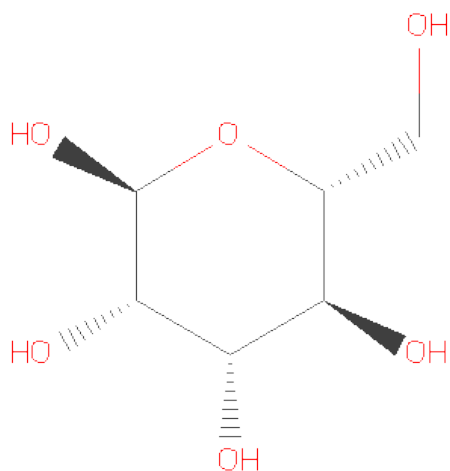
Chain	Residue	Modelled	Actual	Comment	Reference
D	53	SER	-	EXPRESSION TAG	UNP P04850
D	54	PRO	-	EXPRESSION TAG	UNP P04850
D	55	SER	ASN	CONFLICT	UNP P04850
A	53	SER	-	EXPRESSION TAG	UNP P04850
A	54	PRO	-	EXPRESSION TAG	UNP P04850
A	55	SER	ASN	CONFLICT	UNP P04850
A	53	SER	-	EXPRESSION TAG	UNP P04850
A	54	PRO	-	EXPRESSION TAG	UNP P04850
A	55	SER	ASN	CONFLICT	UNP P04850

*Continued on next page...*

Continued from previous page...

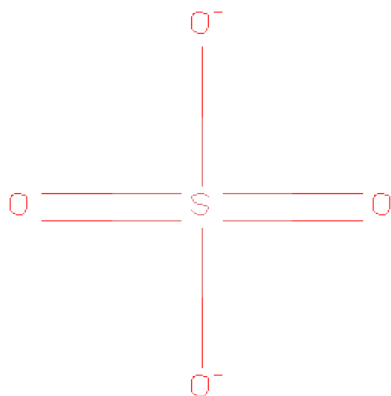
Chain	Residue	Modelled	Actual	Comment	Reference
B	53	SER	-	EXPRESSION TAG	UNP P04850
B	54	PRO	-	EXPRESSION TAG	UNP P04850
B	55	SER	ASN	CONFLICT	UNP P04850

- Molecule 5 is SUGAR (ALPHA-D-MANNOSE) (three-letter code: MAN) (formula:  $C_6H_{12}O_6$ ).



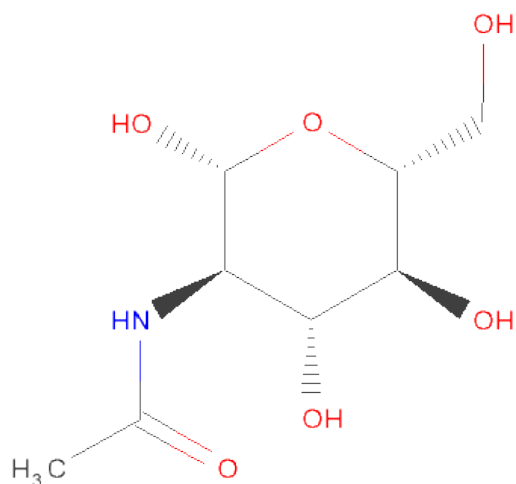
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	D	1	Total	C	O	0	0
			11	6	5		

- Molecule 6 is SULFATE ION (three-letter code: SO4) (formula:  $O_4S$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	D	1	Total	O	S	0	0
			5	4	1		
6	A	1	Total	O	S	0	0
			5	4	1		
6	B	1	Total	O	S	0	0
			5	4	1		
6	C	1	Total	O	S	0	0
			5	4	1		

- Molecule 7 is SUGAR (N-ACETYL-D-GLUCOSAMINE) (three-letter code: NAG) (formula:  $C_8H_{15}NO_6$ ).



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

- Molecule 8 is water.

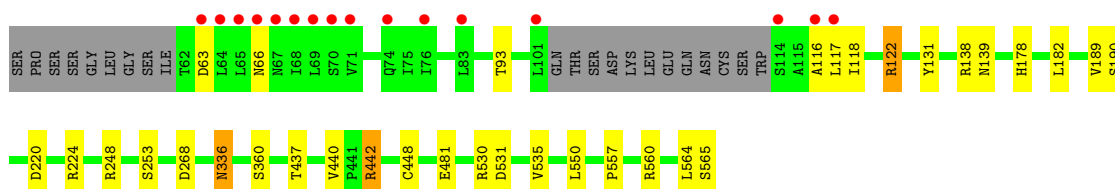
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	D	226	Total	O	0	0
			226	226		
8	A	189	Total	O	0	0
			189	189		
8	B	115	Total	O	0	0
			115	115		
8	C	158	Total	O	0	0
			158	158		

### 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.

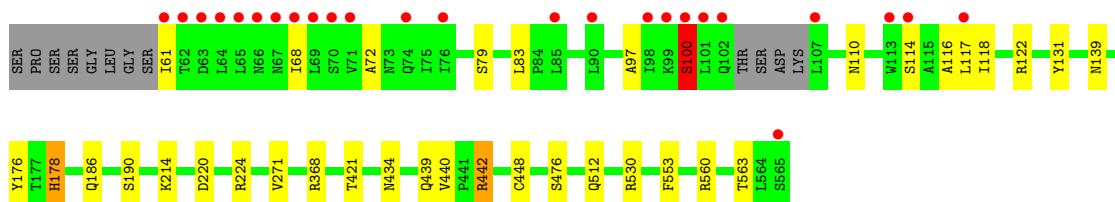
#### • Molecule 1: Hemagglutinin-neuraminidase

Chain D: 



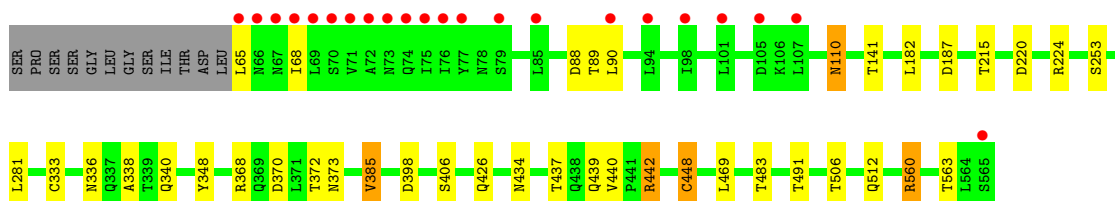
#### • Molecule 1: Hemagglutinin-neuraminidase

Chain A: 



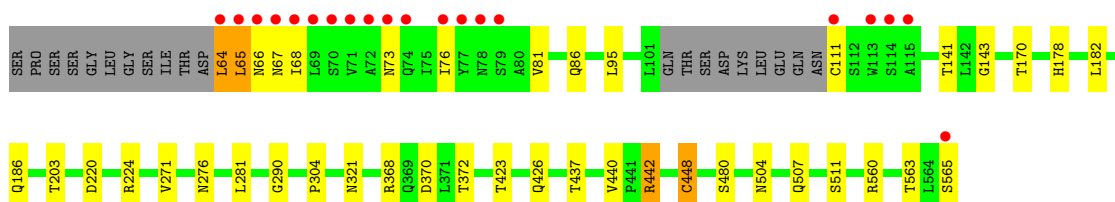
#### • Molecule 1: Hemagglutinin-neuraminidase

Chain B: 



#### • Molecule 1: Hemagglutinin-neuraminidase

Chain C: 





## 4 Data and refinement statistics

Property	Value	Source
Space group	I 4	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	194.39Å 194.39Å 185.98Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	34.36 – 2.50 34.36 – 2.50	Depositor EDS
% Data completeness (in resolution range)	97.8 (34.36-2.50) 97.7 (34.36-2.50)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.68 (at 2.51Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.3_928)	Depositor
R, $R_{free}$	0.166 , 0.207 0.164 , 0.204	Depositor DCC
$R_{free}$ test set	1971 reflections (1.70%)	DCC
Wilson B-factor (Å <sup>2</sup> )	39.7	Xtriage
Anisotropy	0.185	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 30.2	EDS
Estimated twinning fraction	0.028 for l,-k,h 0.029 for -l,-k,-h 0.028 for -h,-l,-k 0.028 for -h,l,k 0.106 for -k,-h,-l	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.27$	Xtriage
Outliers	0 of 118506 reflections	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	16400	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.15% 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: SO4, CA, NAG, MAN

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	0.42	0/3985	0.60	0/5442
1	B	0.39	0/3986	0.55	0/5442
1	C	0.40	0/3919	0.60	2/5351 (0.0%)
1	D	0.43	0/3911	0.60	0/5340
All	All	0.41	0/15801	0.59	2/21575 (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	C	64	LEU	CA-CB-CG	5.77	128.57	115.30
1	C	64	LEU	CB-CG-CD1	-5.47	101.70	111.00

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	3888	0	92	23	0
1	B	3888	0	0	15	0
1	C	3822	0	105	13	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	D	3816	0	28	12	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
3	D	39	0	34	0	0
4	A	56	0	50	0	0
4	B	28	0	25	0	0
4	D	28	0	24	2	0
5	D	11	0	10	2	0
6	A	5	0	0	0	0
6	B	5	0	0	0	0
6	C	5	0	0	0	0
6	D	5	0	0	0	0
7	A	28	0	26	10	0
7	B	14	0	13	0	0
7	C	56	0	52	4	0
7	D	14	0	13	1	0
8	A	189	0	0	4	0
8	B	115	0	0	3	0
8	C	158	0	0	2	0
8	D	226	0	0	2	0
All	All	16400	0	472	67	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 4.

All (67) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:110:ASN:HD21	7:A:608:NAG:C2	1.14	1.58
1:A:110:ASN:HD21	7:A:608:NAG:C1	1.30	1.44
1:A:110:ASN:ND2	7:A:608:NAG:C2	1.91	1.31
1:A:110:ASN:ND2	7:A:608:NAG:H2	1.45	1.25
4:D:605:NAG:O4	5:D:607:MAN:C2	2.20	0.89
1:A:110:ASN:HD22	7:A:608:NAG:H2	1.43	0.82
1:C:442:ARG:NH2	1:C:448:CYS:O	2.15	0.79
7:C:605:NAG:C4	7:C:606:NAG:C1	2.60	0.79
7:C:605:NAG:O4	7:C:606:NAG:C2	2.33	0.76
1:A:442:ARG:NH2	1:A:448:CYS:O	2.18	0.76
1:D:442:ARG:NH2	1:D:448:CYS:O	2.18	0.75
1:A:439:GLN:O	8:A:808:HOH:O	2.06	0.74
1:D:116:ALA:O	1:D:118:ILE:N	2.20	0.74

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:B:442:ARG:NH2	1:B:448:CYS:O	2.21	0.73
1:A:110:ASN:CG	7:A:608:NAG:C1	2.56	0.73
1:D:336:ASN:N	1:D:336:ASN:OD1	2.23	0.72
1:C:111:CYS:N	8:C:835:HOH:O	2.24	0.69
1:B:88:ASP:OD1	1:C:86:GLN:NE2	2.27	0.67
1:B:372:THR:OG1	1:B:373:ASN:N	2.30	0.62
1:B:426:GLN:NE2	8:B:803:HOH:O	2.33	0.61
1:A:100:SER:OG	1:A:100:SER:O	2.17	0.60
1:B:333:CYS:SG	1:B:340:GLN:NE2	2.76	0.59
1:A:110:ASN:ND2	7:A:608:NAG:N2	2.50	0.59
4:D:605:NAG:O4	5:D:607:MAN:H2	2.05	0.57
1:A:116:ALA:O	1:A:118:ILE:N	2.34	0.57
1:A:68:ILE:O	1:A:72:ALA:N	2.39	0.56
1:C:64:LEU:O	1:C:66:ASN:N	2.39	0.55
1:A:214:LYS:NZ	8:A:713:HOH:O	2.40	0.54
1:A:220:ASP:OD1	1:A:224:ARG:NH2	2.41	0.54
1:D:139:ASN:ND2	7:D:609:NAG:C2	2.68	0.53
1:D:122:ARG:NH2	8:D:781:HOH:O	2.42	0.53
1:A:110:ASN:OD1	7:A:608:NAG:C1	2.57	0.53
1:B:560:ARG:NH2	8:B:767:HOH:O	2.41	0.52
1:D:220:ASP:OD1	1:D:224:ARG:NH2	2.44	0.51
1:C:220:ASP:OD1	1:C:224:ARG:NH2	2.44	0.50
1:D:248:ARG:NH2	8:D:804:HOH:O	2.45	0.49
1:B:348:TYR:OH	1:B:385:VAL:O	2.31	0.49
1:B:368:ARG:NH2	1:B:370:ASP:OD2	2.46	0.49
1:D:66:ASN:N	1:D:66:ASN:OD1	2.48	0.47
1:B:439:GLN:NE2	1:B:506:THR:O	2.48	0.47
7:C:605:NAG:H4	7:C:606:NAG:C1	2.42	0.47
1:A:114:SER:HB2	1:B:110:ASN:OD1	2.14	0.47
1:A:434:ASN:ND2	8:A:836:HOH:O	2.47	0.47
1:D:131:TYR:O	1:D:530:ARG:NH2	2.48	0.47
1:B:434:ASN:ND2	8:B:714:HOH:O	2.49	0.46
1:B:220:ASP:OD1	1:B:224:ARG:NH2	2.50	0.45
1:D:138:ARG:NE	1:D:481:GLU:OE2	2.50	0.45
1:A:139:ASN:ND2	7:A:602:NAG:O5	2.41	0.44
1:A:116:ALA:C	1:A:118:ILE:H	2.19	0.44
1:A:139:ASN:ND2	7:A:602:NAG:C2	2.66	0.44
1:A:176:TYR:OH	1:A:178:HIS:ND1	2.51	0.44
1:B:406:SER:O	1:B:442:ARG:NH1	2.51	0.44
1:B:336:ASN:OD1	1:B:338:ALA:N	2.51	0.43
1:D:442:ARG:NE	1:D:442:ARG:O	2.52	0.43
1:C:423:THR:N	1:C:426:GLN:O	2.52	0.43

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:C:290:GLY:N	1:C:304:PRO:O	2.52	0.42
1:A:122:ARG:NH2	8:A:860:HOH:O	2.53	0.42
1:D:122:ARG:NH1	1:D:531:ASP:OD2	2.53	0.42
1:C:426:GLN:NE2	8:C:856:HOH:O	2.53	0.41
7:C:605:NAG:O4	7:C:606:NAG:N2	2.54	0.41
1:C:442:ARG:O	1:C:442:ARG:NE	2.53	0.41
1:C:504:ASN:OD1	1:C:507:GLN:N	2.53	0.41
1:B:469:LEU:N	1:B:483:THR:O	2.54	0.41
1:A:131:TYR:O	1:A:530:ARG:NH2	2.54	0.41
1:C:143:GLY:O	1:C:560:ARG:NH1	2.54	0.41
1:C:276:ASN:N	1:C:372:THR:O	2.54	0.40
1:C:368:ARG:NE	1:C:370:ASP:OD2	2.55	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	497/513 (97%)	463 (93%)	29 (6%)	5 (1%)	22	38
1	B	499/513 (97%)	472 (95%)	25 (5%)	2 (0%)	43	66
1	C	489/513 (95%)	460 (94%)	26 (5%)	3 (1%)	33	55
1	D	488/513 (95%)	466 (96%)	17 (4%)	5 (1%)	22	38
All	All	1973/2052 (96%)	1861 (94%)	97 (5%)	15 (1%)	27	46

All (15) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	117	LEU
1	D	190	SER
1	A	190	SER
1	C	65	LEU
1	A	117	LEU
1	A	97	ALA

Continued on next page...

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	100	SER
1	C	203	THR
1	D	440	VAL
1	B	440	VAL
1	D	535	VAL
1	C	440	VAL
1	D	557	PRO
1	B	68	ILE
1	A	440	VAL

### 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	435/447 (97%)	420 (97%)	15 (3%)	49	75
1	B	435/447 (97%)	416 (96%)	19 (4%)	39	64
1	C	427/447 (96%)	405 (95%)	22 (5%)	32	55
1	D	427/447 (96%)	411 (96%)	16 (4%)	45	72
All	All	1724/1788 (96%)	1652 (96%)	72 (4%)	40	66

All (72) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	63	ASP
1	D	93	THR
1	D	122	ARG
1	D	178	HIS
1	D	182	LEU
1	D	189	VAL
1	D	253	SER
1	D	268	ASP
1	D	336	ASN
1	D	360	SER
1	D	437	THR
1	D	442	ARG
1	D	550	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	D	560	ARG
1	D	564	LEU
1	D	565	SER
1	A	61	ILE
1	A	79	SER
1	A	83	LEU
1	A	100	SER
1	A	178	HIS
1	A	186	GLN
1	A	271	VAL
1	A	368	ARG
1	A	421	THR
1	A	442	ARG
1	A	476	SER
1	A	512	GLN
1	A	553	PHE
1	A	560	ARG
1	A	563	THR
1	B	65	LEU
1	B	89	THR
1	B	90	LEU
1	B	110	ASN
1	B	141	THR
1	B	182	LEU
1	B	187	ASP
1	B	215	THR
1	B	253	SER
1	B	281	LEU
1	B	385	VAL
1	B	398	ASP
1	B	437	THR
1	B	442	ARG
1	B	448	CYS
1	B	491	THR
1	B	512	GLN
1	B	560	ARG
1	B	563	THR
1	C	65	LEU
1	C	67	ASN
1	C	68	ILE
1	C	73	ASN
1	C	76	ILE

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	C	81	VAL
1	C	95	LEU
1	C	141	THR
1	C	170	THR
1	C	178	HIS
1	C	182	LEU
1	C	186	GLN
1	C	271	VAL
1	C	281	LEU
1	C	321	ASN
1	C	437	THR
1	C	442	ARG
1	C	448	CYS
1	C	480	SER
1	C	511	SER
1	C	563	THR
1	C	565	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	110	ASN

### 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 ⓘ

11 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	NAG	A	603	1,4	12,14,15	0.70	0	15,19,21	1.06	1 (6%)
4	NAG	A	604	4	12,14,15	0.66	0	15,19,21	0.95	1 (6%)
4	NAG	A	605	1,4	12,14,15	0.68	0	15,19,21	1.26	1 (6%)
4	NAG	A	606	4	12,14,15	0.74	1 (8%)	15,19,21	0.81	0
4	NAG	B	603	1,4	12,14,15	0.68	0	15,19,21	0.95	1 (6%)
4	NAG	B	604	4	12,14,15	0.57	0	15,19,21	1.06	1 (6%)
3	NAG	D	602	3	12,14,15	0.58	0	15,19,21	0.80	0
3	MAN	D	603	3	10,11,12	0.73	0	11,15,17	1.34	1 (9%)
3	NAG	D	604	1,3	12,14,15	0.74	1 (8%)	15,19,21	1.26	3 (20%)
4	NAG	D	605	5,4	12,14,15	0.70	0	15,19,21	0.96	1 (6%)
4	NAG	D	606	1,4	12,14,15	0.60	0	15,19,21	0.74	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
4	NAG	A	603	1,4	-	0/6/23/26	0/1/1/1
4	NAG	A	604	4	-	0/6/23/26	0/1/1/1
4	NAG	A	605	1,4	-	0/6/23/26	0/1/1/1
4	NAG	A	606	4	-	0/6/23/26	0/1/1/1
4	NAG	B	603	1,4	-	0/6/23/26	0/1/1/1
4	NAG	B	604	4	-	0/6/23/26	0/1/1/1
3	NAG	D	602	3	-	0/6/23/26	0/1/1/1
3	MAN	D	603	3	-	0/2/19/22	0/1/1/1
3	NAG	D	604	1,3	-	0/6/23/26	0/1/1/1
4	NAG	D	605	5,4	-	0/6/23/26	0/1/1/1
4	NAG	D	606	1,4	-	0/6/23/26	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	606	NAG	O5-C5	-2.26	1.41	1.45
3	D	604	NAG	O5-C5	-2.08	1.41	1.45

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	603	MAN	O5-C5-C6	4.24	111.43	106.98
4	B	604	NAG	O5-C5-C6	3.30	110.44	106.98
4	A	605	NAG	C2-N2-C7	-2.84	118.32	123.09
3	D	604	NAG	C2-N2-C7	-2.67	118.61	123.09
4	A	604	NAG	C2-N2-C7	-2.57	118.77	123.09
3	D	604	NAG	O5-C5-C6	2.42	109.52	106.98
3	D	604	NAG	C6-C5-C4	-2.25	107.55	113.00
4	D	605	NAG	C3-C2-N2	-2.07	108.61	111.76
4	B	603	NAG	C3-C2-N2	-2.06	108.62	111.76
4	A	603	NAG	O5-C5-C4	2.02	113.22	110.65

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 5.6 Ligand geometry ⓘ

Of 17 ligands modelled in this entry, 4 are monoatomic - leaving 13 for Mogul analysis.

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	$\# Z  > 2$
7	NAG	A	602	1	12,14,15	0.57	0	15,19,21	0.90	0
6	SO4	A	607	-	4,4,4	0.24	0	6,6,6	0.34	0
7	NAG	A	608	1	12,14,15	0.75	1 (8%)	15,19,21	1.47	2 (13%)
6	SO4	B	602	-	4,4,4	0.19	0	6,6,6	0.15	0
7	NAG	B	605	1	12,14,15	0.73	0	15,19,21	0.85	1 (6%)
6	SO4	C	602	-	4,4,4	0.27	0	6,6,6	0.33	0
7	NAG	C	603	1	12,14,15	0.60	0	15,19,21	1.44	2 (13%)
7	NAG	C	604	1	12,14,15	0.53	0	15,19,21	1.19	2 (13%)
7	NAG	C	605	1,7	12,14,15	0.71	0	15,19,21	1.12	2 (13%)
7	NAG	C	606	7	12,14,15	0.65	0	15,19,21	0.82	0
5	MAN	D	607	4	10,11,12	0.76	0	11,15,17	0.82	0
6	SO4	D	608	-	4,4,4	0.24	0	6,6,6	0.17	0
7	NAG	D	609	1	12,14,15	0.50	0	15,19,21	0.84	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
7	NAG	A	602	1	-	0/6/23/26	0/1/1/1
6	SO4	A	607	-	-	0/0/0/0	0/0/0/0
7	NAG	A	608	1	-	0/6/23/26	0/1/1/1
6	SO4	B	602	-	-	0/0/0/0	0/0/0/0
7	NAG	B	605	1	-	0/6/23/26	0/1/1/1
6	SO4	C	602	-	-	0/0/0/0	0/0/0/0
7	NAG	C	603	1	-	0/6/23/26	0/1/1/1
7	NAG	C	604	1	-	0/6/23/26	0/1/1/1
7	NAG	C	605	1,7	-	0/6/23/26	0/1/1/1
7	NAG	C	606	7	-	0/6/23/26	0/1/1/1
5	MAN	D	607	4	-	0/2/19/22	0/1/1/1
6	SO4	D	608	-	-	0/0/0/0	0/0/0/0
7	NAG	D	609	1	-	0/6/23/26	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	A	608	NAG	O5-C5	-2.23	1.41	1.45

All (9) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
7	A	608	NAG	O5-C5-C6	3.93	111.10	106.98
7	C	603	NAG	C3-C2-N2	3.63	117.29	111.76
7	C	603	NAG	C2-N2-C7	2.52	127.31	123.09
7	A	608	NAG	O5-C5-C4	-2.47	107.52	110.65
7	C	604	NAG	O5-C5-C4	2.40	113.70	110.65
7	C	605	NAG	C2-N2-C7	-2.22	119.37	123.09
7	C	605	NAG	O5-C5-C4	-2.21	107.85	110.65
7	B	605	NAG	C3-C2-N2	-2.09	108.58	111.76
7	C	604	NAG	C2-N2-C7	-2.08	119.59	123.09

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	501/513 (97%)	0.12	25 (4%) 28 28	23, 43, 118, 181	0
1	B	501/513 (97%)	0.14	22 (4%) 33 34	24, 53, 109, 171	0
1	C	493/513 (96%)	0.09	20 (4%) 35 36	22, 45, 98, 169	0
1	D	492/513 (95%)	0.05	16 (3%) 44 45	23, 37, 93, 169	0
All	All	1987/2052 (96%)	0.10	83 (4%) 35 36	22, 44, 106, 181	0

All (83) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	64	LEU	9.6
1	D	64	LEU	8.6
1	A	62	THR	7.8
1	B	65	LEU	7.6
1	C	69	LEU	7.3
1	A	65	LEU	7.2
1	C	70	SER	7.0
1	C	67	ASN	6.7
1	B	69	LEU	6.5
1	B	66	ASN	6.2
1	C	73	ASN	6.1
1	D	76	ILE	6.1
1	D	101	LEU	6.1
1	B	70	SER	6.0
1	B	68	ILE	5.9
1	C	111	CYS	5.8
1	D	66	ASN	5.7
1	C	113	TRP	5.4
1	A	63	ASP	5.3
1	A	61	ILE	5.2
1	B	565	SER	5.1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	D	68	ILE	4.8
1	A	66	ASN	4.7
1	B	75	ILE	4.7
1	C	66	ASN	4.6
1	D	65	LEU	4.6
1	A	67	ASN	4.6
1	B	72	ALA	4.4
1	C	71	VAL	4.2
1	A	69	LEU	4.2
1	B	67	ASN	4.0
1	C	115	ALA	3.8
1	C	68	ILE	3.8
1	C	72	ALA	3.8
1	A	107	LEU	3.7
1	A	71	VAL	3.7
1	B	73	ASN	3.6
1	D	114	SER	3.6
1	C	64	LEU	3.6
1	D	70	SER	3.6
1	D	63	ASP	3.5
1	C	77	TYR	3.5
1	D	71	VAL	3.5
1	A	113	TRP	3.5
1	B	105	ASP	3.4
1	A	101	LEU	3.3
1	B	79	SER	3.2
1	A	68	ILE	3.1
1	A	70	SER	3.0
1	A	90	LEU	3.0
1	D	117	LEU	3.0
1	A	76	ILE	3.0
1	A	102	GLN	3.0
1	A	117	LEU	3.0
1	B	94	LEU	3.0
1	A	565	SER	3.0
1	C	114	SER	3.0
1	B	98	ILE	3.0
1	B	71	VAL	2.9
1	B	74	GLN	2.9
1	C	65	LEU	2.8
1	B	107	LEU	2.8
1	B	85	LEU	2.6

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	D	83	LEU	2.6
1	C	78	ASN	2.5
1	C	79	SER	2.5
1	D	116	ALA	2.5
1	D	69	LEU	2.5
1	A	74	GLN	2.5
1	C	76	ILE	2.5
1	C	565	SER	2.4
1	B	101	LEU	2.4
1	B	77	TYR	2.3
1	A	99	LYS	2.3
1	D	74	GLN	2.3
1	B	76	ILE	2.3
1	A	85	LEU	2.3
1	A	98	ILE	2.2
1	A	100	SER	2.2
1	A	114	SER	2.2
1	D	67	ASN	2.2
1	C	74	GLN	2.1
1	B	90	LEU	2.0

## 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(Å <sup>2</sup> )	Q<0.9
3	NAG	D	604	14/15	0.15	1.64	38,51,63,73	0
4	NAG	A	604	14/15	0.18	1.51	64,77,94,95	0
4	NAG	B	604	14/15	0.18	0.21	88,101,109,112	0
4	NAG	A	605	14/15	0.14	0.19	50,63,73,75	0
4	NAG	A	603	14/15	0.14	0.05	36,46,56,66	0
4	NAG	B	603	14/15	0.12	-0.79	51,57,67,78	0

*Continued on next page...*

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	NAG	D	602	14/15	0.12	-1.05	73,78,82,83	0
4	NAG	D	606	14/15	0.12	-1.15	43,53,61,69	0
3	MAN	D	603	11/12	0.14	-	91,108,111,112	0
4	NAG	A	606	14/15	0.16	-	57,91,97,98	0
4	NAG	D	605	14/15	0.11	-	58,76,91,101	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
6	SO4	C	602	5/5	0.17	4.18	50,57,71,80	0
2	CA	B	601	1/1	0.17	2.10	92,92,92,92	0
6	SO4	B	602	5/5	0.17	1.88	66,72,78,86	0
7	NAG	B	605	14/15	0.21	0.91	45,68,77,80	0
7	NAG	A	602	14/15	0.21	0.20	66,89,100,101	0
6	SO4	D	608	5/5	0.14	-0.14	48,60,61,67	0
7	NAG	C	605	14/15	0.12	-0.42	41,53,62,67	0
7	NAG	D	609	14/15	0.13	-0.55	66,81,91,92	0
7	NAG	C	603	14/15	0.12	-0.60	81,94,100,100	0
6	SO4	A	607	5/5	0.13	-0.66	44,54,65,65	0
7	NAG	C	604	14/15	0.13	-0.68	64,69,77,80	0
7	NAG	A	608	14/15	0.13	-1.19	99,108,112,112	0
2	CA	D	601	1/1	0.11	-1.82	80,80,80,80	0
2	CA	A	601	1/1	0.09	-2.47	74,74,74,74	0
2	CA	C	601	1/1	0.07	-3.67	81,81,81,81	0
5	MAN	D	607	11/12	0.12	-	103,111,115,115	0
7	NAG	C	606	14/15	0.19	-	42,80,99,103	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.