



# Full wwPDB X-ray Structure Validation Report ⓘ

Nov 14, 2023 – 11:11 PM JST

PDB ID : 6ID9  
Title : Crystal structure of H7 hemagglutinin mutant H7-SGTL ( A138S, V186G, P221T) from the influenza virus A/Anhui/1/2013 (H7N9)  
Authors : Gao, G.F.; Xu, Y.; Qi, J.X.  
Deposited on : 2018-09-09  
Resolution : 2.70 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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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

MolProbity : 4.02b-467  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.36  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0158  
CCP4 : 7.0.044 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.36

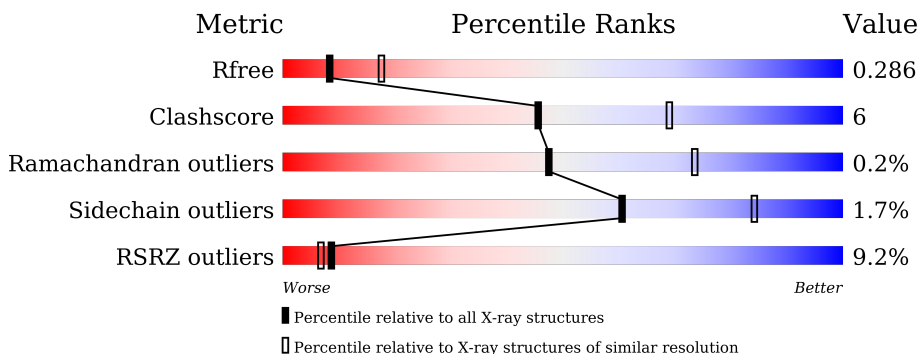
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.70 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))
$R_{free}$	130704	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria respectively. 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	321	 6% 81% 16% ..
2	B	177	 14% 76% 15% • 8%

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	602	-	-	-	X

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 3780 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 HA1 chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	314	2385	1479	432	459	15	0	0	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	128	SER	ALA	engineered mutation	UNP R4NN21
A	177	GLY	VAL	engineered mutation	UNP R4NN21
A	212	THR	PRO	engineered mutation	UNP R4NN21

- Molecule 2 is a protein called Hemagglutinin HA2 chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	163	1328	817	231	273	7	0	0	0

- Molecule 3 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: C<sub>8</sub>H<sub>15</sub>NO<sub>6</sub>).



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

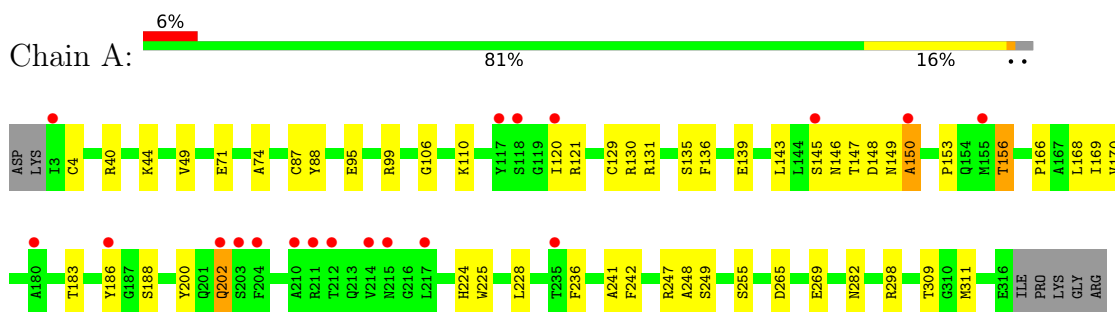
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	14	14	14	0	0
4	B	11	11	11	0	0

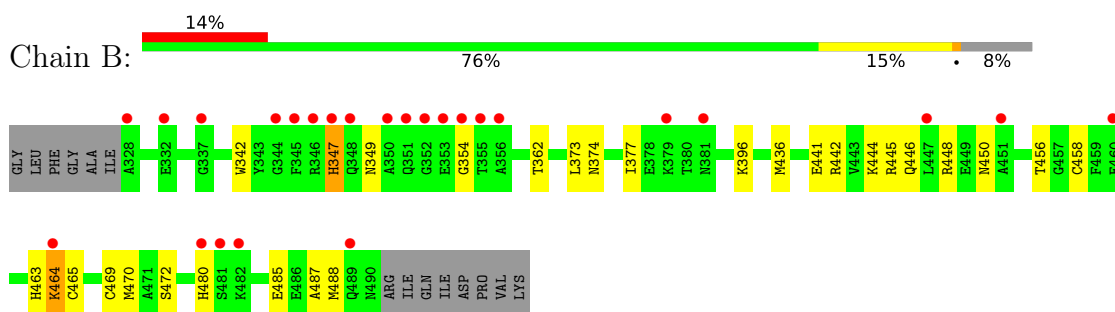
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: Hemagglutinin HA1 chain



- Molecule 2: Hemagglutinin HA2 chain



## 4 Data and refinement statistics i

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	116.27Å 116.27Å 293.84Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	41.53 – 2.70 41.53 – 2.69	Depositor EDS
% Data completeness (in resolution range)	97.4 (41.53-2.70) 97.4 (41.53-2.69)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.91 (at 2.69Å)	Xtrriage
Refinement program	PHENIX (1.11.1_2575: ???)	Depositor
R, $R_{free}$	0.245 , 0.287 0.245 , 0.286	Depositor DCC
$R_{free}$ test set	1102 reflections (5.25%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	61.9	Xtrriage
Anisotropy	0.385	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 57.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.35$	Xtrriage
Estimated twinning fraction	0.004 for $-1/3^*h+1/3^*k+1/3^*l,-k,8/3^*h+4/3^*k+1/3^*l$ 0.012 for $-2/3^*h-1/3^*k-1/3^*l,-1/3^*h-2/3^*k+1/3^*l,-4/3^*h+4/3^*k+1/3^*l$ 0.011 for $-h,1/3^*h-1/3^*k-1/3^*l,-4/3^*h-8/3^*k+1/3^*l$	Xtrriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	3780	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	87.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.28% of the height of the origin peak. No significant pseudotranslation is detected.*

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

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: 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	0.25	0/2430	0.45	0/3285
2	B	0.25	0/1351	0.46	2/1821 (0.1%)
All	All	0.25	0/3781	0.45	2/5106 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
2	B	464	LYS	CB-CA-C	7.29	124.99	110.40
2	B	464	LYS	N-CA-C	-6.77	92.71	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	202	GLN	Peptide

### 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	2385	0	2331	32	0
2	B	1328	0	1222	18	0
3	A	28	0	26	0	0
3	B	14	0	13	0	0
4	A	14	0	0	1	0
4	B	11	0	0	2	0
All	All	3780	0	3592	47	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 6.

All (47) 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:A:169:ILE:O	1:A:225:TRP:HA	1.95	0.66
1:A:71:GLU:OE2	1:A:247:ARG:NH2	2.30	0.64
2:B:450:ASN:ND2	4:B:601:HOH:O	2.33	0.62
1:A:168:LEU:HB3	1:A:249:SER:HB2	1.84	0.60
1:A:121:ARG:NH1	1:A:145:SER:O	2.36	0.58
1:A:147:THR:HG23	1:A:149:ASN:O	2.05	0.56
1:A:106:GLY:HA2	1:A:255:SER:HB3	1.88	0.55
1:A:309:THR:HG22	2:B:373:LEU:HD11	1.87	0.55
2:B:444:LYS:NZ	4:B:602:HOH:O	2.40	0.54
1:A:200:TYR:CE2	1:A:202:GLN:HB2	2.43	0.54
2:B:374:ASN:HA	2:B:377:ILE:HG22	1.91	0.53
2:B:463:HIS:HB3	2:B:487:ALA:HB2	1.90	0.53
1:A:170:VAL:HG22	1:A:225:TRP:HB3	1.90	0.53
2:B:441:GLU:HB3	2:B:445:ARG:NH1	2.25	0.52
1:A:110:LYS:HA	1:A:248:ALA:O	2.10	0.51
1:A:147:THR:OG1	1:A:148:ASP:N	2.43	0.51
1:A:49:VAL:HG23	1:A:74:ALA:HB2	1.92	0.50
1:A:87:CYS:SG	1:A:88:TYR:N	2.85	0.50
1:A:120:ILE:HD11	1:A:153:PRO:HD2	1.94	0.49
1:A:156:THR:HA	1:A:236:PHE:O	2.12	0.49
1:A:143:LEU:HD12	1:A:242:PHE:HD2	1.78	0.49
2:B:463:HIS:O	2:B:464:LYS:C	2.52	0.47
1:A:40:ARG:HD2	1:A:265:ASP:HB2	1.97	0.45
1:A:149:ASN:HB3	1:A:150:ALA:H	1.53	0.45
2:B:448:ARG:HD3	2:B:480:HIS:HB2	1.98	0.45

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:170:VAL:HA	1:A:224:HIS:O	2.16	0.45
2:B:347:HIS:HE1	2:B:354:GLY:HA3	1.81	0.45
2:B:342:TRP:HB2	2:B:362:THR:HG22	1.99	0.44
1:A:44:LYS:HE3	1:A:269:GLU:HB2	2.00	0.44
2:B:485:GLU:HA	2:B:488:MET:HB2	1.99	0.44
1:A:183:THR:HG22	1:A:188:SER:HA	2.00	0.43
2:B:463:HIS:ND1	2:B:464:LYS:O	2.51	0.43
1:A:166:PRO:HA	1:A:228:LEU:O	2.18	0.43
1:A:130:ARG:NE	1:A:131:ARG:O	2.45	0.42
2:B:349:ASN:OD1	2:B:349:ASN:N	2.44	0.42
1:A:95:GLU:HB3	1:A:99:ARG:NH1	2.34	0.42
1:A:121:ARG:NH1	1:A:146:ASN:O	2.53	0.42
1:A:139:GLU:OE1	1:A:247:ARG:HD3	2.20	0.42
2:B:396:LYS:HE2	2:B:396:LYS:HB2	1.88	0.41
1:A:135:SER:OG	1:A:136:PHE:N	2.49	0.41
2:B:442:ARG:NH1	2:B:446:GLN:OE1	2.53	0.41
1:A:282:ASN:HB2	2:B:377:ILE:HD11	2.01	0.41
1:A:186:TYR:CZ	1:A:241:ALA:HA	2.55	0.41
1:A:298:ARG:NH2	4:A:706:HOH:O	2.53	0.41
1:A:311:MET:HE3	1:A:311:MET:HB3	1.88	0.41
2:B:465:CYS:SG	2:B:470:MET:HG3	2.61	0.41
1:A:4:CYS:HA	2:B:458:CYS:HA	2.02	0.40

There are no symmetry-related clashes.

## 5.3 Torsion angles [i](#)

### 5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	312/321 (97%)	288 (92%)	23 (7%)	1 (0%)	41 66
2	B	161/177 (91%)	153 (95%)	8 (5%)	0	100 100
All	All	473/498 (95%)	441 (93%)	31 (7%)	1 (0%)	47 73

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	150	ALA

### 5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	261/269 (97%)	259 (99%)	2 (1%)	81 93
2	B	141/152 (93%)	136 (96%)	5 (4%)	36 65
All	All	402/421 (96%)	395 (98%)	7 (2%)	60 84

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	129	CYS
1	A	156	THR
2	B	347	HIS
2	B	436	MET
2	B	456	THR
2	B	469	CYS
2	B	472	SER

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

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

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

## 5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

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	B	501	2	14,14,15	0.27	0	17,19,21	0.46	0
3	NAG	A	602	1	14,14,15	0.21	0	17,19,21	0.42	0
3	NAG	A	601	1	14,14,15	0.24	0	17,19,21	0.42	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	B	501	2	-	0/6/23/26	0/1/1/1
3	NAG	A	602	1	-	2/6/23/26	0/1/1/1
3	NAG	A	601	1	-	1/6/23/26	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	602	NAG	O5-C5-C6-O6
3	A	602	NAG	C4-C5-C6-O6
3	A	601	NAG	O5-C5-C6-O6

There are no ring outliers.

No monomer is involved in short contacts.

## 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	314/321 (97%)	0.39	19 (6%) 21 20	34, 74, 138, 200	0
2	B	163/177 (92%)	0.85	25 (15%) 2 1	36, 100, 159, 171	0
All	All	477/498 (95%)	0.55	44 (9%) 9 7	34, 81, 149, 200	0

All (44) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	356	ALA	5.7
2	B	460	GLU	5.5
1	A	212	THR	4.8
2	B	379	LYS	4.4
2	B	447	LEU	4.3
2	B	352	GLY	4.2
2	B	353	GLU	4.1
1	A	117	TYR	4.1
1	A	203	SER	3.9
2	B	348	GLN	3.8
1	A	214	VAL	3.6
1	A	150	ALA	3.4
1	A	211	ARG	3.3
2	B	328	ALA	3.2
1	A	3	ILE	3.2
2	B	347	HIS	3.1
2	B	337	GLY	3.1
2	B	355	THR	3.0
1	A	118	SER	3.0
1	A	120	ILE	3.0
1	A	215	ASN	3.0
1	A	186	TYR	3.0
2	B	350	ALA	3.0
2	B	451	ALA	2.9

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	B	354	GLY	2.8
2	B	345	PHE	2.7
1	A	145	SER	2.7
2	B	464	LYS	2.7
2	B	332	GLU	2.6
1	A	235	THR	2.5
2	B	381	ASN	2.5
1	A	155	MET	2.5
1	A	180	ALA	2.5
1	A	217	LEU	2.5
2	B	344	GLY	2.4
1	A	202	GLN	2.4
2	B	489	GLN	2.3
2	B	351	GLN	2.3
1	A	210	ALA	2.3
2	B	480	HIS	2.2
1	A	204	PHE	2.2
2	B	482	LYS	2.1
2	B	481	SER	2.0
2	B	346	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](#)

There are no monosaccharides in this entry.

## 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. 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	B-factors(Å <sup>2</sup> )	Q<0.9
3	NAG	A	602	14/15	0.72	0.47	112,132,143,148	0
3	NAG	A	601	14/15	0.74	0.33	101,118,131,135	0
3	NAG	B	501	14/15	0.88	0.14	69,84,91,95	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.