



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 17, 2016 – 10:41 PM EDT

PDB ID : 5E66  
Title : The complex structure of Hemagglutinin-esterase-fusion mutant protein from the influenza D virus with receptor analog 9-N-Ac-Sia  
Authors : Song, H.; Qi, J.; Shi, Y.; Gao, G.F.  
Deposited on : 2015-10-09  
Resolution : 3.10 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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 : 1.7.1 (RC1), CSD as537be (2016)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20027107  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0122  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20027107

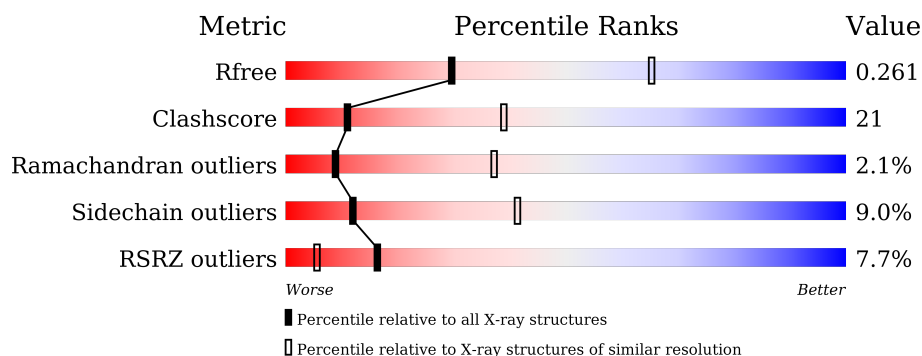
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.10 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	1114 (3.14-3.06)
Clashscore	102246	1222 (3.14-3.06)
Ramachandran outliers	100387	1174 (3.14-3.06)
Sidechain outliers	100360	1174 (3.14-3.06)
RSRZ outliers	91569	1119 (3.14-3.06)

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	427	<div> <div>3%</div> <div>59%</div> <div>37%</div> <div>..</div> </div>
2	B	166	<div> <div>19%</div> <div>49%</div> <div>32%</div> <div>9%</div> <div>10%</div> </div>

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	CAC	A	501	-	-	X	X

## 2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 4575 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.

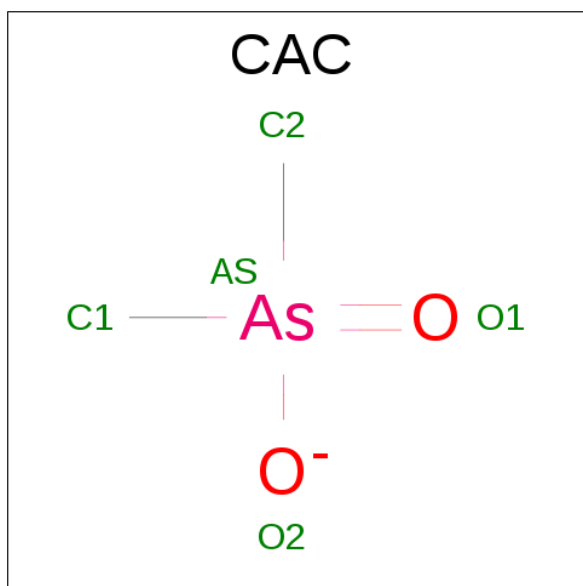
- Molecule 1 is a protein called Hemagglutinin-esterase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	424	Total	C	N	O	S	0	0	0
			3249	2043	542	642	22			

- Molecule 2 is a protein called Hemagglutinin-esterase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	149	Total	C	N	O	S	0	0	0
			1104	689	194	217	4			

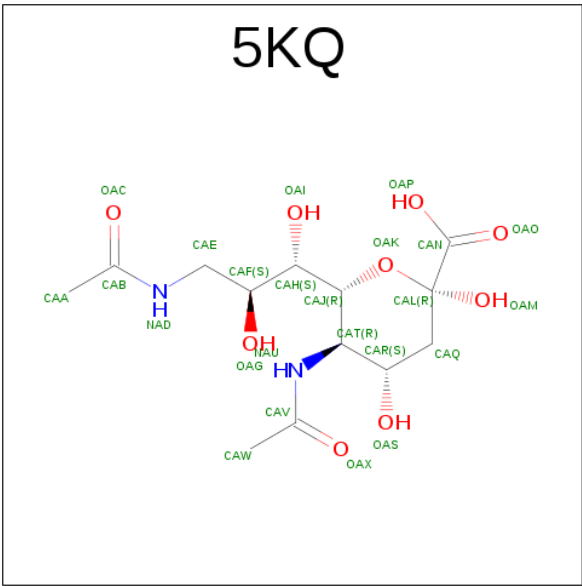
- Molecule 3 is CACODYLATE ION (three-letter code: CAC) (formula:  $C_2H_6AsO_2$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	As	C	O	0	0
			4	1	2	1		

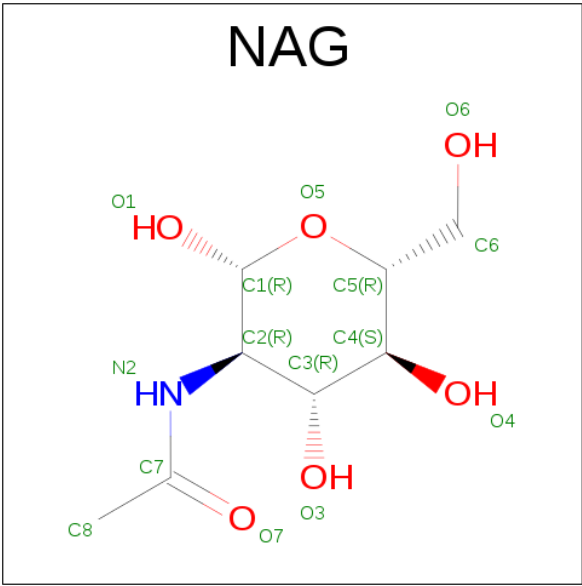
- Molecule 4 is (6R)-5-(acetylamino)-6-[(1S,2S)-3-(acetylamino)-1,2-dihydroxypropyl]-3,5-dideoxy-beta-L-threo-hex-2-ulopyranosonic acid (three-letter code: 5KQ) (formula:

C<sub>13</sub>H<sub>22</sub>N<sub>2</sub>O<sub>9</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total	C	N	O	0	0
			24	13	2	9		
4	A	1	Total	C	N	O	0	0
			24	13	2	9		

- Molecule 5 is N-ACETYL-D-GLUCOSAMINE (three-letter code: NAG) (formula: C<sub>8</sub>H<sub>15</sub>NO<sub>6</sub>).



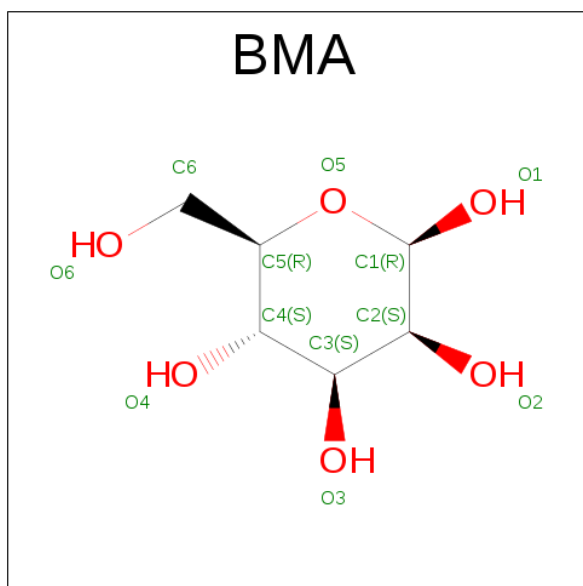
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	1	Total	C	N	O	0	0
			14	8	1	5		

Continued on next page...

Continued from previous page...

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

- Molecule 6 is BETA-D-MANNOSE (three-letter code: BMA) (formula:  $C_6H_{12}O_6$ ).



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

- Molecule 7 is ALPHA-D-MANNOSE (three-letter code: MAN) (formula:  $C_6H_{12}O_6$ ).

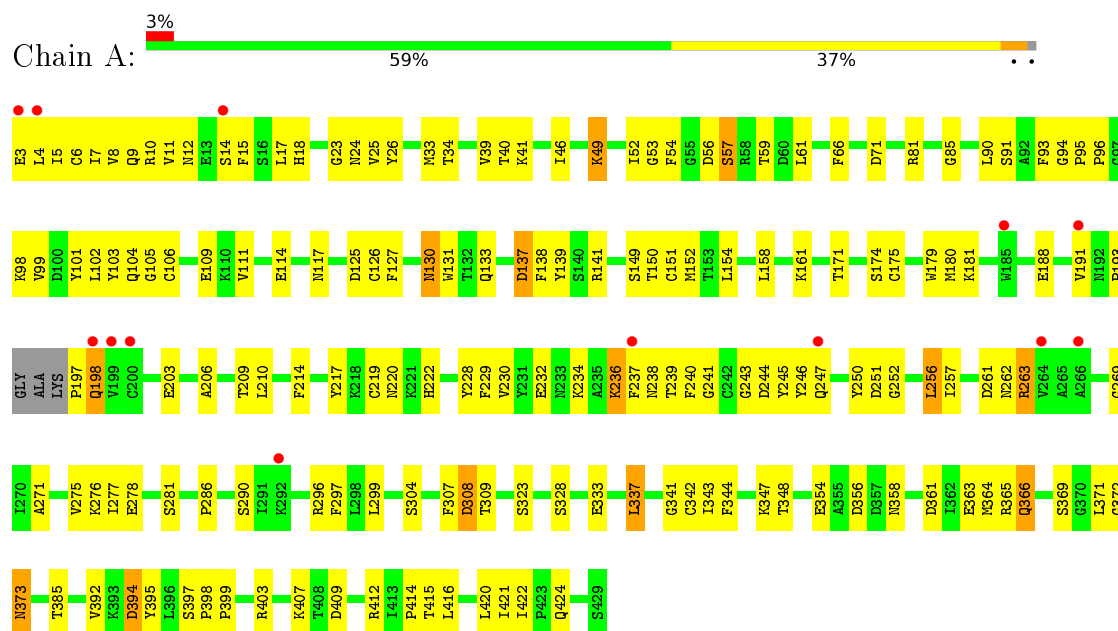


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
7	A	1	Total	C	O	0	0
			11	6	5		
7	A	1	Total	C	O	0	0
			11	6	5		
7	A	1	Total	C	O	0	0
			11	6	5		

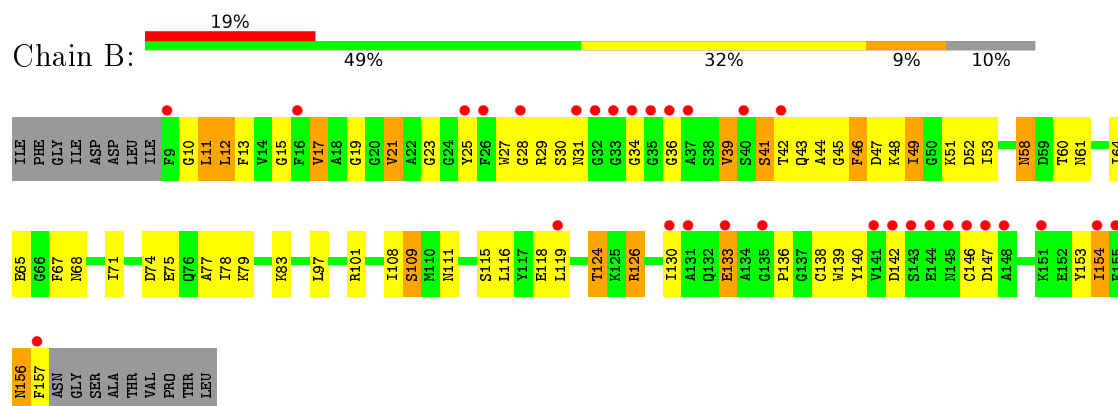
### 3 Residue-property plots

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-esterase



#### • Molecule 2: Hemagglutinin-esterase





## 4 Data and refinement statistics

Property	Value	Source
Space group	I 21 3	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	165.30Å 165.30Å 165.30Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	38.96 – 3.10 38.96 – 3.09	Depositor EDS
% Data completeness (in resolution range)	99.9 (38.96-3.10) 94.7 (38.96-3.09)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	0.08	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.96 (at 3.12Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, $R_{free}$	0.234 , 0.265 0.231 , 0.261	Depositor DCC
$R_{free}$ test set	629 reflections (4.81%)	DCC
Wilson B-factor (Å <sup>2</sup> )	93.7	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.26 , 57.3	EDS
Estimated twinning fraction	0.034 for -l,-k,-h	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 13871 reflections	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	4575	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	117.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.17% 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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: CAC, BMA, NAG, 5KQ, 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	0.55	0/3325	0.70	0/4507
2	B	0.52	0/1120	0.67	0/1506
All	All	0.55	0/4445	0.69	0/6013

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

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	3249	0	3114	137	1
2	B	1104	0	1060	65	0
3	A	4	0	0	9	0
4	A	48	0	0	8	1
5	A	84	0	74	3	0
5	B	42	0	38	1	0
6	A	11	0	9	0	0
7	A	33	0	28	0	0
All	All	4575	0	4323	186	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 21.

The worst 5 of 186 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:415:THR:O	2:B:101:ARG:NH1	2.03	0.91
5:A:511:NAG:O6	2:B:23:GLY:N	2.05	0.89
1:A:297:PHE:HE2	4:A:513:5KQ:OAC	1.55	0.88
1:A:354:GLU:OE2	1:A:365:ARG:NH2	2.11	0.84
3:A:501:CAC:C1	4:A:502:5KQ:OAG	2.28	0.81

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:14:SER:OG	4:A:502:5KQ:OAS[14_545]	1.89	0.31

## 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	420/427 (98%)	371 (88%)	47 (11%)	2 (0%)	34	72
2	B	147/166 (89%)	112 (76%)	25 (17%)	10 (7%)	1	8
All	All	567/593 (96%)	483 (85%)	72 (13%)	12 (2%)	9	37

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	46	PHE
2	B	109	SER
2	B	49	ILE
2	B	133	GLU
2	B	154	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	360/361 (100%)	330 (92%)	30 (8%)	14	46
2	B	108/123 (88%)	96 (89%)	12 (11%)	8	29
All	All	468/484 (97%)	426 (91%)	42 (9%)	12	41

5 of 42 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	304	SER
1	A	364	MET
2	B	126	ARG
1	A	308	ASP
1	A	328	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	238	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry

16 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	CAC	A	501	-	0,3,4	0.00	-	0,3,6	0.00	-
4	5KQ	A	502	-	20,24,24	1.01	1 (5%)	21,35,35	1.05	1 (4%)
5	NAG	A	503	1,5	14,14,15	0.87	2 (14%)	15,19,21	0.35	0
5	NAG	A	504	5,6	14,14,15	0.73	1 (7%)	15,19,21	0.65	0
6	BMA	A	505	5,7	11,11,12	2.10	6 (54%)	15,15,17	2.44	8 (53%)
7	MAN	A	506	7,6	11,11,12	1.03	0	15,15,17	1.59	2 (13%)
7	MAN	A	507	7	11,11,12	1.37	3 (27%)	15,15,17	1.26	1 (6%)
7	MAN	A	508	7	11,11,12	2.17	4 (36%)	15,15,17	2.35	4 (26%)
5	NAG	A	509	1,5	14,14,15	0.36	0	15,19,21	0.54	0
5	NAG	A	510	5	14,14,15	0.39	0	15,19,21	1.05	2 (13%)
5	NAG	A	511	1,5	14,14,15	0.44	0	15,19,21	0.80	1 (6%)
5	NAG	A	512	5	14,14,15	1.20	2 (14%)	15,19,21	0.68	1 (6%)
4	5KQ	A	513	-	20,24,24	1.11	1 (5%)	21,35,35	1.40	2 (9%)
5	NAG	B	701	2,5	14,14,15	1.02	2 (14%)	15,19,21	0.55	0
5	NAG	B	702	5	14,14,15	0.41	0	15,19,21	0.56	0
5	NAG	B	703	2	14,14,15	0.44	0	15,19,21	0.60	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	CAC	A	501	-	-	0/0/0/0	0/0/0/0
4	5KQ	A	502	-	-	0/17/41/41	0/1/1/1
5	NAG	A	503	1,5	-	0/6/23/26	0/1/1/1
5	NAG	A	504	5,6	-	0/6/23/26	0/1/1/1
6	BMA	A	505	5,7	-	0/2/19/22	0/1/1/1
7	MAN	A	506	7,6	-	0/2/19/22	0/1/1/1

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
7	MAN	A	507	7	-	0/2/19/22	0/1/1/1
7	MAN	A	508	7	-	0/2/19/22	0/1/1/1
5	NAG	A	509	1,5	-	0/6/23/26	0/1/1/1
5	NAG	A	510	5	-	0/6/23/26	0/1/1/1
5	NAG	A	511	1,5	-	0/6/23/26	0/1/1/1
5	NAG	A	512	5	-	0/6/23/26	0/1/1/1
4	5KQ	A	513	-	-	0/17/41/41	0/1/1/1
5	NAG	B	701	2,5	-	0/6/23/26	0/1/1/1
5	NAG	B	702	5	-	0/6/23/26	0/1/1/1
5	NAG	B	703	2	-	0/6/23/26	0/1/1/1

The worst 5 of 22 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	B	701	NAG	O5-C1	-2.96	1.38	1.43
7	A	507	MAN	O5-C1	-2.81	1.39	1.43
6	A	505	BMA	O5-C1	-2.55	1.39	1.43
5	A	504	NAG	O5-C1	-2.54	1.39	1.43
5	A	503	NAG	O5-C1	-2.29	1.40	1.43

The worst 5 of 22 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
7	A	508	MAN	C1-C2-C3	-7.03	101.03	109.55
6	A	505	BMA	C3-C4-C5	-3.93	103.22	110.23
4	A	502	5KQ	CAH-CAJ-CAT	-3.24	109.59	114.06
7	A	507	MAN	C1-C2-C3	-3.17	105.71	109.55
6	A	505	BMA	C2-C3-C4	-2.90	105.99	111.05

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

6 monomers are involved in 19 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	501	CAC	9	0
4	A	502	5KQ	3	1
5	A	511	NAG	3	0
4	A	513	5KQ	5	0
5	B	701	NAG	1	0
5	B	702	NAG	1	0

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

## 6 Fit of model and data [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

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	424/427 (99%)	-0.07	13 (3%)	52 28	61, 94, 168, 259	0
2	B	149/166 (89%)	0.88	31 (20%)	1 0	61, 154, 255, 358	0
All	All	573/593 (96%)	0.18	44 (7%)	16 5	61, 101, 220, 358	0

The worst 5 of 44 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	141	VAL	9.1
2	B	142	ASP	8.3
2	B	147	ASP	7.0
2	B	31	ASN	6.5
2	B	32	GLY	5.7

### 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 carbohydrates 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. 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( $\text{\AA}^2$ )	Q<0.9
3	CAC	A	501	4/5	0.87	0.42	2.61	135,227,251,410	0
4	5KQ	A	502	24/24	0.79	0.27	0.13	72,153,168,181	0
4	5KQ	A	513	24/24	0.88	0.26	0.07	83,126,141,156	0
5	NAG	B	702	14/15	0.81	0.20	-0.33	122,146,158,161	0
5	NAG	A	511	14/15	0.92	0.22	-0.52	123,140,230,232	0
5	NAG	B	701	14/15	0.95	0.16	-0.99	106,130,149,155	0
5	NAG	A	512	14/15	0.87	0.26	-	103,115,170,179	0
7	MAN	A	508	11/12	0.84	0.23	-	76,103,135,143	0
5	NAG	A	504	14/15	0.94	0.18	-	77,95,106,109	0
5	NAG	A	503	14/15	0.96	0.12	-	78,87,116,125	0
5	NAG	A	510	14/15	0.82	0.20	-	90,110,120,123	0
6	BMA	A	505	11/12	0.85	0.17	-	75,78,102,114	0
5	NAG	A	509	14/15	0.92	0.14	-	82,106,120,124	0
7	MAN	A	506	11/12	0.97	0.12	-	76,88,99,99	0
5	NAG	B	703	14/15	0.79	0.24	-	134,158,184,185	0
7	MAN	A	507	11/12	0.91	0.16	-	102,117,129,130	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.