W O R L D W I D E

wwPDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 06:09 PM GMT

PDB ID	:	4KPS
Title	:	Structure and receptor binding specificity of the hemagglutinin H13 from avian influenza A virus H13N6
Authors	:	Lu, X.; Qi, J.; Shi, Y.; Gao, G.
Deposited on	:	2013-05-14
Resolution	:	2.59 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
A user guide is available at
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the ⓘ symbol.

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

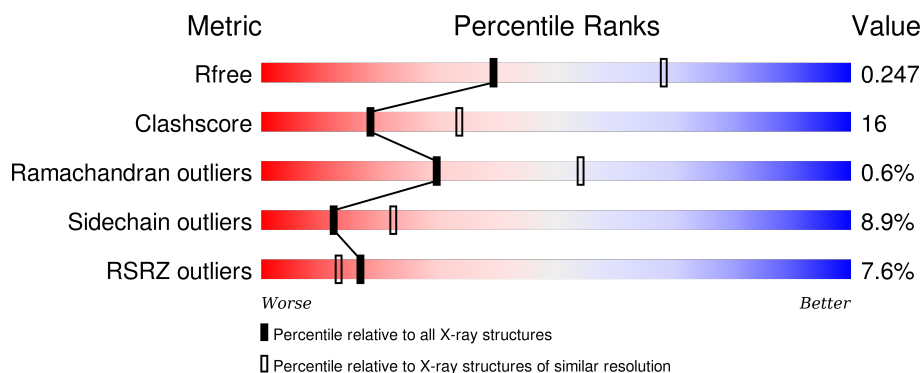|		
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.59 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	Percentile Ranks	Value
Rfree		0.247
Clashscore		16
Ramachandran outliers		0.6%
Sidechain outliers		8.9%
RSRZ outliers		7.6%

Worse   Better

■ Percentile relative to all X-ray structures
▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	91344	2636 (2.60-2.56)
Clashscore	102246	3003 (2.60-2.56)
Ramachandran outliers	100387	2956 (2.60-2.56)
Sidechain outliers	100360	2956 (2.60-2.56)
RSRZ outliers	91569	2642 (2.60-2.56)

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 >=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 <=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	324	5% / 71% / 25% / •
1	C	324	9% / 73% / 23% / •
1	E	324	4% / 72% / 24% / •
2	B	165	7% / 67% / 30% / •
2	D	165	8% / 73% / 24% / •

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	165	

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 11796 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Hemagglutinin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	324	Total 2555	C 1610	N 448	O 485	S 12	0	0	0
1	C	324	Total 2555	C 1610	N 448	O 485	S 12	0	0	0
1	E	324	Total 2555	C 1610	N 448	O 485	S 12	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	4	ILE	-	EXPRESSION TAG	UNP P13103
A	5	GLN	-	EXPRESSION TAG	UNP P13103
C	4	ILE	-	EXPRESSION TAG	UNP P13103
C	5	GLN	-	EXPRESSION TAG	UNP P13103
E	4	ILE	-	EXPRESSION TAG	UNP P13103
E	5	GLN	-	EXPRESSION TAG	UNP P13103

- Molecule 2 is a protein called Hemagglutinin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	165	Total 1312	C 811	N 229	O 265	S 7	0	0	0
2	D	165	Total 1312	C 811	N 229	O 265	S 7	0	0	0
2	F	165	Total 1312	C 811	N 229	O 265	S 7	0	0	0

- Molecule 3 is a polymer of unknown type called SUGAR (2-MER).

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

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

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

- Molecule 5 is a polymer of unknown type called SUGAR (3-MER).

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

- Molecule 6 is water.

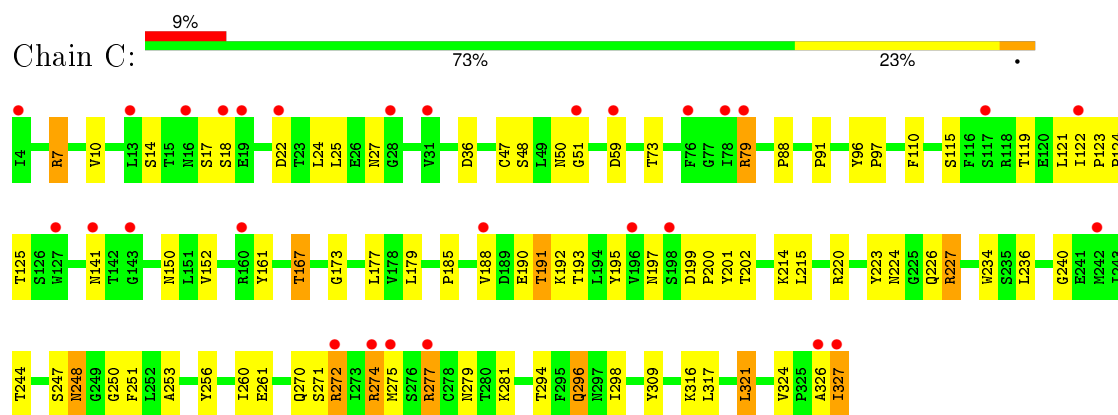| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
6	A	15	15	15	0	0
6	B	19	19	19	0	0
6	C	20	20	20	0	0
6	D	13	13	13	0	0
6	E	25	25	25	0	0
6	F	11	11	11	0	0

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.
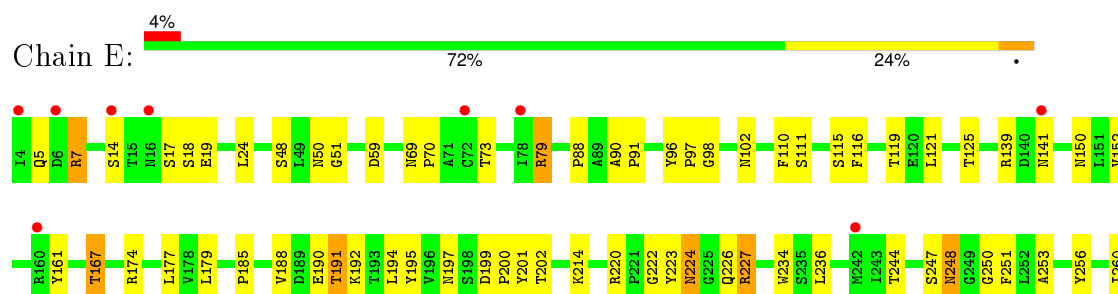
- Molecule 1: Hemagglutinin

- Molecule 1: Hemagglutinin
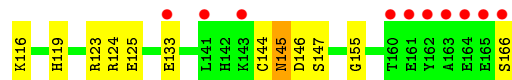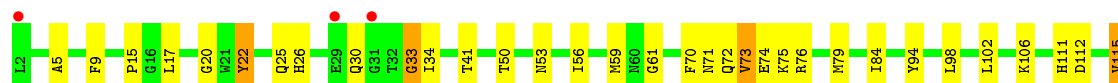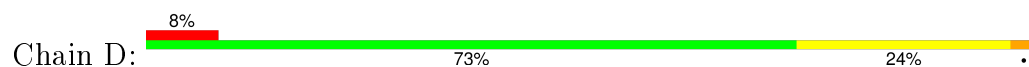

- Molecule 1: Hemagglutinin

E261 R267 Q270 S271 R272 I273 R274 R277 C278 N279 T280 K281 N292 R293 T294 F295 N297 D299 P307 K308 Y309 L317 L321 R322 N323 V324 P325 A326 I327

• Molecule 2: Hemagglutinin

Chain B:

7%

67%   30%

L2 F3 G4 A5 F9 I10 E11 P15 G16 L17 I18 N19 G20 W21 Y22 G23 F24 Q25 H26 E29 Q30 G31 T32 G33 I34 T41 I48 I52 I56 M59 R67 G68 E69 F70 N71 Q72 V73 E74 K75 M79 D85 D86 A87 V88 I91 Y94

L98 L102 K106 D109 M110 H111 D112 A113 N114 V115 K116 R123 R124 E125 C144 N145 D146 S147 C148 G155 E161 Y162 A163 E164 E165 S166

• Molecule 2: Hemagglutinin

Chain D:

8%

73%   24%

L2 A5 F9 P15 G16 L17 G20 W21 Y22 Q25 H26 E29 Q30 G31 T32 G33 I34 T41 T50 N53 I56 M59 N60 G61 F70 N71 Q72 V73 E74 K75 R76 M79 I84 Y94 L98 L102 K106 H111 D112 V115

K116 H119 R123 R124 E125 E133 L141 H142 K143 C144 N145 D146 S147 G155 T160 E161 Y162 A163 E164 E165 S166

• Molecule 2: Hemagglutinin

Chain F:

17%

72%   24%

L2 F3 F9 P15 G16 L17 I18 N19 G20 W21 Y22 Q25 H26 E29 Q30 G31 T32 G33 I34 T41 A44 Q47 I48 K51 I52 I56 M59 F70 N71 Q72 V73 E74 K75 R76 M79 I84 Y94 L98 L102 D105 K106 H111

V115 K116 R123 R124 E125 I131 D132 E133 F138 E139 L140 C144 N145 D146 S147 C148 M149 G155 T156 Y157 D158 H159 T160 E161 Y162 A163 E164 E165 S166

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	75.64Å 76.80Å 76.38Å 85.60° 82.67° 87.52°	Depositor
Resolution (Å)	42.45 – 2.59 42.46 – 2.59	Depositor EDS
% Data completeness (in resolution range)	90.3 (42.45-2.59) 93.2 (42.46-2.59)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.12 (at 2.58Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.5_2)	Depositor
R, R_{free}	0.216 , 0.255 0.209 , 0.247	Depositor DCC
R_{free} test set	2573 reflections (5.10%)	DCC
Wilson B-factor (Å^2)	49.8	Xtriage
Anisotropy	0.510	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 50.9	EDS
Estimated twinning fraction	0.010 for -l,-k,-h	Xtriage
L-test for twinning [2]	$<|L|> = 0.50$, $<L^2> = 0.33$	Xtriage
Outliers	0 of 50524 reflections	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	11796	wwPDB-VP
Average B, all atoms (Å^2)	76.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.16% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $<|L|>$, $<L^2>$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: SIA, GAL, NAG

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$|Z|$>5	RMSZ	#$|Z|$>5
1	A	0.37	0/2619	0.46	0/3560
1	C	0.27	0/2619	0.43	0/3560
1	E	0.36	0/2619	0.44	0/3560
2	B	0.35	0/1333	0.43	0/1798
2	D	0.24	0/1333	0.42	0/1798
2	F	0.23	0/1333	0.42	0/1798
All	All	0.32	0/11856	0.44	0/16074

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2555	0	2476	85	0
1	C	2555	0	2476	76	0
1	E	2555	0	2474	81	0
2	B	1312	0	1240	67	0
2	D	1312	0	1240	37	0
2	F	1312	0	1240	40	0
3	A	32	0	27	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	E	14	0	13	0	0
5	E	46	0	39	2	0
6	A	15	0	0	24	0
6	B	19	0	0	37	0
6	C	20	0	0	22	0
6	D	13	0	0	3	0
6	E	25	0	0	21	0
6	F	11	0	0	11	0
All	All	11796	0	11225	356	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 16.

The worst 5 of 356 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:C:47:CYS:HB2	6:C:409:HOH:O	1.14	1.32
2:B:106:LYS:HB2	6:B:218:HOH:O	1.15	1.29
2:B:30:GLN:HG2	6:B:207:HOH:O	1.30	1.25
6:A:903:HOH:O	2:D:50:THR:HG21	1.38	1.24
2:B:147:SER:HB3	6:B:215:HOH:O	1.42	1.19

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	322/324 (99%)	304 (94%)	16 (5%)	2 (1%)	30	54
1	C	322/324 (99%)	307 (95%)	15 (5%)	0	100	100
1	E	322/324 (99%)	306 (95%)	15 (5%)	1 (0%)	46	70

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	163/165 (99%)	153 (94%)	8 (5%)	2 (1%)	16	33
2	D	163/165 (99%)	152 (93%)	9 (6%)	2 (1%)	16	33
2	F	163/165 (99%)	153 (94%)	8 (5%)	2 (1%)	16	33
All	All	1455/1467 (99%)	1375 (94%)	71 (5%)	9 (1%)	30	54

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	224	ASN
2	B	145	ASN
2	D	145	ASN
2	F	145	ASN
2	B	33	GLY

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	285/285 (100%)	258 (90%)	27 (10%)	11	19
1	C	285/285 (100%)	260 (91%)	25 (9%)	12	23
1	E	285/285 (100%)	257 (90%)	28 (10%)	10	18
2	B	140/140 (100%)	128 (91%)	12 (9%)	13	24
2	D	140/140 (100%)	129 (92%)	11 (8%)	15	29
2	F	140/140 (100%)	129 (92%)	11 (8%)	15	29
All	All	1275/1275 (100%)	1161 (91%)	114 (9%)	12	22

5 of 114 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	192	LYS
1	C	327	ILE
2	F	30	GLN
1	C	227	ARG

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	274	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 38 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	248	ASN
2	D	25	GLN
2	F	30	GLN
1	C	279	ASN
2	D	30	GLN

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)

5 carbohydrates are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#|Z| > 2$	Counts	RMSZ	$\#|Z| > 2$
3	SIA	A	801	3	16,20,21	2.13	6 (37%)	18,28,31	2.51	5 (27%)
3	GAL	A	802	3	12,12,12	0.48	0	17,17,17	1.48	3 (17%)
5	SIA	E	602	5	16,20,21	2.13	6 (37%)	18,28,31	2.66	3 (16%)
5	GAL	E	603	5	11,11,12	0.61	0	14,15,17	1.70	2 (14%)
5	NAG	E	604	5	15,15,15	0.45	0	17,21,21	0.91	1 (5%)

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

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	SIA	A	801	3	-	0/14/34/38	0/1/1/1
3	GAL	A	802	3	-	0/2/22/22	0/1/1/1
5	SIA	E	602	5	-	0/14/34/38	0/1/1/1
5	GAL	E	603	5	-	0/2/19/22	0/1/1/1
5	NAG	E	604	5	-	0/6/26/26	0/1/1/1

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	801	SIA	C7-C6	-4.49	1.47	1.52
5	E	602	SIA	C7-C6	-4.00	1.47	1.52
5	E	602	SIA	O8-C8	-3.35	1.35	1.43
3	A	801	SIA	O8-C8	-3.20	1.36	1.43
3	A	801	SIA	C6-C5	-2.62	1.48	1.53

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	E	602	SIA	O6-C2-C3	-9.88	90.87	109.86
3	A	801	SIA	O6-C2-C3	-8.75	93.04	109.86
3	A	802	GAL	C4-C3-C2	-3.61	104.05	110.79
3	A	801	SIA	C7-C6-C5	-3.24	109.41	114.32
5	E	603	GAL	O3-C3-C2	-3.02	104.54	110.00

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	801	SIA	1	0
5	E	602	SIA	2	0

5.6 Ligand geometry (i)

1 ligand is 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	E	601	1	14,14,15	0.28	0	15,19,21	0.53	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	E	601	1	-	0/6/23/26	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers (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, 95th 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^2)	Q<0.9
1	A	324/324 (100%)	0.50	17 (5%)	31	26	38, 70, 111, 147	0
1	C	324/324 (100%)	0.69	29 (8%)	12	9	43, 80, 115, 156	0
1	E	324/324 (100%)	0.42	12 (3%)	45	40	41, 71, 107, 181	0
2	B	165/165 (100%)	0.81	12 (7%)	18	14	35, 61, 106, 280	0
2	D	165/165 (100%)	0.79	13 (7%)	15	12	38, 67, 119, 189	0
2	F	165/165 (100%)	1.16	28 (16%)	2	1	36, 72, 148, 211	0
All	All	1467/1467 (100%)	0.67	111 (7%)	17	13	35, 72, 118, 280	0

The worst 5 of 111 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	166	SER	15.6
2	D	163	ALA	12.6
1	E	4	ILE	11.5
2	B	164	GLU	11.4
2	D	166	SER	10.2

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

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

6.3 Carbohydrates (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å^2)	Q<0.9
5	SIA	E	602	20/21	0.89	0.18	0.36	72,85,96,96	0
3	SIA	A	801	20/21	0.94	0.14	-1.18	90,102,109,111	0
5	GAL	E	603	11/12	0.85	0.18	-	121,134,150,152	0
5	NAG	E	604	15/15	0.66	0.41	-	163,175,183,185	0
3	GAL	A	802	12/12	0.79	0.30	-	132,146,159,163	0

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, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å^2)	Q<0.9
4	NAG	E	601	14/15	0.77	0.32	-	137,145,151,152	0

6.5 Other polymers (i)

There are no such residues in this entry.