



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 22, 2016 – 07:24 PM EDT

PDB ID : 4YEL  
Title : Tailspike protein double mutant D339A/E372A of E. coli bacteriophage HK620  
in complex with hexasaccharide  
Authors : Gohlke, U.; Broecker, N.K.; Heinemann, U.; Seckler, R.; Barbirz, S.  
Deposited on : 2015-02-24  
Resolution : 1.72 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

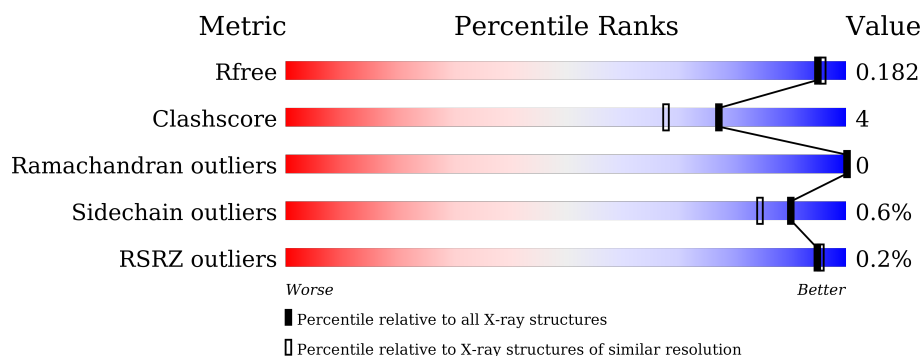
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.72 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	3998 (1.74-1.70)
Clashscore	102246	4425 (1.74-1.70)
Ramachandran outliers	100387	4360 (1.74-1.70)
Sidechain outliers	100360	4360 (1.74-1.70)
RSRZ outliers	91569	4010 (1.74-1.70)

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. 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	596	<div style="display: flex; align-items: center;"> <div style="width: 92%; height: 10px; background: linear-gradient(to right, red, orange, yellow, green);"></div> <div style="margin-left: 10px;">92%</div> <div style="width: 7%; height: 10px; background: yellow;"></div> <div style="margin-left: 10px;">7%</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
7	TRS	A	1007	-	-	-	X
8	FMT	A	1010	-	-	-	X

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
8	FMT	A	1012	-	-	-	X
9	NA	A	1013	-	-	-	X
9	NA	A	1014	-	-	-	X

## 2 Entry composition [i](#)

There are 10 unique types of molecules in this entry. The entry contains 5529 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.

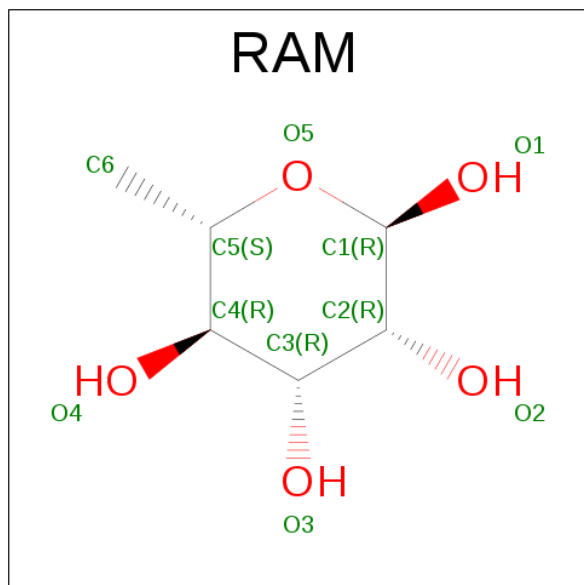
- Molecule 1 is a protein called Tail spike protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	596	4653	2907	807	915	24	0	19	0

There are 2 discrepancies between the modelled and reference sequences:

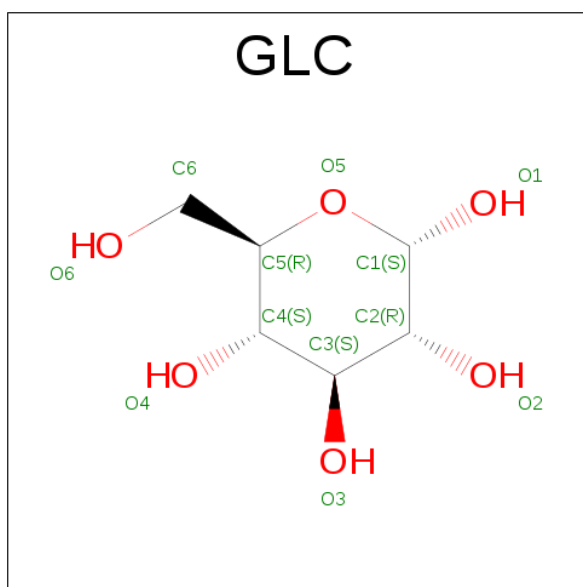
Chain	Residue	Modelled	Actual	Comment	Reference
A	339	ALA	ASP	engineered mutation	UNP Q9AYY6
A	372	ALA	GLU	engineered mutation	UNP Q9AYY6

- Molecule 2 is ALPHA-L-RHAMNOSE (three-letter code: RAM) (formula: C<sub>6</sub>H<sub>12</sub>O<sub>5</sub>).



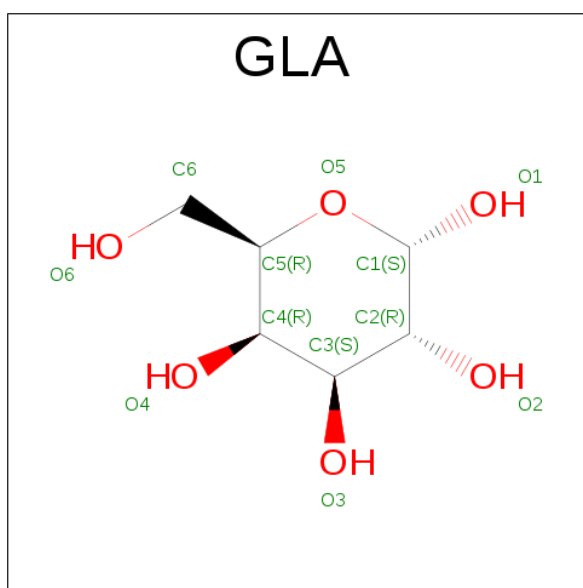
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
2	A	1	10	6	4	0	0

- Molecule 3 is ALPHA-D-GLUCOSE (three-letter code: GLC) (formula: C<sub>6</sub>H<sub>12</sub>O<sub>6</sub>).



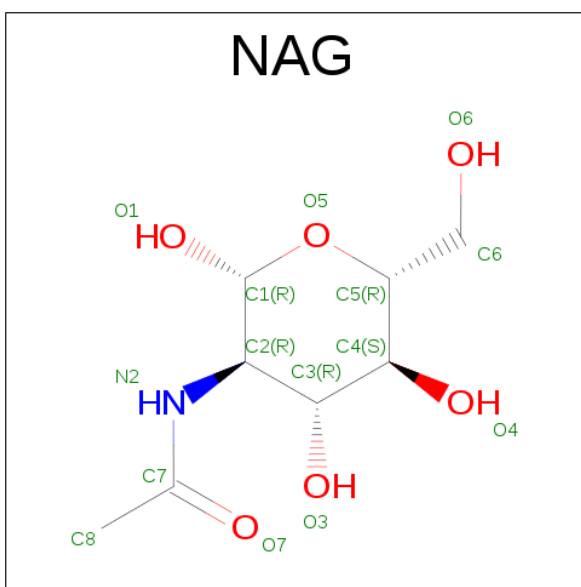
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			11	6	5		
3	A	1	Total	C	O	0	0
			11	6	5		

- Molecule 4 is ALPHA D-GALACTOSE (three-letter code: GLA) (formula:  $C_6H_{12}O_6$ ).



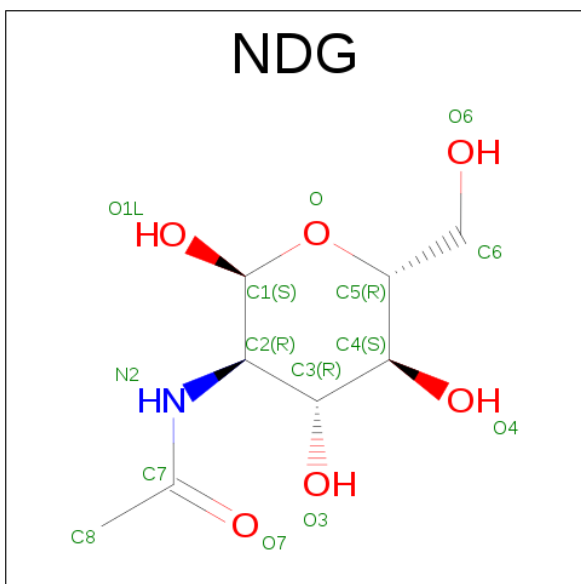
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			11	6	5		

- Molecule 5 is N-ACETYL-D-GLUCOSAMINE (three-letter code: NAG) (formula:  $C_8H_{15}NO_6$ ).



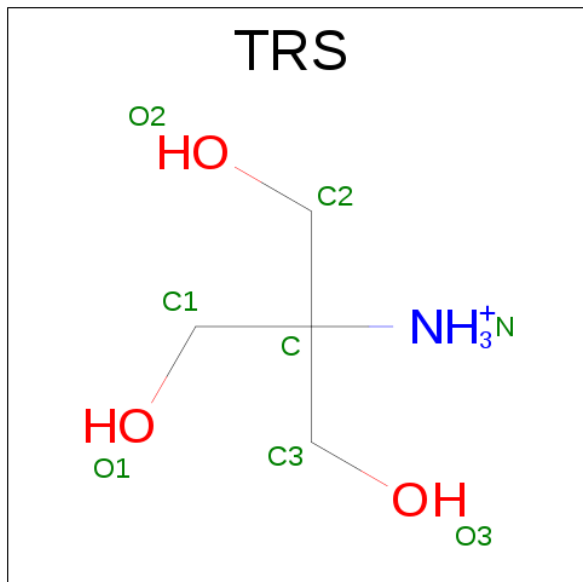
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	1	Total	C	N	O	0	0
			14	8	1	5		
5	A	1	Total	C	N	O	0	1
			15	8	1	6		

- Molecule 6 is 2-(ACETYLAMINO)-2-DEOXY-A-D-GLUCOPYRANOSE (three-letter code: NDG) (formula:  $C_8H_{15}NO_6$ ).



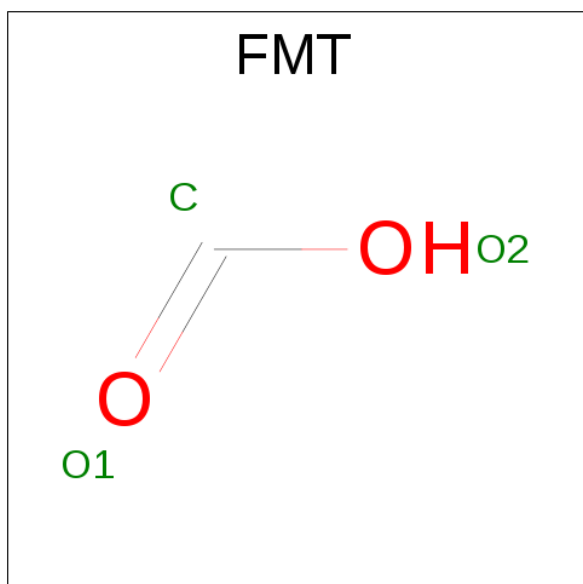
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
6	A	1	Total	C	N	O	0	1
			15	8	1	6		

- Molecule 7 is 2-AMINO-2-HYDROXYMETHYL-PROPANE-1,3-DIOL (three-letter code: TRS) (formula:  $C_4H_{12}NO_3$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
7	A	1	Total	C	N	O	0	0
			8	4	1	3		

- Molecule 8 is FORMIC ACID (three-letter code: FMT) (formula:  $CH_2O_2$ ).



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

*Continued on next page...*

*Continued from previous page...*

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

- Molecule 9 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	A	3	Total	Na	0	0
			3	3		

- Molecule 10 is water.

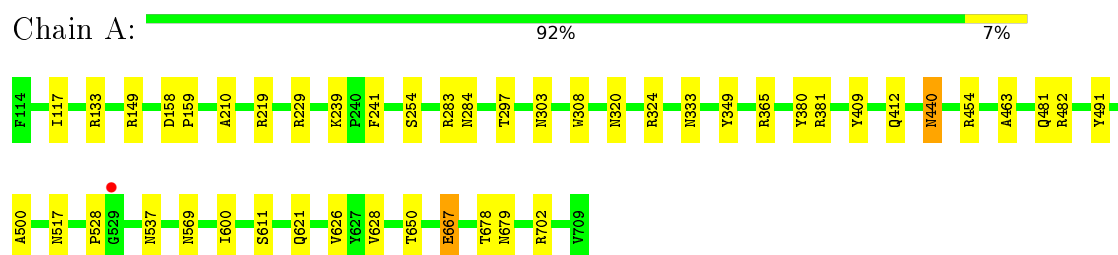
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
10	A	763	Total	O	0	0
			763	763		



### 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: Tail spike protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 3 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	74.36 Å 74.36 Å 174.68 Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	43.67 – 1.72 43.67 – 1.72	Depositor EDS
% Data completeness (in resolution range)	99.9 (43.67-1.72) 99.9 (43.67-1.72)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.55 (at 1.72 Å)	Xtriage
Refinement program	REFMAC	Depositor
R, $R_{free}$	0.135 , 0.171 0.148 , 0.182	Depositor DCC
$R_{free}$ test set	3016 reflections (5.26%)	DCC
Wilson B-factor (Å <sup>2</sup> )	13.3	Xtriage
Anisotropy	0.295	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 36.1	EDS
Estimated twinning fraction	0.041 for -h,-k,l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Outliers	6 of 60325 reflections (0.010%)	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	5529	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	17.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.93% 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: RAM, NAG, GLA, FMT, GLC, NDG, NA, TRS

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	0.95	3/4760 (0.1%)	0.90	6/6485 (0.1%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	667	GLU	CD-OE2	5.92	1.32	1.25
1	A	491	TYR	CE1-CZ	-5.49	1.31	1.38
1	A	380	TYR	CE1-CZ	-5.28	1.31	1.38

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	482	ARG	NE-CZ-NH1	5.80	123.20	120.30
1	A	229	ARG	NE-CZ-NH1	5.55	123.08	120.30
1	A	133	ARG	NE-CZ-NH1	5.51	123.06	120.30
1	A	219	ARG	NE-CZ-NH2	-5.10	117.75	120.30
1	A	454	ARG	CG-CD-NE	5.07	122.44	111.80

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	4653	0	4375	33	0
2	A	10	0	8	0	0
3	A	22	0	19	0	0
4	A	11	0	8	0	0
5	A	29	0	26	1	0
6	A	15	0	13	0	0
7	A	8	0	12	0	0
8	A	15	0	6	0	0
9	A	3	0	0	0	0
10	A	763	0	0	18	3
All	All	5529	0	4467	34	3

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 4.

The worst 5 of 34 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:303[B]:ASN:OD1	10:A:1851:HOH:O	1.62	1.15
1:A:528:PRO:O	10:A:1861:HOH:O	1.90	0.89
1:A:678[A]:THR:OG1	10:A:1101:HOH:O	1.86	0.87
1:A:381[A]:ARG:HD2	10:A:1274:HOH:O	1.86	0.76
1:A:381[A]:ARG:CD	10:A:1274:HOH:O	2.40	0.66

All (3) 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)
10:A:1398:HOH:O	10:A:1398:HOH:O[6_765]	1.85	0.35
10:A:1220:HOH:O	10:A:1327:HOH:O[3_665]	1.96	0.24
10:A:1103:HOH:O	10:A:1313:HOH:O[2_655]	2.11	0.09

## 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	613/596 (103%)	601 (98%)	12 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 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	505/486 (104%)	502 (99%)	3 (1%)	90	84

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	149	ARG
1	A	440	ASN
1	A	517	ASN

Some 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 carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 16 ligands modelled in this entry, 3 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
2	RAM	A	1000	9,3	10,10,11	0.97	0	13,14,16	1.54	3 (23%)
3	GLC	A	1001	2,4	11,11,12	1.33	2 (18%)	15,15,17	1.42	2 (13%)
4	GLA	A	1002	3,5,6	11,11,12	1.40	2 (18%)	15,15,17	2.33	6 (40%)
5	NAG	A	1003	4	14,14,15	0.78	0	15,19,21	1.47	3 (20%)
6	NDG	A	1004[A]	3,4	15,15,15	0.83	0	17,21,21	1.20	2 (11%)
5	NAG	A	1005[B]	3,4	15,15,15	0.64	0	17,21,21	0.81	0
3	GLC	A	1006	5,6	11,11,12	0.84	0	15,15,17	1.10	2 (13%)
7	TRS	A	1007	-	7,7,7	0.78	0	9,9,9	1.56	3 (33%)
8	FMT	A	1008	-	0,2,2	0.00	-	0,1,1	0.00	-
8	FMT	A	1009	-	0,2,2	0.00	-	0,1,1	0.00	-
8	FMT	A	1010	-	0,2,2	0.00	-	0,1,1	0.00	-
8	FMT	A	1011	-	0,2,2	0.00	-	0,1,1	0.00	-
8	FMT	A	1012	-	0,2,2	0.00	-	0,1,1	0.00	-

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	RAM	A	1000	9,3	-	0/0/17/20	0/1/1/1
3	GLC	A	1001	2,4	-	0/2/19/22	0/1/1/1
4	GLA	A	1002	3,5,6	-	0/2/19/22	0/1/1/1
5	NAG	A	1003	4	-	0/6/23/26	0/1/1/1
6	NDG	A	1004[A]	3,4	-	0/6/26/26	0/1/1/1
5	NAG	A	1005[B]	3,4	-	0/6/26/26	0/1/1/1
3	GLC	A	1006	5,6	-	0/2/19/22	0/1/1/1
7	TRS	A	1007	-	-	0/9/9/9	0/0/0/0
8	FMT	A	1008	-	-	0/0/0/0	0/0/0/0
8	FMT	A	1009	-	-	0/0/0/0	0/0/0/0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
8	FMT	A	1010	-	-	0/0/0/0	0/0/0/0
8	FMT	A	1011	-	-	0/0/0/0	0/0/0/0
8	FMT	A	1012	-	-	0/0/0/0	0/0/0/0

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	1002	GLA	O5-C1	-2.73	1.39	1.43
3	A	1001	GLC	O2-C2	-2.29	1.38	1.43
3	A	1001	GLC	O5-C1	-2.07	1.40	1.43
4	A	1002	GLA	C2-C3	2.59	1.56	1.52

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	1002	GLA	C1-C2-C3	-4.19	104.48	109.55
3	A	1001	GLC	C2-C3-C4	-3.20	105.46	111.05
4	A	1002	GLA	O3-C3-C2	-2.99	104.53	110.01
4	A	1002	GLA	C1-O5-C5	-2.92	107.84	112.14
3	A	1001	GLC	O5-C1-C2	-2.84	106.35	110.89

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	1005[B]	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	596/596 (100%)	-0.66	1 (0%) 95 95	8, 14, 24, 48	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	529	GLY	4.2

### 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](#)

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(Å <sup>2</sup> )	Q<0.9
8	FMT	A	1010	3/3	0.92	0.18	8.80	34,34,37,37	0
8	FMT	A	1012	3/3	0.83	0.13	8.33	35,35,36,41	0
9	NA	A	1013	1/1	0.97	0.19	6.02	36,36,36,36	0
9	NA	A	1014	1/1	0.99	0.13	2.77	18,18,18,18	0

*Continued on next page...*



*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
7	TRS	A	1007	8/8	0.95	0.08	2.44	13,15,17,19	0
8	FMT	A	1011	3/3	0.88	0.14	1.71	24,24,29,31	0
5	NAG	A	1003	14/15	0.96	0.06	0.44	13,15,26,27	0
5	NAG	A	1005[B]	15/15	0.98	0.05	-0.14	12,12,13,13	15
2	RAM	A	1000	10/11	0.97	0.06	-0.29	15,17,20,21	0
4	GLA	A	1002	11/12	0.97	0.05	-0.30	12,13,16,19	0
6	NDG	A	1004[A]	15/15	0.97	0.05	-0.41	12,12,14,15	15
3	GLC	A	1006	11/12	0.98	0.04	-1.24	11,12,13,14	0
8	FMT	A	1009	3/3	0.99	0.03	-2.63	17,17,18,19	0
8	FMT	A	1008	3/3	0.98	0.07	-	21,21,22,25	0
9	NA	A	1015	1/1	0.97	0.13	-	26,26,26,26	0
3	GLC	A	1001	11/12	0.96	0.06	-	14,17,19,20	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.