



Full wwPDB X-ray Structure Validation Report ⓘ

Feb 23, 2016 – 12:27 PM GMT

PDB ID : 5I4D
Title : 1.75 Angstrom Crystal Structure of Superantigen-like Protein, Exotoxin from Staphylococcus aureus, in Complex with Sialyl-LewisX.
Authors : Minasov, G.; Nocadello, S.; Shuvalova, L.; Filippova, E.V.; Halavaty, A.; Dubrovskaya, I.; Bagnoli, F.; Falugi, F.; Bottomley, M.; Grandi, G.; Anderson, W.F.; Center for Structural Genomics of Infectious Diseases (CSGID)
Deposited on : 2016-02-11
Resolution : 1.75 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
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-20026982
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) : rb-20026982

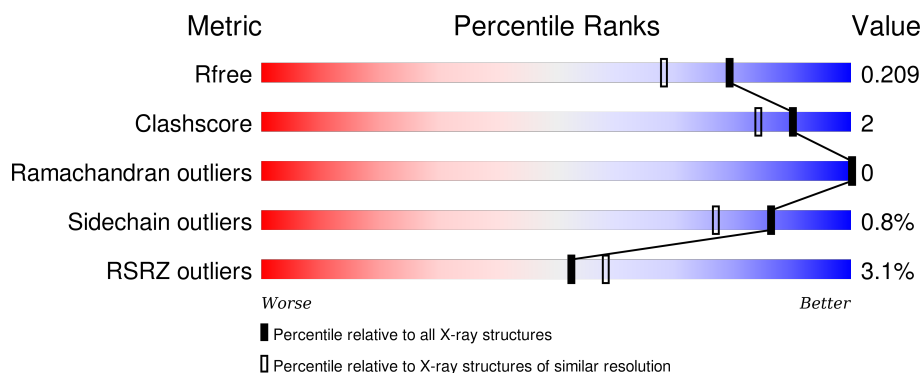
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.75 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	1609 (1.76-1.76)
Clashscore	102246	1730 (1.76-1.76)
Ramachandran outliers	100387	1711 (1.76-1.76)
Sidechain outliers	100360	1711 (1.76-1.76)
RSRZ outliers	91569	1610 (1.76-1.76)

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 $\leq 5\%$. 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	343	<div> <div></div> <div>52%</div> <div>•</div> <div>44%</div> </div>
1	B	343	<div> <div>3%</div> <div>54%</div> <div>•</div> <div>44%</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
6	PGE	A	405	-	-	-	X

2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 4379 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 Superantigen-like protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	192	Total	C	N	O	S	0	25	0
			1787	1154	298	326	9			
1	B	193	Total	C	N	O	S	0	19	0
			1746	1122	294	319	11			

There are 34 discrepancies between the modelled and reference sequences:

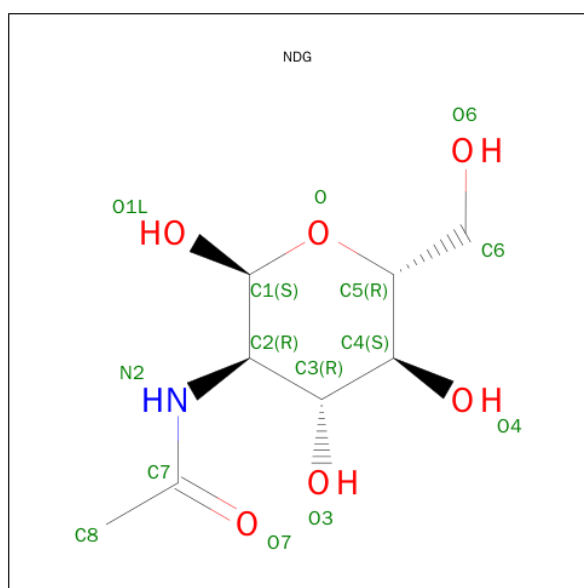
Chain	Residue	Modelled	Actual	Comment	Reference
A	14	MET	-	initiating methionine	UNP A0A0D6GBT0
A	15	GLY	-	expression tag	UNP A0A0D6GBT0
A	16	SER	-	expression tag	UNP A0A0D6GBT0
A	17	SER	-	expression tag	UNP A0A0D6GBT0
A	18	HIS	-	expression tag	UNP A0A0D6GBT0
A	19	HIS	-	expression tag	UNP A0A0D6GBT0
A	20	HIS	-	expression tag	UNP A0A0D6GBT0
A	21	HIS	-	expression tag	UNP A0A0D6GBT0
A	22	HIS	-	expression tag	UNP A0A0D6GBT0
A	23	HIS	-	expression tag	UNP A0A0D6GBT0
A	24	GLU	-	expression tag	UNP A0A0D6GBT0
A	25	ASN	-	expression tag	UNP A0A0D6GBT0
A	26	LEU	-	expression tag	UNP A0A0D6GBT0
A	27	TYR	-	expression tag	UNP A0A0D6GBT0
A	28	PHE	-	expression tag	UNP A0A0D6GBT0
A	29	GLN	-	expression tag	UNP A0A0D6GBT0
A	30	GLY	-	expression tag	UNP A0A0D6GBT0
B	14	MET	-	initiating methionine	UNP A0A0D6GBT0
B	15	GLY	-	expression tag	UNP A0A0D6GBT0
B	16	SER	-	expression tag	UNP A0A0D6GBT0
B	17	SER	-	expression tag	UNP A0A0D6GBT0
B	18	HIS	-	expression tag	UNP A0A0D6GBT0
B	19	HIS	-	expression tag	UNP A0A0D6GBT0
B	20	HIS	-	expression tag	UNP A0A0D6GBT0
B	21	HIS	-	expression tag	UNP A0A0D6GBT0

Continued on next page...

Continued from previous page...

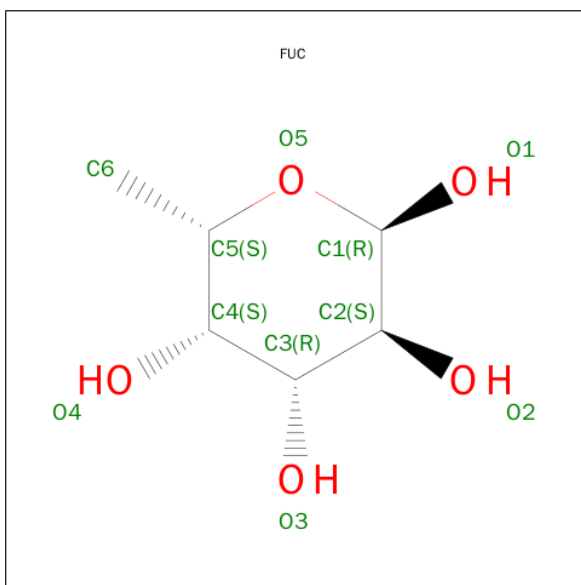
Chain	Residue	Modelled	Actual	Comment	Reference
B	22	HIS	-	expression tag	UNP A0A0D6GBT0
B	23	HIS	-	expression tag	UNP A0A0D6GBT0
B	24	GLU	-	expression tag	UNP A0A0D6GBT0
B	25	ASN	-	expression tag	UNP A0A0D6GBT0
B	26	LEU	-	expression tag	UNP A0A0D6GBT0
B	27	TYR	-	expression tag	UNP A0A0D6GBT0
B	28	PHE	-	expression tag	UNP A0A0D6GBT0
B	29	GLN	-	expression tag	UNP A0A0D6GBT0
B	30	GLY	-	expression tag	UNP A0A0D6GBT0

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



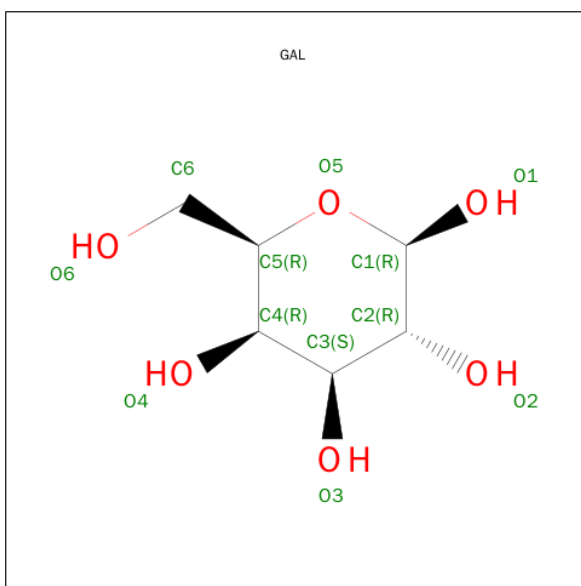
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			15	8	1	6		
2	B	1	Total	C	N	O	0	1
			15	8	1	6		

- Molecule 3 is ALPHA-L-FUCOSE (three-letter code: FUC) (formula: $C_6H_{12}O_5$).



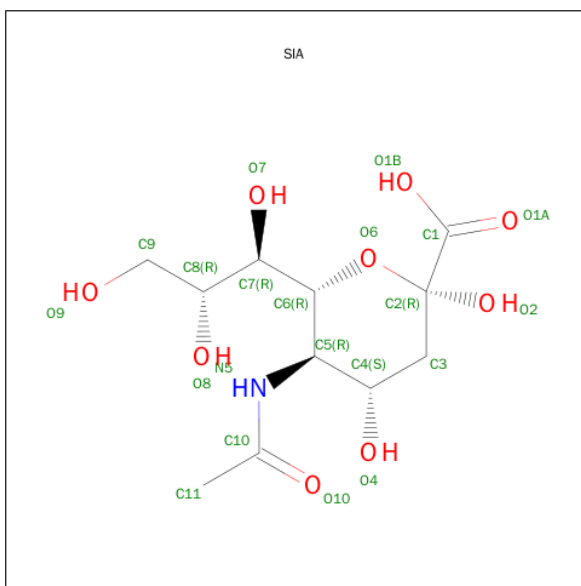
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			10	6	4		
3	B	1	Total	C	O	0	1
			11	7	4		

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



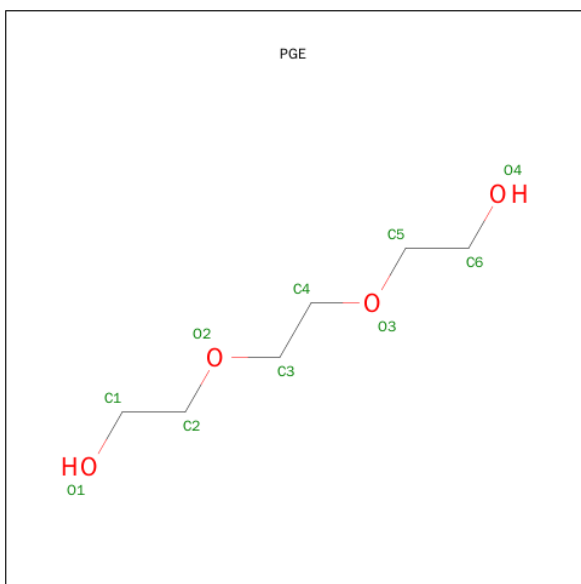
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			11	6	5		
4	B	1	Total	C	O	0	1
			20	11	9		

- Molecule 5 is O-SIALIC ACID (three-letter code: SIA) (formula: $C_{11}H_{19}NO_9$).



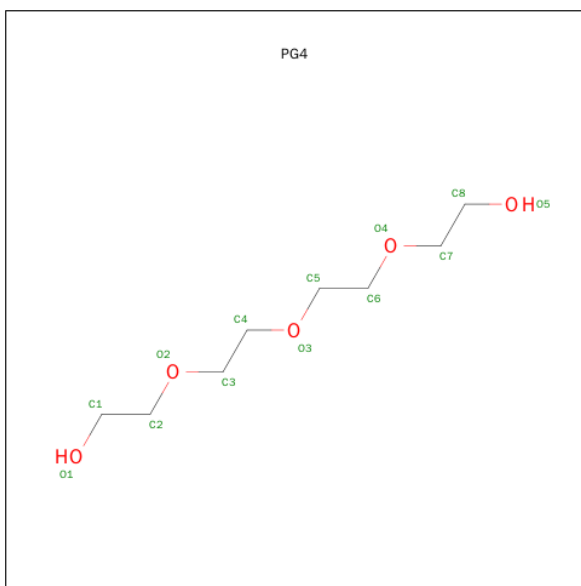
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	1	Total	C	N	O	0	0
			20	11	1	8		
5	B	1	Total	C	N	O	0	0
			20	11	1	8		

- Molecule 6 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula: $C_6H_{14}O_4$).



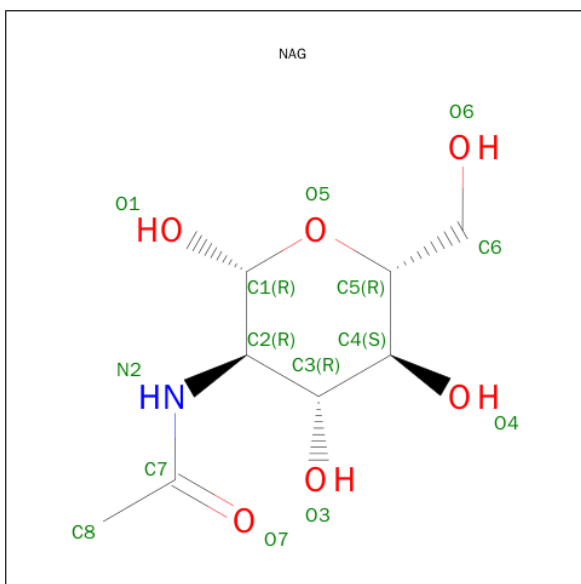
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	A	1	Total	C	O	0	0
			10	6	4		

- Molecule 7 is TETRAETHYLENE GLYCOL (three-letter code: PG4) (formula: $C_8H_{18}O_5$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
7	A	1	Total	C	O	0	0
			13	8	5		
7	B	1	Total	C	O	0	0
			13	8	5		

- Molecule 8 is N-ACETYL-D-GLUCOSAMINE (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).

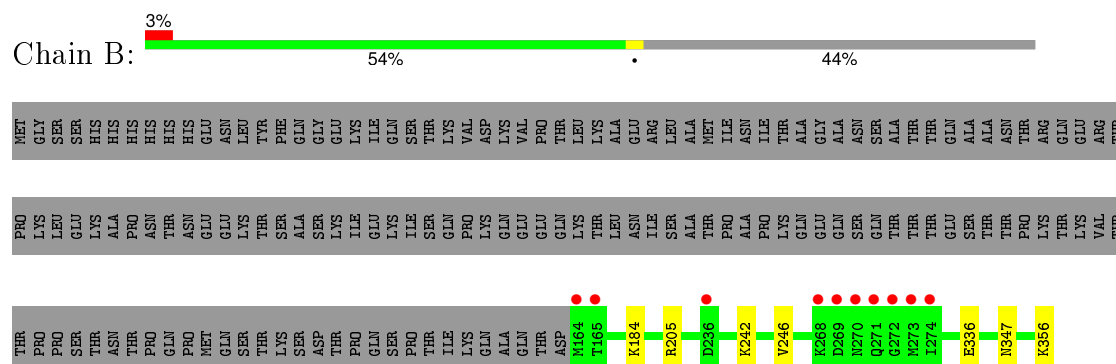


Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
8	B	1	Total	C	N	O	0	1
			15	8	1	6		

- Molecule 9 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	A	326	Total 349	O 349	0	33
9	B	315	Total 324	O 324	0	22

- Molecule 1: Superantigen-like protein



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α , β , γ	85.97Å 166.88Å 34.14Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.94 – 1.75 29.94 – 1.75	Depositor EDS
% Data completeness (in resolution range)	99.6 (29.94-1.75) 99.7 (29.94-1.75)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	0.09	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.11 (at 1.75Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
R, R_{free}	0.161 , 0.202 0.170 , 0.209	Depositor DCC
R_{free} test set	2563 reflections (5.36%)	DCC
Wilson B-factor (Å ²)	18.6	Xtriage
Anisotropy	0.170	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 51.3	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	1 of 50393 reflections (0.002%)	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	4379	wwPDB-VP
Average B, all atoms (Å ²)	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.15% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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 [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: PGE, NAG, SIA, NDG, PG4, GAL, FUC

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	0.57	0/1838	0.74	1/2455 (0.0%)
1	B	0.57	0/1785	0.75	0/2379
All	All	0.57	0/3623	0.75	1/4834 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	235	GLU	CB-CA-C	-5.26	99.87	110.40

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	1787	0	1837	14	0
1	B	1746	0	1805	3	0
2	A	15	0	13	0	0
2	B	15	0	13	0	0
3	A	10	0	10	0	0
3	B	11	0	4	0	0
4	A	11	0	9	0	0
4	B	20	0	18	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	A	20	0	17	0	0
5	B	20	0	17	0	0
6	A	10	0	14	0	0
7	A	13	0	18	0	0
7	B	13	0	18	0	0
8	B	15	0	13	0	0
9	A	349	0	0	3	0
9	B	324	0	0	2	0
All	All	4379	0	3806	17	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 2.

All (17) 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:260[B]:LYS:HA	1:A:260[B]:LYS:HE3	1.67	0.74
1:A:210:ILE:CD1	1:A:230[B]:VAL:HG11	2.22	0.69
1:A:260[B]:LYS:NZ	1:A:282[B]:MET:HG3	2.10	0.66
1:A:260[B]:LYS:HZ1	1:A:282[B]:MET:HG3	1.62	0.64
1:A:210:ILE:HD11	1:A:230[B]:VAL:HG11	1.80	0.62
1:A:260[B]:LYS:HZ3	1:A:282[B]:MET:CG	2.18	0.57
1:A:240[B]:GLN:HA	1:A:240[B]:GLN:OE1	2.05	0.56
1:B:246:VAL:HG11	1:B:336[A]:GLU:HG3	1.89	0.54
1:A:260[B]:LYS:NZ	1:A:282[B]:MET:CG	2.70	0.54
1:A:210:ILE:HD13	1:A:230[B]:VAL:HG11	1.88	0.52
1:A:269:ASP:HA	1:A:311:MET:SD	2.54	0.48
1:B:347[A]:ASN:ND2	9:B:501:HOH:O	2.46	0.47
1:A:167:LYS:NZ	9:A:503:HOH:O	2.44	0.47
1:A:255:LYS:NZ	9:A:506:HOH:O	2.49	0.45
1:B:184[A]:LYS:HB3	9:B:650[A]:HOH:O	2.17	0.43
1:A:193:TRP:CZ2	1:A:198[B]:PHE:CE1	3.09	0.40
1:A:276[B]:ARG:NH2	9:A:510:HOH:O	2.54	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	215/343 (63%)	208 (97%)	7 (3%)	0	100	100
1	B	210/343 (61%)	205 (98%)	5 (2%)	0	100	100
All	All	425/686 (62%)	413 (97%)	12 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	202/313 (64%)	202 (100%)	0	100	100
1	B	197/313 (63%)	193 (98%)	4 (2%)	63	39
All	All	399/626 (64%)	395 (99%)	4 (1%)	86	69

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	205	ARG
1	B	242[A]	LYS
1	B	242[B]	LYS
1	B	356	LYS

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

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 ⓘ

14 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	$\# Z > 2$
2	NDG	A	401	3,4	15,15,15	0.49	0	17,21,21	0.67	0
3	FUC	A	402	2	10,10,11	0.49	0	13,14,16	0.84	1 (7%)
4	GAL	A	403	2,5	11,11,12	0.42	0	15,15,17	0.99	0
5	SIA	A	404	4	17,20,21	0.25	0	18,28,31	0.57	0
6	PGE	A	405	-	9,9,9	0.59	0	8,8,8	0.46	0
7	PG4	A	406	-	12,12,12	0.50	0	11,11,11	0.31	0
2	NDG	B	401[A]	3,4	15,15,15	0.42	0	17,21,21	0.78	0
8	NAG	B	402[B]	3,4	15,15,15	0.41	0	17,21,21	0.67	0
3	FUC	B	403[A]	2	10,10,11	0.41	0	13,14,16	0.87	0
3	FUC	B	403[B]	8	10,10,11	0.42	0	13,14,16	0.89	0
4	GAL	B	404[A]	2	11,11,12	0.51	0	15,15,17	1.00	1 (6%)
4	GAL	B	404[B]	8	11,11,12	0.51	0	15,15,17	0.99	0
5	SIA	B	405	4	17,20,21	0.46	0	18,28,31	0.74	0
7	PG4	B	406	-	12,12,12	0.52	0	11,11,11	0.34	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
2	NDG	A	401	3,4	-	0/6/26/26	0/1/1/1
3	FUC	A	402	2	-	0/0/17/20	0/1/1/1
4	GAL	A	403	2,5	-	0/2/19/22	0/1/1/1
5	SIA	A	404	4	-	0/14/34/38	0/1/1/1
6	PGE	A	405	-	-	0/7/7/7	0/0/0/0
7	PG4	A	406	-	-	0/10/10/10	0/0/0/0
2	NDG	B	401[A]	3,4	-	0/6/26/26	0/1/1/1
8	NAG	B	402[B]	3,4	-	0/6/26/26	0/1/1/1
3	FUC	B	403[A]	2	-	0/0/17/20	0/1/1/1
3	FUC	B	403[B]	8	-	0/0/17/20	0/1/1/1
4	GAL	B	404[A]	2	-	0/2/19/22	0/1/1/1
4	GAL	B	404[B]	8	-	0/2/19/22	0/1/1/1
5	SIA	B	405	4	-	0/14/34/38	0/1/1/1
7	PG4	B	406	-	-	0/10/10/10	0/0/0/0

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	B	404[A]	GAL	C1-O5-C5	2.03	115.12	112.14
3	A	402	FUC	C1-C2-C3	2.28	112.31	109.55

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	192/343 (55%)	-0.10	2 (1%) 84 89	11, 19, 37, 74	0
1	B	193/343 (56%)	0.13	10 (5%) 31 36	11, 20, 43, 62	0
All	All	385/686 (56%)	0.02	12 (3%) 52 58	11, 20, 39, 74	0

All (12) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	164	MET	7.7
1	A	165	THR	7.1
1	B	273	MET	6.6
1	B	165	THR	5.6
1	B	271	GLN	4.9
1	B	270	ASN	4.1
1	B	269	ASP	3.5
1	A	166	PRO	3.4
1	B	272	GLY	2.9
1	B	236	ASP	2.5
1	B	268[A]	LYS	2.4
1	B	274	ILE	2.3

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, 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
6	PGE	A	405	10/10	0.69	0.37	5.27	41,54,68,71	0
7	PG4	A	406	13/13	0.83	0.15	1.24	48,50,60,60	0
7	PG4	B	406	13/13	0.90	0.13	0.59	37,42,46,49	0
2	NDG	A	401	15/15	0.94	0.08	-0.31	15,18,22,22	0
5	SIA	A	404	20/21	0.97	0.08	-0.45	12,14,16,17	0
2	NDG	B	401[A]	15/15	0.95	0.09	-0.48	15,17,17,18	15
4	GAL	B	404[B]	11/12	0.97	0.06	-0.72	14,14,15,15	9
4	GAL	B	404[A]	11/12	0.97	0.06	-0.72	13,14,15,15	9
4	GAL	A	403	11/12	0.98	0.06	-0.82	13,14,15,15	0
5	SIA	B	405	20/21	0.97	0.06	-1.11	13,13,17,19	0
3	FUC	A	402	10/11	0.97	0.05	-1.59	15,17,19,19	0
3	FUC	B	403[A]	10/11	0.96	0.06	-	18,19,20,22	1
3	FUC	B	403[B]	10/11	0.96	0.06	-	18,19,20,22	1
8	NAG	B	402[B]	15/15	0.94	0.10	-	15,18,22,24	15

6.5 Other polymers [i](#)

There are no such residues in this entry.