

wwPDB X-ray Structure Validation Summary Report (i)

Nov 13, 2023 – 04:50 PM JST

5XL7
The structure of hemagglutinin Q226L mutant from an avian-origin H4N6
influenza virus in complex with human receptor analog LSTc
Song, H.; Qi, J.; Gao, G.F.
2017-05-10
2.10 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 A user guide is available at https://www.wwpdb.org/validation/2017/XrayValidationReportHelp with specific help available everywhere you see the (i) 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 (1)) were used in the production of this report:

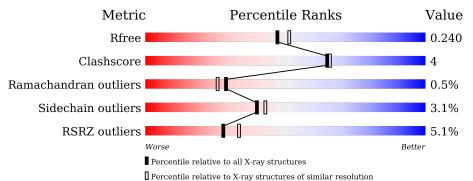
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA,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

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure: $X\text{-}RAY \, DIFFRACTION$

The reported resolution of this entry is 2.10 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	$\begin{array}{l} \textbf{Whole archive} \\ (\#\textbf{Entries}) \end{array}$	${f Similar\ resolution}\ (\#{ m Entries,\ resolution\ range}({ m \AA}))$
R _{free}	130704	5197(2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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 >=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 <=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	А	327	<u>2%</u>	88%	8% ••	
-	11	021	11%	0070	070 •••	
1	В	327		83%	13% ••	
2	С	176	3%	90%	7% •	
2	D	176	.% •	89%	7% ••	
3	Е	3	33%	33%	33%	
3	F	3	33%	67'	%	



2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 8439 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	А	319	Total 2458	C 1541	N 436	0 469	S 12	0	0	0
1	В	319	Total 2458	C 1541	N 436	O 469	S 12	0	0	0

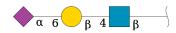
There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
А	223	LEU	GLN	engineered mutation	UNP A3KF09
В	223	LEU	GLN	engineered mutation	UNP A3KF09

• Molecule 2 is a protein called Hemagglutinin.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
9	2 C 172	Total	С	Ν	0	S	0	0	0	
		172	1404	871	250	279	4	0	0	0
0	Л	172	Total	С	Ν	0	S	0	0	0
	2 D		1404	871	250	279	4	0	0	U

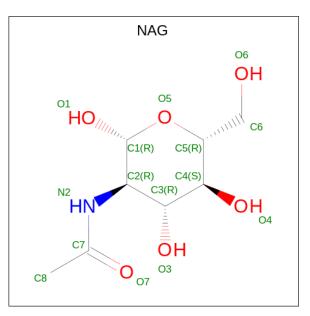
• Molecule 3 is an oligosaccharide called N-acetyl-alpha-neuraminic acid-(2-6)-beta-D-galacto pyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose.



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
3	Е	3	Total C N O 46 25 2 19	0	0	0
3	F	3	Total C N O 46 25 2 19	0	0	0



• Molecule 4 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



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

• Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	А	173	Total O 173 173	0	0
5	С	165	Total O 165 165	0	0
5	В	90	Total O 90 90	0	0
5	D	139	Total O 139 139	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

 \bullet Molecule 3: N-acetyl-alpha-neuraminic acid-(2-6)-beta-D-galactopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose



Chain E:	33%	33%	33%	
NAG1 GAL2 SIA3				
	: N-acetyl-alpha-)-glucopyranose	neuraminic acid-(2-6)-be	ta-D-galactopyranos	se-(1-4)-2-acetamido-2

Chain F:	33%	67%
NAG1 GAL2 SIA3		



4 Data and refinement statistics (i)

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants	100.94Å 100.94Å 688.15Å	Depositor
a, b, c, α , β , γ	90.00° 90.00° 120.00°	Depositor
Resolution (Å)	42.84 - 2.10	Depositor
Resolution (A)	42.84 - 2.10	EDS
% Data completeness	99.0 (42.84-2.10)	Depositor
(in resolution range)	92.9 (42.84-2.10)	EDS
R _{merge}	(Not available)	Depositor
R _{sym}	(Not available)	Depositor
$< I/\sigma(I) > 1$	$5.22 (at 2.10 \text{\AA})$	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
D D	0.206 , 0.240	Depositor
R, R_{free}	0.209 , 0.240	DCC
R_{free} test set	4006 reflections $(5.02%)$	wwPDB-VP
Wilson B-factor $(Å^2)$	31.5	Xtriage
Anisotropy	0.536	Xtriage
Bulk solvent $k_{sol}(e/Å^3), B_{sol}(Å^2)$	0.36 , 52.9	EDS
L-test for twinning ²	$ \langle L \rangle = 0.50, \langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	8439	wwPDB-VP
Average B, all atoms $(Å^2)$	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The analyses of the Patterson function reveals a significant off-origin peak that is 46.72 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.1002e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.

²Theoretical 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.



¹Intensities estimated from amplitudes.

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, NAG, GAL

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	Bo	ond angles
IVIOI	Chain	RMSZ	# Z > 5	RMSZ	# Z > 5
1	А	0.31	0/2512	0.49	0/3420
1	В	0.29	0/2512	0.52	2/3420~(0.1%)
2	С	0.36	0/1428	0.52	0/1922
2	D	0.33	0/1428	0.51	0/1922
All	All	0.32	0/7880	0.51	2/10684~(0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$\mathbf{Observed}(^{o})$	$Ideal(^{o})$
1	В	124	ASN	N-CA-C	5.37	125.49	111.00
1	В	124	ASN	CB-CA-C	-5.12	100.17	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	А	2458	0	2411	18	0
1	В	2458	0	2411	26	0
2	С	1404	0	1320	9	0
2	D	1404	0	1320	10	0
3	Е	46	0	40	1	0



Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	F	46	0	40	0	0
4	А	28	0	26	2	0
4	В	28	0	26	5	0
5	А	173	0	0	4	1
5	В	90	0	0	2	0
5	С	165	0	0	3	1
5	D	139	0	0	1	0
All	All	8439	0	7594	64	1

Continued from previous page...

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 64 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:B:162:ASN:O	4:B:601:NAG:O7	1.85	0.94
2:C:369:GLN:OE1	5:C:701:HOH:O	1.90	0.90
1:A:244:THR:HG22	1:A:246:GLY:H	1.39	0.87
1:B:244:THR:HG22	1:B:246:GLY:H	1.38	0.87
1:A:166:ILE:HA	1:A:239:LEU:HB3	1.67	0.77

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)
5:A:803:HOH:O	5:C:862:HOH:O[2_10125]	2.16	0.04

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	Perce	ntiles
1	А	317/327~(97%)	304 (96%)	11 (4%)	2(1%)	25	21
1	В	317/327~(97%)	301 (95%)	15 (5%)	1 (0%)	41	41
2	С	170/176~(97%)	161 (95%)	9~(5%)	0	100	100
2	D	170/176~(97%)	159~(94%)	9~(5%)	2(1%)	13	8
All	All	974/1006~(97%)	925~(95%)	44 (4%)	5(0%)	29	26

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	А	125	THR
2	D	384	GLU
1	В	193	LYS
1	А	193	LYS
2	D	383	ILE

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	Percei	ntiles
1	А	275/282~(98%)	265~(96%)	10 (4%)	35	36
1	В	275/282~(98%)	262~(95%)	13~(5%)	26	25
2	С	147/150~(98%)	146 (99%)	1 (1%)	84	88
2	D	147/150~(98%)	145~(99%)	2(1%)	67	73
All	All	844/864 (98%)	818 (97%)	26 (3%)	40	43

5 of 26 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	В	139	ASN
1	В	183	SER
2	D	386	THR
1	В	160	LEU
1	В	189	THR



Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	В	74	HIS

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)

6 monosaccharides 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	Iol Type Chain Re		Res	5 Link	Bo	ond leng	ths	Bond angles		
MIOI	Type	Ullaili	nes		Counts	RMSZ	# Z >2	Counts	RMSZ	# Z >2
3	NAG	Е	1	3	$15,\!15,\!15$	0.57	0	21,21,21	0.96	1 (4%)
3	GAL	Е	2	3	11,11,12	0.54	0	$15,\!15,\!17$	0.87	0
3	SIA	Е	3	3	20,20,21	0.67	0	24,28,31	3.43	7 (29%)
3	NAG	F	1	3	15,15,15	0.48	0	21,21,21	1.55	1 (4%)
3	GAL	F	2	3	11,11,12	0.68	0	$15,\!15,\!17$	0.95	0
3	SIA	F	3	3	20,20,21	0.70	0	$24,\!28,\!31$	3.48	7 (29%)

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	Е	1	3	-	2/6/26/26	0/1/1/1



Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	GAL	Ε	2	3	-	0/2/19/22	0/1/1/1
3	SIA	Е	3	3	-	3/18/34/38	0/1/1/1
3	NAG	F	1	3	-	4/6/26/26	0/1/1/1
3	GAL	F	2	3	-	0/2/19/22	0/1/1/1
3	SIA	F	3	3	-	3/18/34/38	0/1/1/1

Continued from previous page...

There are no bond length outliers.

The worst 5 of 16 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$Observed(^{o})$	$Ideal(^{o})$
3	F	3	SIA	O6-C2-C3	-13.52	91.86	110.46
3	Е	3	SIA	O6-C2-C3	-13.04	92.51	110.46
3	F	3	SIA	O6-C2-C1	8.10	123.59	107.70
3	Е	3	SIA	O6-C2-C1	7.82	123.03	107.70
3	F	1	NAG	O5-C1-C2	5.77	115.32	109.52

There are no chirality outliers.

5 of 12 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	Ε	1	NAG	C8-C7-N2-C2
3	Е	1	NAG	O7-C7-N2-C2
3	Е	3	SIA	O1A-C1-C2-O6
3	F	1	NAG	C8-C7-N2-C2
3	F	1	NAG	O7-C7-N2-C2

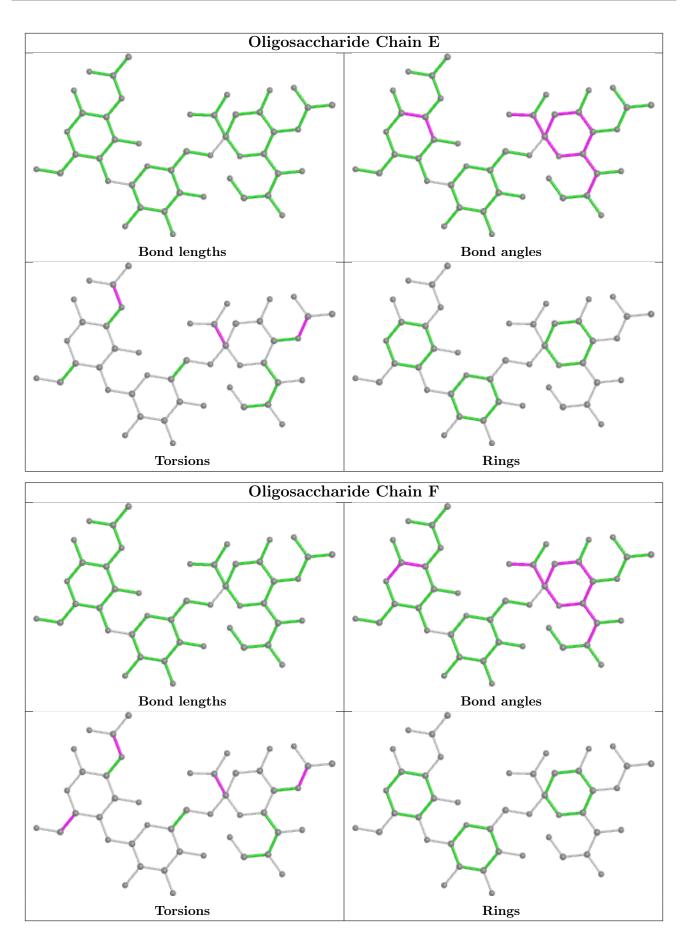
There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	Ε	3	SIA	1	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for oligosaccharide.







5.6 Ligand geometry (i)

4 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	ol Type Chain Res Linl		Link	Bo	ond leng	\mathbf{ths}	Bond angles			
	Type	Chain	nes		Counts	RMSZ	# Z >2	Counts	RMSZ	# Z > 2
4	NAG	А	601	1	14,14,15	0.29	0	17,19,21	0.62	0
4	NAG	В	602	1	14,14,15	0.56	0	$17,\!19,\!21$	1.18	1 (5%)
4	NAG	А	602	1	14,14,15	0.49	0	17,19,21	1.01	1 (5%)
4	NAG	В	601	1	14,14,15	0.30	0	17,19,21	0.61	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	А	601	1	-	3/6/23/26	0/1/1/1
4	NAG	В	602	1	-	0/6/23/26	0/1/1/1
4	NAG	А	602	1	-	0/6/23/26	0/1/1/1
4	NAG	В	601	1	-	2/6/23/26	0/1/1/1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$\mathbf{Observed}(^{o})$	$Ideal(^{o})$
4	В	602	NAG	C1-O5-C5	4.38	118.12	112.19
4	А	602	NAG	C1-O5-C5	3.24	116.58	112.19

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	А	601	NAG	C8-C7-N2-C2
4	А	601	NAG	O7-C7-N2-C2



Continued from previous page...

Mol	Chain	Res	Type	Atoms
4	В	601	NAG	C8-C7-N2-C2
4	В	601	NAG	O7-C7-N2-C2
4	А	601	NAG	C4-C5-C6-O6

There are no ring outliers.

3 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	А	601	NAG	2	0
4	В	602	NAG	1	0
4	В	601	NAG	4	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^{th} 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	$\mathbf{OWAB}(\mathrm{\AA}^2)$	Q < 0.9
1	А	319/327~(97%)	-0.01	7 (2%) 62 66	19, 39, 61, 90	0
1	В	319/327~(97%)	0.63	36 (11%) 5 6	24, 62, 98, 118	0
2	С	172/176~(97%)	0.15	5 (2%) 51 57	21, 29, 47, 108	0
2	D	172/176~(97%)	0.04	2 (1%) 79 82	23, 33, 60, 118	0
All	All	982/1006~(97%)	0.24	50 (5%) 28 33	19, 40, 88, 118	0

The worst 5 of 50 RSRZ outliers are listed below:

Mol	Chain	\mathbf{Res}	Type	RSRZ
1	В	157	ALA	8.2
1	В	158	TYR	5.7
2	D	383	ILE	5.2
1	В	192	TYR	5.1
1	В	151	VAL	5.0

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)

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^{th} 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.

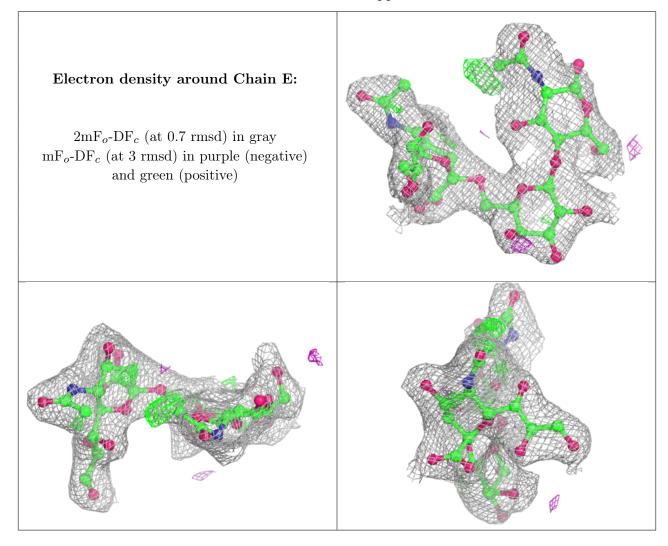
Mol	Type	Chain	Res	Atoms	RSCC	RSR	$B-factors(Å^2)$	Q < 0.9
3	NAG	Ε	1	15/15	0.64	0.18	$92,\!98,\!102,\!102$	0
3	NAG	F	1	15/15	0.65	0.39	146,158,166,168	0



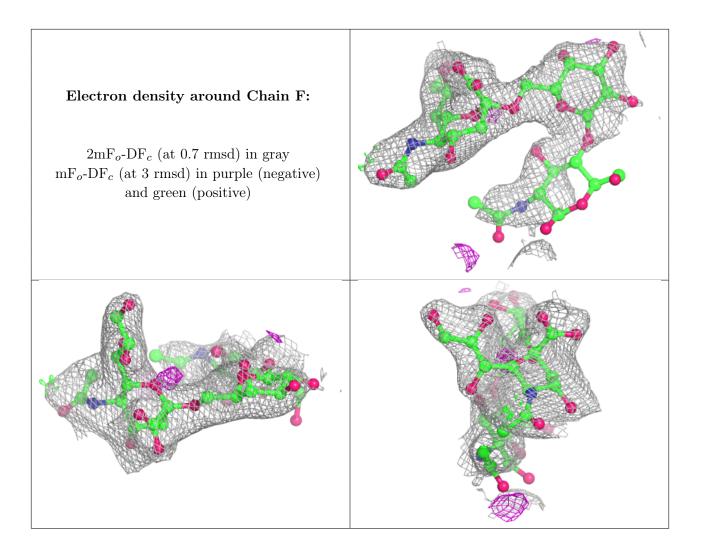
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å ²)	Q<0.9
3	SIA	F	3	20/21	0.86	0.20	72,77,82,83	0
3	GAL	F	2	11/12	0.88	0.28	90,99,108,109	0
3	SIA	Е	3	20/21	0.89	0.13	36,47,51,52	0
3	GAL	Е	2	11/12	0.92	0.15	$53,\!66,\!76,\!79$	0

Continued from previous page...

The following is a graphical depiction of the model fit to experimental electron density for oligosaccharide. Each fit is shown from different orientation to approximate a three-dimensional view.







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^{th} 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	$B-factors(Å^2)$	Q < 0.9
4	NAG	В	601	14/15	0.69	0.34	101,108,110,114	0
4	NAG	А	601	14/15	0.73	0.23	91,95,97,99	0
4	NAG	А	602	14/15	0.94	0.10	30,37,41,46	0
4	NAG	В	602	14/15	0.94	0.12	44,48,52,53	0

6.5 Other polymers (i)

There are no such residues in this entry.

