wwPDB X-ray Structure Validation Summary Report (i)

Oct 10, 2023 – 06:10 AM EDT

PDB ID	:	7SO7
Title	:	Novel structural insights for a pair of monoclonal antibodies recognizing non-overlapping epitopes of the glucosyltransferase domain of Clostridium difficile toxin B
Authors	:	Liu, J.
Deposited on	:	2021-10-29
Resolution	:	3.59 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
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:

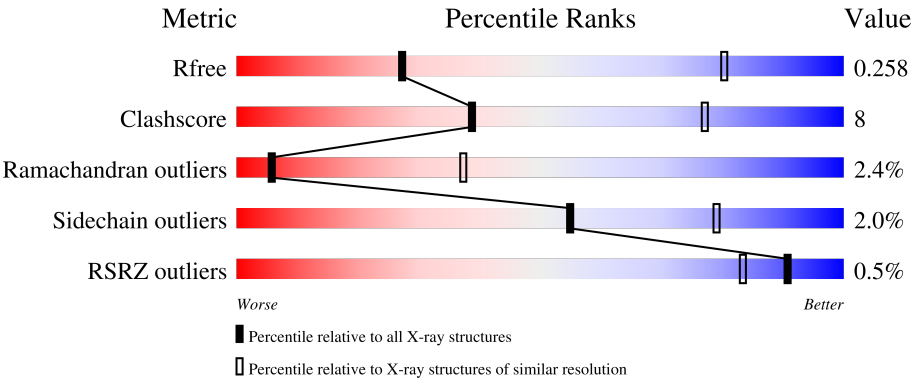
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 3.59 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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1257 (3.70-3.50)
Clashscore	141614	1353 (3.70-3.50)
Ramachandran outliers	138981	1307 (3.70-3.50)
Sidechain outliers	138945	1307 (3.70-3.50)
RSRZ outliers	127900	1161 (3.70-3.50)

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	A	538	% 78% 19%
1	B	538	83% 15%
2	H	213	76% 21%
2	X	213	74% 25%
3	L	218	76% 23%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	Y	218	

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 15130 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 Toxin B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	528	Total	C	N	O	S	0	0	0
			4315	2737	696	865	17			
1	B	529	Total	C	N	O	S	0	0	0
			4326	2743	701	865	17			

- Molecule 2 is a protein called Fab B1 HC.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	X	213	Total	C	N	O	S	0	0	0
			1603	1012	270	313	8			
2	H	213	Total	C	N	O	S	0	0	0
			1603	1012	270	313	8			

- Molecule 3 is a protein called FAB B1 LC.

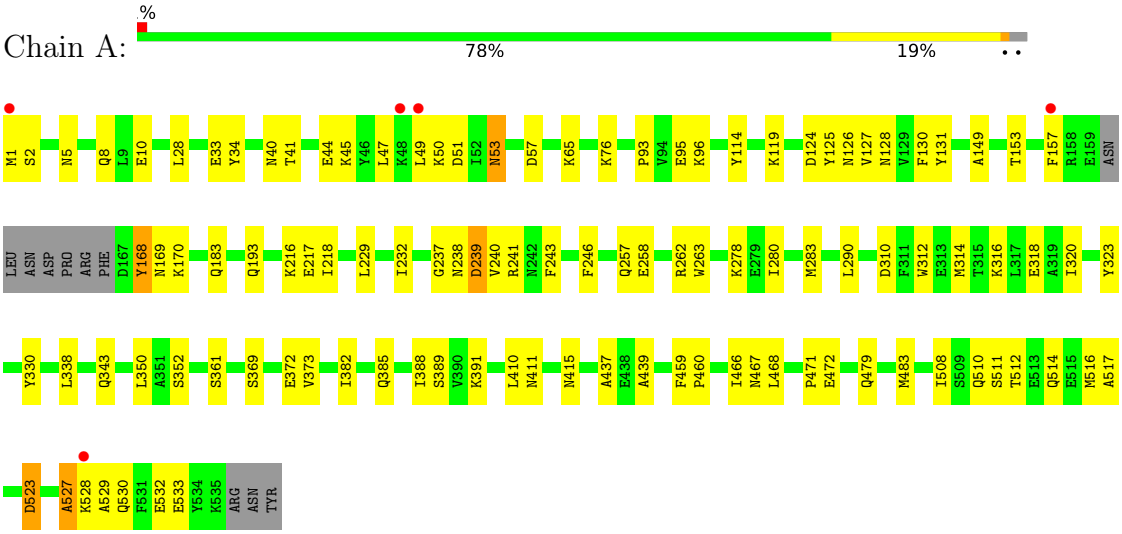
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	Y	218	Total	C	N	O	S	0	0	0
			1642	1037	278	323	4			
3	L	218	Total	C	N	O	S	0	0	0
			1641	1036	278	323	4			

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.

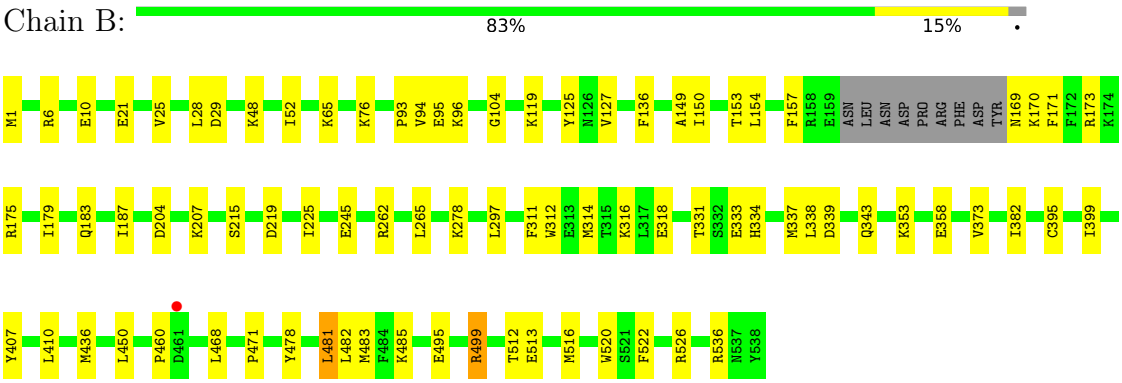
• Molecule 1: Toxin B

Chain A:



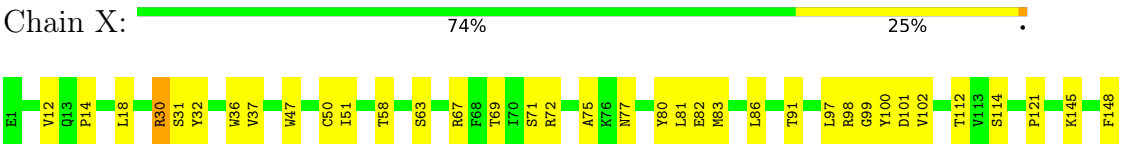
• Molecule 1: Toxin B

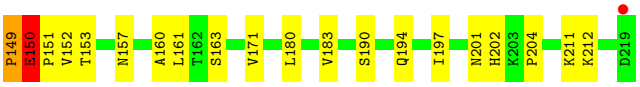
Chain B:



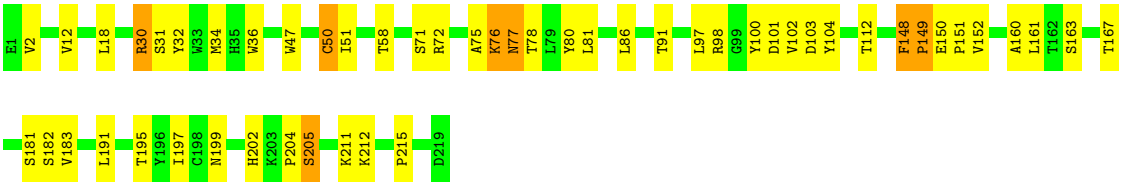
• Molecule 2: Fab B1 HC

Chain X:

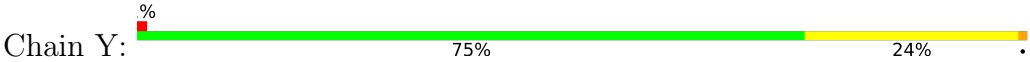


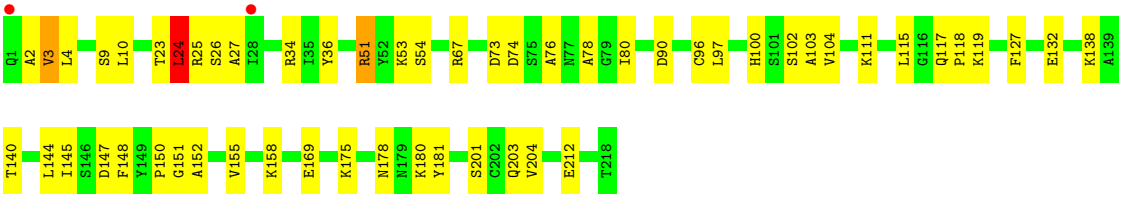


• Molecule 2: Fab B1 HC

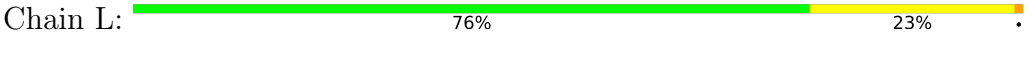


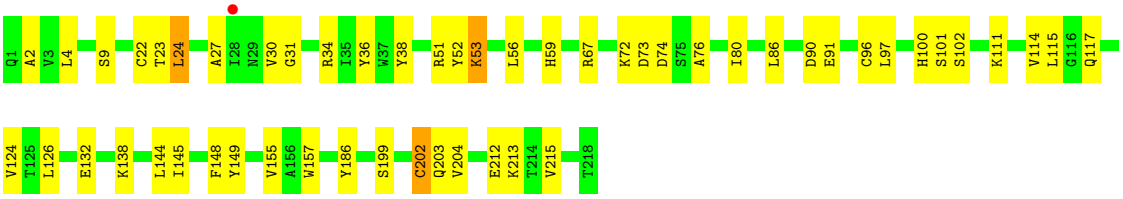
• Molecule 3: FAB B1 LC





• Molecule 3: FAB B1 LC





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	210.54Å 320.08Å 65.61Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	74.80 – 3.59 74.80 – 3.59	Depositor EDS
% Data completeness (in resolution range)	99.9 (74.80-3.59) 99.9 (74.80-3.59)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.06 (at 3.58Å)	Xtriage
Refinement program	PHENIX 1.12_2829	Depositor
R, R_{free}	0.209 , 0.258 0.209 , 0.258	Depositor DCC
R_{free} test set	1998 reflections (3.76%)	wwPDB-VP
Wilson B-factor (Å^2)	91.4	Xtriage
Anisotropy	0.410	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 59.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	15130	wwPDB-VP
Average B, all atoms (Å^2)	93.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 13.09% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

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	0.26	0/4391	0.43	0/5923
1	B	0.26	0/4402	0.41	0/5937
2	H	0.27	0/1641	0.51	0/2234
2	X	0.27	0/1641	0.52	0/2234
3	L	0.28	0/1683	0.52	0/2293
3	Y	0.27	0/1684	0.51	1/2295 (0.0%)
All	All	0.26	0/15442	0.46	1/20916 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
2	H	0	1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	Y	24	LEU	CA-CB-CG	5.19	127.23	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	H	148	PHE	Peptide

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	4315	0	4208	60	0
1	B	4326	0	4223	43	0
2	H	1603	0	1568	35	0
2	X	1603	0	1568	37	0
3	L	1641	0	1611	34	0
3	Y	1642	0	1615	34	0
All	All	15130	0	14793	225	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 8.

The worst 5 of 225 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)
2:H:211:LYS:NZ	3:L:132:GLU:OE2	1.81	1.12
3:L:34:ARG:HG2	3:L:53:LYS:HB2	1.55	0.88
2:X:30:ARG:O	2:X:32:TYR:N	2.17	0.78
1:B:95:GLU:HG3	3:Y:34:ARG:HH22	1.50	0.77
3:L:73:ASP:HB2	3:L:80:ILE:HD11	1.67	0.76

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	524/538 (97%)	471 (90%)	43 (8%)	10 (2%)	8 42

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	B	525/538 (98%)	478 (91%)	44 (8%)	3 (1%)	25 64
2	H	209/213 (98%)	187 (90%)	14 (7%)	8 (4%)	3 27
2	X	209/213 (98%)	190 (91%)	11 (5%)	8 (4%)	3 27
3	L	216/218 (99%)	181 (84%)	28 (13%)	7 (3%)	4 31
3	Y	216/218 (99%)	174 (81%)	33 (15%)	9 (4%)	3 25
All	All	1899/1938 (98%)	1681 (88%)	173 (9%)	45 (2%)	6 37

5 of 45 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	168	TYR
1	A	217	GLU
1	A	240	VAL
1	A	241	ARG
2	X	30	ARG

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	483/493 (98%)	471 (98%)	12 (2%)	47 75
1	B	484/493 (98%)	476 (98%)	8 (2%)	60 82
2	H	180/180 (100%)	177 (98%)	3 (2%)	60 82
2	X	180/180 (100%)	176 (98%)	4 (2%)	52 77
3	L	182/182 (100%)	179 (98%)	3 (2%)	62 83
3	Y	182/182 (100%)	179 (98%)	3 (2%)	62 83
All	All	1691/1710 (99%)	1658 (98%)	33 (2%)	55 79

5 of 33 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	H	199	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	H	205	SER
3	L	199	SER
1	B	1	MET
1	A	361	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	126	ASN
1	A	238	ASN
1	A	257	GLN
1	A	385	GLN
1	A	411	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 monosaccharides in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

The following chains have linkage breaks:

Mol	Chain	Number of breaks
2	X	1
2	H	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	X	128:PRO	C	135:GLY	N	9.41
1	H	128:PRO	C	135:GLY	N	8.56

6 Fit of model and data

6.1 Protein, DNA and RNA chains

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	OWAB(Å^2)	Q<0.9
1	A	528/538 (98%)	-0.13	5 (0%) 84 73	53, 91, 151, 208	0
1	B	529/538 (98%)	-0.17	1 (0%) 95 91	51, 89, 153, 211	0
2	H	213/213 (100%)	0.00	0 100 100	55, 88, 126, 180	0
2	X	213/213 (100%)	0.04	1 (0%) 91 83	54, 87, 132, 185	0
3	L	218/218 (100%)	-0.05	1 (0%) 91 83	58, 85, 132, 225	0
3	Y	218/218 (100%)	-0.04	2 (0%) 84 73	56, 84, 128, 195	0
All	All	1919/1938 (99%)	-0.09	10 (0%) 91 83	51, 88, 145, 225	0

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	48	LYS	2.8
1	A	49	LEU	2.7
1	A	1	MET	2.5
1	A	157	PHE	2.3
1	A	528	LYS	2.3

6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.