wwPDB X-ray Structure Validation Summary Report (i)

Sep 24, 2023 – 07:43 AM EDT

PDB ID	:	5UMI
Title	:	Clostridium difficile TcdA-CROPs bound to PA50 Fab
Authors	:	Kroh, H.K.; Chandrasekaran, R.; Rosenthal, K.; Woods, R.; Jin, X.; Ohi, M.D.; Nyborg, A.C.; Rainey, G.J.; Warrener, P.; Spiller, B.W.; Lacy, D.B.
Deposited on	:	2017-01-27
Resolution	:	3.23 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 (i)) were used in the production of this report:

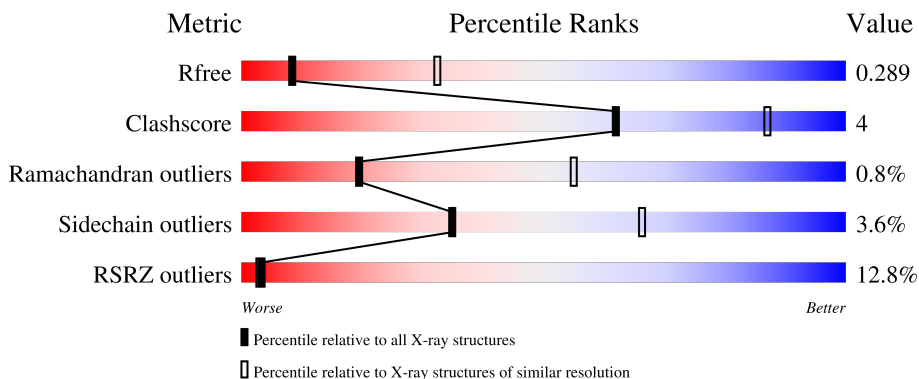|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

1 Overall quality at a glance ⓘ

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

The reported resolution of this entry is 3.23 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		0.289
Clashscore		4
Ramachandran outliers		0.8%
Sidechain outliers		3.6%
RSRZ outliers		12.8%

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$_{free}$	130704	1619 (3.28-3.20)
Clashscore	141614	1755 (3.28-3.20)
Ramachandran outliers	138981	1728 (3.28-3.20)
Sidechain outliers	138945	1727 (3.28-3.20)
RSRZ outliers	127900	1567 (3.28-3.20)

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	C	271	4% / 81% 8% · 10%
2	B	211	17% / 38% 8% · 52%
2	L	211	7% / 91% 9%
3	A	221	13% / 42% 10% · 47%
3	H	221	11% / 85% 11% ·

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 13425 atoms, of which 6570 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 A.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	C	245	3793	1265	1839	324	361	4	0	0	0

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	2440	MET	-	initiating methionine	UNP P16154
C	2441	GLY	-	expression tag	UNP P16154
C	2442	SER	-	expression tag	UNP P16154
C	2443	SER	-	expression tag	UNP P16154
C	2444	HIS	-	expression tag	UNP P16154
C	2445	HIS	-	expression tag	UNP P16154
C	2446	HIS	-	expression tag	UNP P16154
C	2447	HIS	-	expression tag	UNP P16154
C	2448	HIS	-	expression tag	UNP P16154
C	2449	HIS	-	expression tag	UNP P16154
C	2450	SER	-	expression tag	UNP P16154
C	2451	SER	-	expression tag	UNP P16154
C	2452	GLY	-	expression tag	UNP P16154
C	2453	LEU	-	expression tag	UNP P16154
C	2454	VAL	-	expression tag	UNP P16154
C	2455	PRO	-	expression tag	UNP P16154
C	2456	ARG	-	expression tag	UNP P16154
C	2457	GLY	-	expression tag	UNP P16154
C	2458	SER	-	expression tag	UNP P16154
C	2459	HIS	-	expression tag	UNP P16154
C	2460	MET	-	expression tag	UNP P16154

- Molecule 2 is a protein called PA50 Fab Light chain.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	L	210	Total 3171	C 1007	H 1561	N 272	O 326	S 5	0	0	0
2	B	101	Total 1499	C 480	H 735	N 129	O 152	S 3	0	0	0

- Molecule 3 is a protein called PA50 Fab Heavy chain.

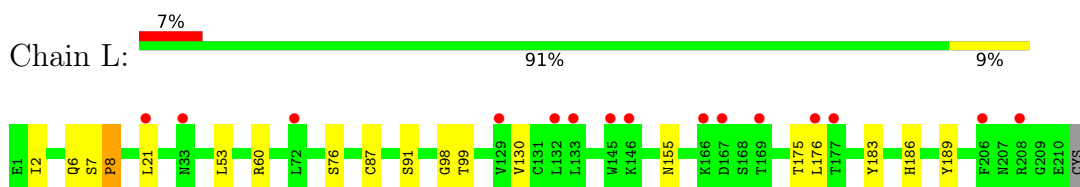| Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	H	213	Total 3161	C 1013	H 1553	N 271	O 316	S 8	0	0	0
3	A	117	Total 1801	C 578	H 882	N 158	O 177	S 6	0	0	0

3 Residue-property plots ⓘ

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.
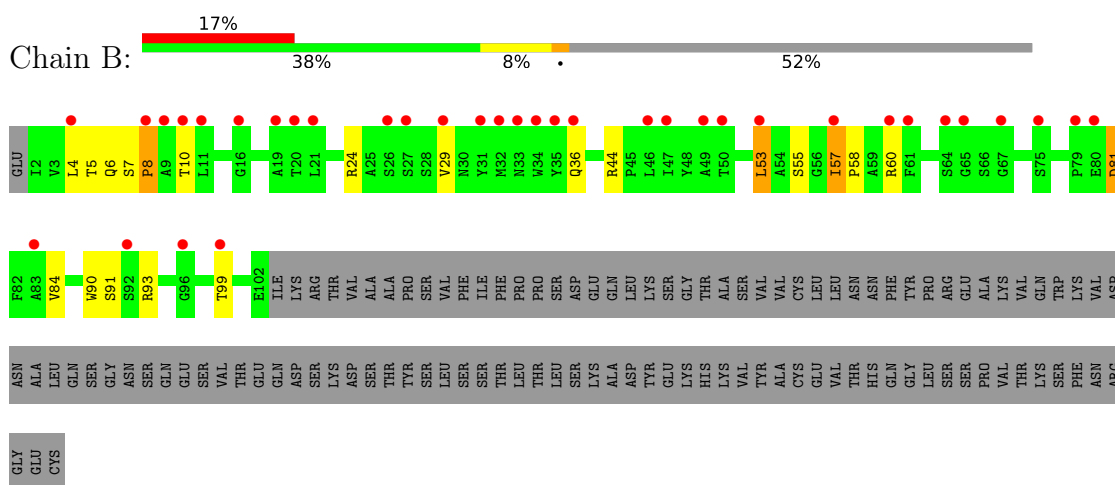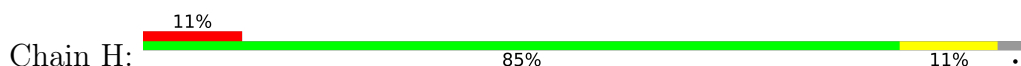
- Molecule 1: Toxin A

Chain C:

- Molecule 2: PA50 Fab Light chain

Chain L:

- Molecule 2: PA50 Fab Light chain

Chain B:

- Molecule 3: PA50 Fab Heavy chain

Chain H:

GLN V2 E10 V18 K19 C22 T30 P53 K54 Y55 D56 I57 M65 G66 R67 R72 T78 Y104 F105 D106 V107 T112 L113 V114 S125 V126 F127 P131 S132 S133 K134 S135 T136 S137 A141 A142 L143 G144 C145 N160 L164 L175 V189

P190 S191 S192 Y199 I200 V203 N204 K211 V212 D213 K214 ARG VAL GLU PRO LYS SER CYS

• Molecule 3: PA50 Fab Heavy chain

13%

Chain A:　　　42%　　　10%　·　　47%

GLN V2 Q3 P14 V18 K19 V20 Y27 M34 D35 W36 V37 R38 Q39 L45 E46 W47 D56 I57 I58 G59 H60 N61 P62 K63 F64 M65 V68 T69 I70 T78 L83 S84 S85 D90 Y94 Y95 C96 A97 R98 S99 D100 R101 G102 W103 Y104 F105 D106 V107

W108 G109 Q110 V114 T115 V116 S117 S118 ALA SER THR LYS GLY PRO SER VAL PHE PRO LEU ALA PRO SER SER LYS SER THR SER GLY GLY THR ALA ALA LEU GLY CYS LEU VAL LYS ASP TYR PHE PRO GLU PRO VAL THR VAL SER TRP ASN SER GLY ALA LEU THR SER GLY VAL HIS

THR PHE PRO ALA VAL LEU GLN SER SER GLY LEU TYR SER LEU SER SER VAL VAL THR VAL PRO SER SER SER LEU GLY THR GLN THR TYR ILE CYS ASN VAL ASN HIS LYS PRO SER ASN THR LYS VAL ASP LYS ARG VAL GLU PRO LYS SER CYS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	93.69Å 77.33Å 97.78Å 90.00° 90.91° 90.00°	Depositor
Resolution (Å)	48.88 – 3.23 93.67 – 3.23	Depositor EDS
% Data completeness (in resolution range)	95.0 (48.88-3.23) 96.0 (93.67-3.23)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.53 (at 3.26Å)	Xtriage
Refinement program	PHENIX 1.11.1_2575	Depositor
R, R_{free}	0.270 , 0.290 0.271 , 0.289	Depositor DCC
R_{free} test set	1126 reflections (5.17%)	wwPDB-VP
Wilson B-factor (Å^2)	70.4	Xtriage
Anisotropy	0.428	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 43.4	EDS
L-test for twinning[2]	$< |L| > = 0.47, < L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.014 for l,k,-h 0.036 for h,-k,-l 0.025 for l,-k,h	Xtriage
F_o,F_c correlation	0.85	EDS
Total number of atoms	13425	wwPDB-VP
Average B, all atoms (Å^2)	90.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.84% 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	C	0.27	0/2011	0.44	0/2725
2	B	0.26	0/783	0.48	0/1066
2	L	0.27	0/1646	0.49	0/2237
3	A	0.27	0/941	0.48	0/1276
3	H	0.28	0/1648	0.50	0/2246
All	All	0.27	0/7029	0.48	0/9550

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
1	C	0	1
3	A	0	1
All	All	0	2

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	A	103	TRP	Peptide
1	C	2477	ALA	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	C	1954	1839	1839	14	0
2	B	764	735	735	12	0
2	L	1610	1561	1562	11	0
3	A	919	882	882	20	0
3	H	1608	1553	1568	8	0
All	All	6855	6570	6586	59	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 4.

The worst 5 of 59 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)
3:H:53:PRO:O	3:H:72:ARG:NH1	2.07	0.88
3:A:47:TRP:O	3:A:61:ASN:ND2	2.11	0.83
1:C:2645:ARG:NH2	2:B:91:SER:O	2.12	0.82
2:L:6:GLN:NE2	2:L:87:CYS:SG	2.58	0.77
2:B:60:ARG:NH2	2:B:81:ASP:OD2	2.18	0.77

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	C	243/271 (90%)	228 (94%)	13 (5%)	2 (1%)	19	55
2	B	99/211 (47%)	91 (92%)	7 (7%)	1 (1%)	15	50
2	L	208/211 (99%)	198 (95%)	9 (4%)	1 (0%)	29	64
3	A	115/221 (52%)	103 (90%)	11 (10%)	1 (1%)	17	52
3	H	211/221 (96%)	196 (93%)	13 (6%)	2 (1%)	17	52

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	876/1135 (77%)	816 (93%)	53 (6%)	7 (1%)	19 55

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	2536	PRO
1	C	2475	ASN
3	H	190	PRO
3	H	131	PRO
2	L	8	PRO

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	C	194/215 (90%)	191 (98%)	3 (2%)	65 83
2	B	84/183 (46%)	78 (93%)	6 (7%)	14 45
2	L	182/183 (100%)	180 (99%)	2 (1%)	73 87
3	A	99/189 (52%)	96 (97%)	3 (3%)	41 70
3	H	181/189 (96%)	168 (93%)	13 (7%)	14 44
All	All	740/959 (77%)	713 (96%)	27 (4%)	35 66

5 of 27 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	H	145	CYS
3	H	213	ASP
3	A	65	MET
3	H	200	ILE
2	B	4	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
2	L	6	GLN
3	H	60	HIS
3	A	110	GLN

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 ⓘ

There are no chain breaks in this entry.

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, 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(Å²)	Q<0.9
1	C	245/271 (90%)	0.82	10 (4%)	37	27	58, 79, 102, 133	0
2	B	101/211 (47%)	1.76	36 (35%)	0	0	92, 121, 149, 157	0
2	L	210/211 (99%)	0.81	15 (7%)	16	11	56, 81, 129, 158	0
3	A	117/221 (52%)	1.34	28 (23%)	0	1	88, 106, 128, 136	0
3	H	213/221 (96%)	0.97	24 (11%)	5	4	48, 70, 133, 151	0
All	All	886/1135 (78%)	1.03	113 (12%)	3	3	48, 88, 133, 158	0

The worst 5 of 113 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	75	SER	7.4
3	H	126	VAL	5.4
2	B	36	GLN	5.1
2	B	21	LEU	4.7
3	A	94	TYR	4.5

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 ⓘ

There are no such residues in this entry.