



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 19, 2016 – 07:14 PM GMT

PDB ID : 4R04  
Title : Clostridium difficile Toxin A (TcdA)  
Authors : Lacy, D.B.; Spiller, B.W.; Rutherford, S.A.  
Deposited on : 2014-07-29  
Resolution : 3.26 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 : unknown  
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

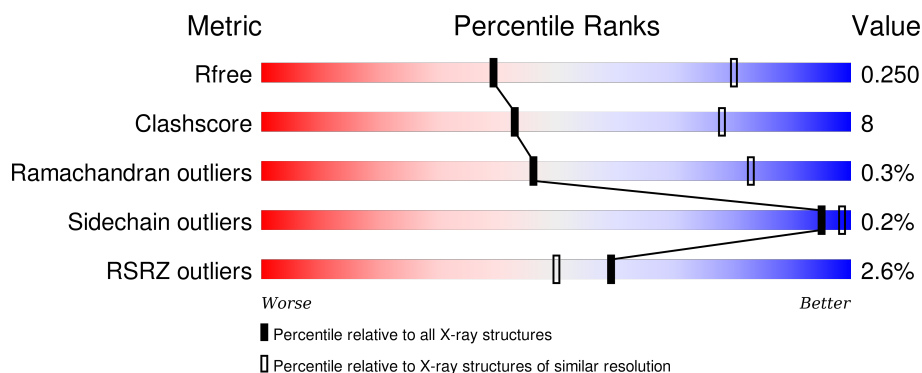
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.26 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	1624 (3.32-3.20)
Clashscore	102246	1806 (3.32-3.20)
Ramachandran outliers	100387	1773 (3.32-3.20)
Sidechain outliers	100360	1771 (3.32-3.20)
RSRZ outliers	91569	1632 (3.32-3.20)

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 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	1838	<div> <div>3%</div> <div>76%</div> <div>21%</div> <div>.</div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 14410 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 A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1793	Total	C	N	O	S	0	0	0
			14409	9199	2332	2852	26			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1833	HIS	-	EXPRESSION TAG	UNP P16154
A	1834	HIS	-	EXPRESSION TAG	UNP P16154
A	1835	HIS	-	EXPRESSION TAG	UNP P16154
A	1836	HIS	-	EXPRESSION TAG	UNP P16154
A	1837	HIS	-	EXPRESSION TAG	UNP P16154
A	1838	HIS	-	EXPRESSION TAG	UNP P16154

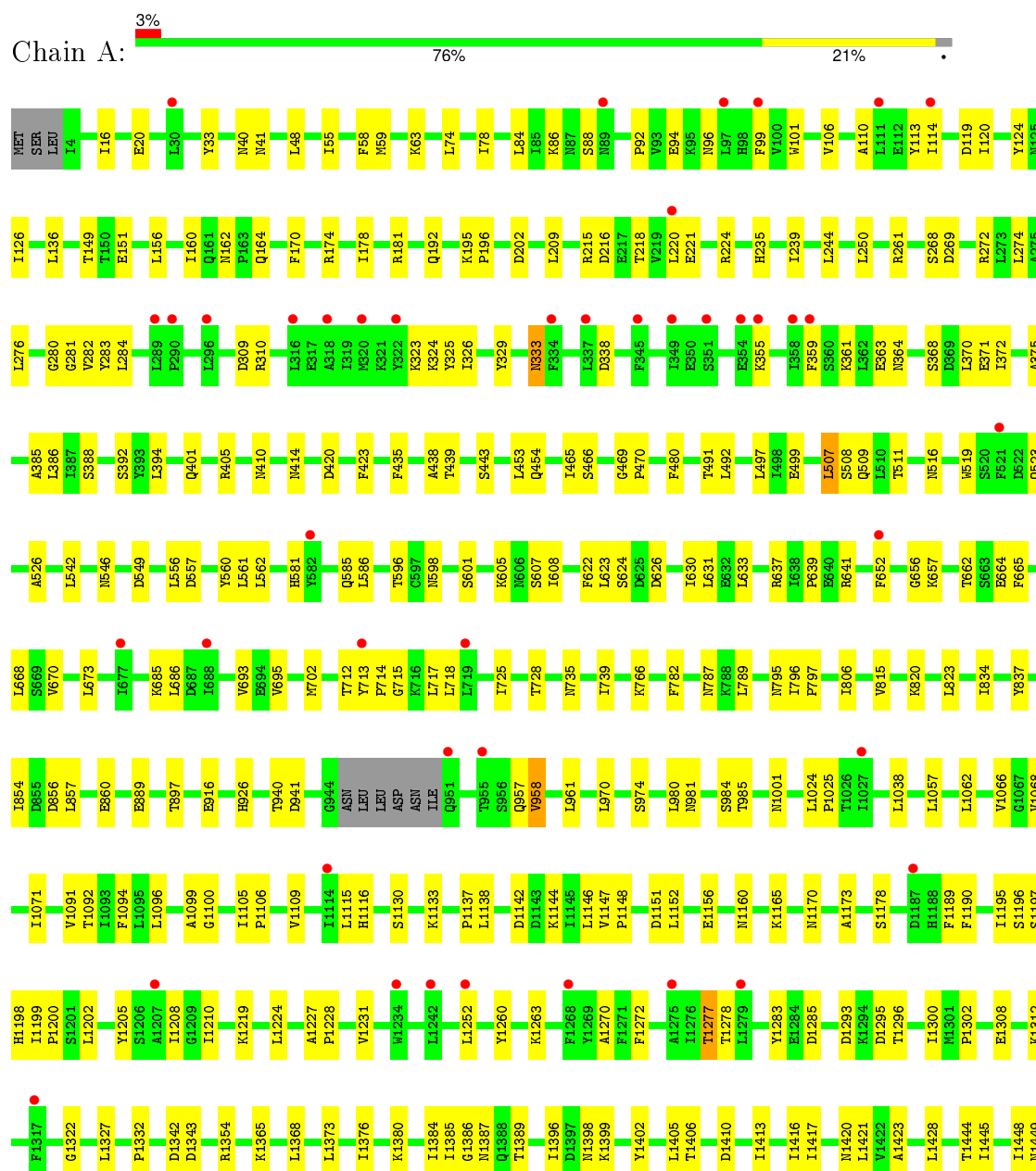
- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Zn	0	0
			1	1		

### 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 ( $\text{RSRZ} > 2$ ). 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



LYS	L1715	S1590	T1450
LEU	H1723	K1593	L1451
VAL	LYS	I1604	G1452
LYS	L1730	G1612	L1453
GLY	F1735	K1613	D1454
HIS	E1736	T1614	S1455
HIS	Y1737	M1615	Y1470
HIS	S1744	L1616	A1473
HIS	D1745	G1617	K1487
		K1627	I1492
		Y1632	L1493
		R1750	E1494
		K1762	F1495
			L1501
		K1766	E1502
		G1767	F1503
		I1768	N1504
			S1505
		N1771	K1518
		T1772	
		Q1773	I1521
		K1777	V1530
		D1781	D1535
		L1790	K1536
		M1801	D1539
		S1802	S1546
		GLU	Q1549
		ASN	Y1556
		GLU	L1557
		LEU	N1558
		LEU	Y1562
		ASP	D1567
		ARG	F1568
		ASP	V1569
		ASP	S1572
		HIS	D1573
		LEU	G1574
		GLY	H1575
		PHE	T1578
		LYS	N1579
		ILE	S1579
		ILE	F1585
		Y1695	I1589
		T1696	
		ASP	
		ASN	
		LYS	
		THR	
		TYR	
		TYR	
		ASP	
		GLU	
		ASP	
		SER	

## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	303.49Å 124.54Å 75.95Å 90.00° 97.50° 90.00°	Depositor
Resolution (Å)	61.62 – 3.26 62.27 – 3.26	Depositor EDS
% Data completeness (in resolution range)	99.9 (61.62-3.26) 94.3 (62.27-3.26)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.32 (at 3.26Å)	Xtriage
Refinement program	PHENIX (phenix.refine: dev_1565)	Depositor
R, $R_{free}$	0.221 , 0.249 0.221 , 0.250	Depositor DCC
$R_{free}$ test set	3928 reflections (4.57%)	DCC
Wilson B-factor (Å <sup>2</sup> )	100.7	Xtriage
Anisotropy	0.227	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 58.0	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Outliers	0 of 43828 reflections	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	14410	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	122.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.14% 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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	0.22	0/14671	0.39	1/19845 (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	507	LEU	CA-CB-CG	5.04	126.88	115.30

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	14409	0	14349	235	0
2	A	1	0	0	0	0
All	All	14410	0	14349	235	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 235 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:371:GLU:HA	1:A:394:LEU:HD13	1.61	0.81
1:A:1495:PHE:HB3	1:A:1503:PHE:HB3	1.63	0.78
1:A:1156:GLU:HB3	1:A:1165:LYS:HB2	1.65	0.78
1:A:1178:SER:HB2	1:A:1190:PHE:HB3	1.74	0.69
1:A:239:ILE:HG23	1:A:244:LEU:HD12	1.74	0.68

There are no symmetry-related clashes.

## 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	1789/1838 (97%)	1656 (93%)	127 (7%)	6 (0%)	46	83

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	420	ASP
1	A	549	ASP
1	A	787	ASN
1	A	338	ASP
1	A	958	VAL

### 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.

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
-----	-------	----------	-----------	----------	-------------

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	1652/1696 (97%)	1648 (100%)	4 (0%)	95 98

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

Mol	Chain	Res	Type
1	A	333	ASN
1	A	1109	VAL
1	A	1277	THR
1	A	1578	THR

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

Mol	Chain	Res	Type
1	A	701	ASN

### 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 [i](#)

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

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

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 [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	1793/1838 (97%)	0.26	47 (2%)	59 50	58, 115, 195, 286	0

The worst 5 of 47 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	349	ILE	4.6
1	A	354	GLU	4.6
1	A	290	PRO	4.1
1	A	337	LEU	4.0
1	A	322	TYR	3.6

### 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
2	ZN	A	1901	1/1	0.96	0.12	-1.18	132,132,132,132	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.