



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 22, 2024 – 03:21 AM EST

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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
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

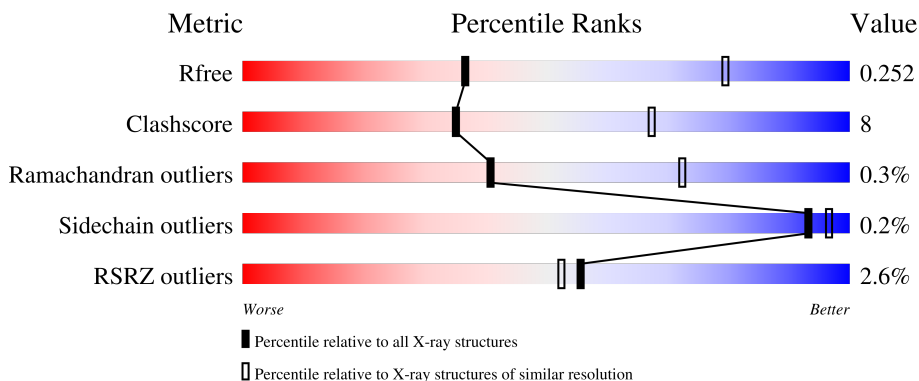
# 1 Overall quality at a glance

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))
$R_{free}$	130704	1191 (3.30-3.22)
Clashscore	141614	1251 (3.30-3.22)
Ramachandran outliers	138981	1229 (3.30-3.22)
Sidechain outliers	138945	1228 (3.30-3.22)
RSRZ outliers	127900	1154 (3.30-3.22)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	

## 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
			Total	C	N	O	S			
1	A	1793	14409	9199	2332	2852	26	0	0	0

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

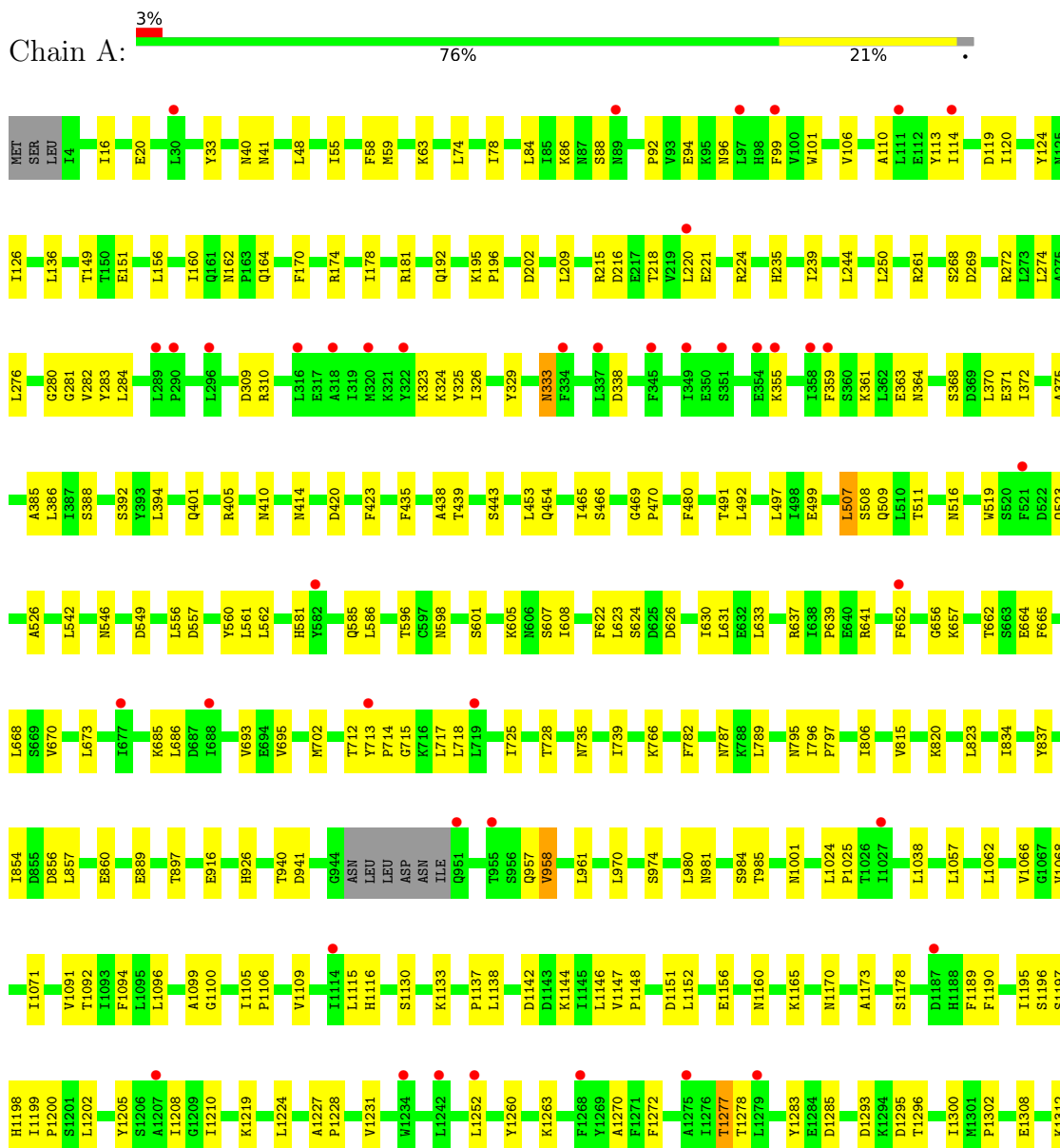
- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

### 3 Residue-property plots [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 A



LYS	F1317
LEU	G1322
VAL	L1327
LYS	P1332
GLY	D1342
HIS	D1343
HIS	R1354
HIS	K1365
HIS	L1368
HIS	L1373
HIS	I1376
HIS	K1380
HIS	I1384
HIS	I1385
HIS	G1386
HIS	M1387
HIS	Q1388
HIS	T1389
HIS	I1396
HIS	D1397
HIS	M1398
HIS	K1399
HIS	Y1402
HIS	L1405
HIS	T1406
HIS	D1410
HIS	I1413
HIS	I1416
HIS	I1417
HIS	M1420
HIS	L1421
HIS	Y1422
HIS	A1423
HIS	L1428
HIS	T1444
HIS	I1445
HIS	I1448
HIS	M1449
HIS	T1450
HIS	L1451
HIS	G1452
HIS	L1453
HIS	D1454
HIS	S1455
HIS	Y1470
HIS	A1473
HIS	K1487
HIS	I1492
HIS	L1493
HIS	E1494
HIS	F1495
HIS	L1501
HIS	E1502
HIS	F1503
HIS	M1504
HIS	S1505
HIS	K1518
HIS	I1521
HIS	V1530
HIS	D1535
HIS	K1536
HIS	D1539
HIS	L1546
HIS	Q1549
HIS	Y1556
HIS	L1557
HIS	M1558
HIS	Y1562
HIS	D1567
HIS	F1568
HIS	V1569
HIS	S1572
HIS	D1573
HIS	G1574
HIS	H1575
HIS	T1578
HIS	S1579
HIS	F1585
HIS	I1589
HIS	S1690
HIS	K1593
HIS	I1604
HIS	G1612
HIS	K1613
HIS	T1614
HIS	M1615
HIS	L1616
HIS	G1617
HIS	K1627
HIS	Y1632
HIS	W1636
HIS	S1640
HIS	M1649
HIS	G1650
HIS	R1651
HIS	V1655
HIS	Y1669
HIS	M1660
HIS	P1661
HIS	G1664
HIS	E1665
HIS	T1669
HIS	S1670
HIS	L1678
HIS	Y1679
HIS	Y1684
HIS	I1685
HIS	L1688
HIS	I1690
HIS	D1693
HIS	L1694
HIS	Y1695
HIS	T1696
HIS	I1701
HIS	Y1705
HIS	Y1706
HIS	S1707
HIS	M1708
HIS	Y1711
HIS	I1715
HIS	H1723
HIS	M1729
HIS	L1730
HIS	F1735
HIS	E1736
HIS	Y1737
HIS	S1744
HIS	D1745
HIS	V1749
HIS	R1750
HIS	K1762
HIS	K1766
HIS	G1767
HIS	I1768
HIS	M1771
HIS	T1772
HIS	Q1773
HIS	K1777
HIS	D1781
HIS	L1790
HIS	M1801
HIS	S1802
HIS	GLU
HIS	ASN
HIS	GLU
HIS	LEU
HIS	ASP
HIS	ARG
HIS	ASP
HIS	HIS
HIS	LEU
HIS	GLY
HIS	PHE
HIS	LYS
HIS	ILE
HIS	ILE
HIS	ASP
HIS	ASN
HIS	LYS
HIS	THR
HIS	TYR
HIS	TYR
HIS	ASP
HIS	GLU
HIS	ASP
HIS	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>	1.66 (at 3.26Å)	Xtrriage
Refinement program	PHENIX (phenix.refine: dev_1565)	Depositor
R, $R_{free}$	0.221 , 0.249 0.223 , 0.252	Depositor DCC
$R_{free}$ test set	2000 reflections (4.56%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	99.6	Xtrriage
Anisotropy	0.235	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 58.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	14410	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	122.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.15% 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, 0.333 respectively for untwinned datasets, and 0.375, 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%)	41 72

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
1	A	1652/1696 (97%)	1648 (100%)	4 (0%)	93 96

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

Sometimes 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 monosaccharides 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%) 56 52	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 monosaccharides 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. 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	B-factors(Å <sup>2</sup> )	Q<0.9
2	ZN	A	1901	1/1	0.96	0.12	132,132,132,132	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.