



wwPDB X-ray Structure Validation Summary Report ⓘ

May 28, 2020 – 09:17 pm BST

PDB ID : 2A97
Title : Crystal structure of catalytic domain of Clostridium botulinum neurotoxin serotype F
Authors : Agarwal, R.; Binz, T.; Swaminathan, S.
Deposited on : 2005-07-11
Resolution : 1.80 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 ⓘ symbol.

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	:	2.11
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)	:	2.11

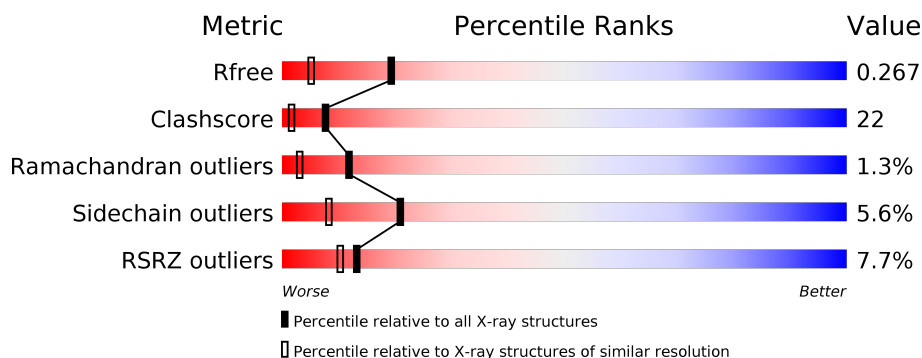
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.80 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	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)
RSRZ outliers	127900	5850 (1.80-1.80)

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 ≥ 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	439	<div> <div>6%</div> <div> <div></div> <div>59%</div> <div>27%</div> <div>•</div> <div>11%</div> </div> </div>
1	B	439	<div> <div>8%</div> <div> <div></div> <div>62%</div> <div>23%</div> <div>•</div> <div>11%</div> </div> </div>

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 6586 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 Botulinum neurotoxin type F.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	392	Total	C	N	O	S	0	0	0
			3168	2038	505	618	7			
1	B	392	Total	C	N	O	S	0	0	0
			3168	2038	505	618	7			

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

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

- Molecule 3 is CADMIUM ION (three-letter code: CD) (formula: Cd).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	2	Total	Cd	0	0
			2	2		
3	A	5	Total	Cd	0	0
			5	5		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	109	Total	O	0	0
			109	109		
4	B	132	Total	O	0	0
			132	132		

These plots are drawn for all protein, RNA and DNA 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.

Chain A:

Chain B:

8% 62% 23% 11%

Label	Color	Category
MET	Grey	Category 4
P2	Green	Category 2
N11	Yellow	Category 3
D12	Green	Category 2
V14	Green	Category 2
P13	Green	Category 2
N15	Yellow	Category 3
D16	Green	Category 2
D17	Green	Category 2
T18	Green	Category 2
I19	Green	Category 2
L20	Green	Category 2
M22	Green	Category 2
Q23	Green	Category 2
T24	Green	Category 2
P25	Green	Category 2
Y26	Green	Category 2
K29	Yellow	Category 3
S30	Green	Category 2
K31	Yellow	Category 3
K32	Green	Category 2
Y33	Green	Category 2
M40	Green	Category 2
R41	Yellow	Category 3
I46	Green	Category 2
P47	Green	Category 2
E48	Green	Category 2
T51	Yellow	Category 3
F52	Green	Category 2
G53	Green	Category 2
T54	Green	Category 2
N55	Yellow	Category 3
P56	Yellow	Category 3
S57	Green	Category 2
D58	Green	Category 2
F59	Green	Category 2
D60	Green	Category 2
P61	Yellow	Category 3
P62	Green	Category 2
A63	Green	Category 2
S64	Yellow	Category 3
L65	Yellow	Category 3
R66	Yellow	Category 3
R67	Red	Category 1
G68	Green	Category 2
S69	Yellow	Category 3
S70	Yellow	Category 3
A71	Yellow	Category 3
Y72	Yellow	Category 3
Y73	Yellow	Category 3
R86	Yellow	Category 3
Y90	Yellow	Category 3
M101	Green	Category 2
P102	Green	Category 2
A103	Yellow	Category 3
V106	Yellow	Category 3
E110	Green	Category 2
I111	Green	Category 2
S112	Green	Category 2
K115	Yellow	Category 3
D121	Green	Category 2
P124	Yellow	Category 3
E127	Yellow	Category 3
R133	Yellow	Category 3
K140	Yellow	Category 3
L141	Yellow	Category 3
S142	Yellow	Category 3
T143	Green	Category 2
M144	Green	Category 2
V145	Green	Category 2
H149	Yellow	Category 3
L150	Yellow	Category 3
L151	Yellow	Category 3
M152	Yellow	Category 3
L153	Yellow	Category 3
L154	Yellow	Category 3
P160	Green	Category 2
D161	Green	Category 2
I162	Yellow	Category 3
G167	Green	Category 2
Y168	Green	Category 2
P169	Green	Category 2
V170	Green	Category 2
I174	Yellow	Category 3
D175	Yellow	Category 3
I277	Green	Category 2
I278	Green	Category 2
L282	Green	Category 2
M282	Green	Category 2
S415	Yellow	Category 3
I1E	Green	Category 2
PRO	Green	Category 2
ASP	Green	Category 2
LYS	Green	Category 2
GLY	Green	Category 2
VAL	Green	Category 2
GLU	Green	Category 2
LYS	Green	Category 2
Y201	Green	Category 2
T202	Green	Category 2
N204	Green	Category 2
ASP	Grey	Category 4
I1E	Green	Category 2
SER	Green	Category 2
GLY	Green	Category 2
HIS	Green	Category 2
ASN	Green	Category 2
SER	Green	Category 2
SER	Green	Category 2
THR	Green	Category 2
E215	Green	Category 2
I218	Yellow	Category 3
A222	Green	Category 2
I223	Green	Category 2
S224	Green	Category 2
L225	Green	Category 2
A226	Yellow	Category 3
H227	Green	Category 2
E246	Green	Category 2
F374	Green	Category 2
T247	Green	Category 2
I248	Green	Category 2
GLU	Grey	Category 4
VAL	Grey	Category 4
LYS	Green	Category 2
GLN	Green	Category 2
ALA	Green	Category 2
PRO	Green	Category 2
LEU	Green	Category 2
MET	Green	Category 2
I1E	Green	Category 2
ALA	Green	Category 2
GLU	Green	Category 2
LYS	Green	Category 2
P261	Green	Category 2
I262	Green	Category 2
R263	Green	Category 2
M276	Green	Category 2
I277	Green	Category 2
K411	Green	Category 2
I412	Green	Category 2
I413	Green	Category 2
D414	Red	Category 1
S415	Yellow	Category 3
I1E	Green	Category 2
PRO	Green	Category 2
ASP	Green	Category 2
LYS	Green	Category 2
GLY	Green	Category 2
VAL	Green	Category 2
GLU	Green	Category 2
LYS	Green	Category 2
Y214	Green	Category 2

VAL
LYS
PHE
CYS
LYS
SER
VAL
ILE
PRO
ARG
LYS
GLY
THR
LYS

4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	173.40Å 53.24Å 113.87Å 90.00° 119.17° 90.00°	Depositor
Resolution (Å)	26.05 – 1.80 26.05 – 1.80	Depositor EDS
% Data completeness (in resolution range)	76.6 (26.05-1.80) 76.7 (26.05-1.80)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.28 (at 1.80Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.241 , 0.269 0.240 , 0.267	Depositor DCC
R_{free} test set	1974 reflections (2.83%)	wwPDB-VP
Wilson B-factor (Å ²)	19.7	Xtriage
Anisotropy	0.374	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.40 , 54.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	6586	wwPDB-VP
Average B, all atoms (Å ²)	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.57% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

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.

5 Model quality

5.1 Standard geometry

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

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	0.34	0/3245	0.62	0/4407
1	B	0.36	0/3245	0.61	0/4407
All	All	0.35	0/6490	0.61	0/8814

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	3168	0	3090	155	0
1	B	3168	0	3090	124	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	5	0	0	0	0
3	B	2	0	0	0	0
4	A	109	0	0	14	0
4	B	132	0	0	13	0
All	All	6586	0	6180	274	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 22.

The worst 5 of 274 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:86:ARG:HH11	1:B:86:ARG:HB2	1.04	1.11
1:A:174:ILE:HG22	1:A:176:PRO:HD2	1.30	1.09
1:A:365:ARG:HD3	1:A:370:ILE:HD13	1.44	0.97
1:B:363:LYS:H	1:B:404:GLN:HE22	1.04	0.96
1:A:67:ASN:HD21	1:A:73:TYR:H	1.05	0.96

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	386/439 (88%)	373 (97%)	11 (3%)	2 (0%)	29	15
1	B	386/439 (88%)	358 (93%)	20 (5%)	8 (2%)	7	1
All	All	772/878 (88%)	731 (95%)	31 (4%)	10 (1%)	12	3

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	67	ASN
1	B	69	SER
1	B	374	PHE
1	B	413	ILE
1	B	414	ASP

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	354/395 (90%)	334 (94%)	20 (6%)	21	8
1	B	354/395 (90%)	334 (94%)	20 (6%)	21	8
All	All	708/790 (90%)	668 (94%)	40 (6%)	21	8

5 of 40 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	369	PHE
1	B	66	LYS
1	B	369	PHE
1	B	15	ASN
1	B	67	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 23 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	11	ASN
1	B	55	ASN
1	B	404	GLN
1	B	15	ASN
1	B	67	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 carbohydrates in this entry.

5.6 Ligand geometry

Of 9 ligands modelled in this entry, 9 are 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, 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(Å ²)	Q<0.9
1	A	392/439 (89%)	0.46	27 (6%) 16 13	16, 28, 41, 52	0
1	B	392/439 (89%)	0.49	33 (8%) 11 8	13, 26, 43, 51	0
All	All	784/878 (89%)	0.47	60 (7%) 13 10	13, 27, 42, 52	0

The worst 5 of 60 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	369	PHE	13.0
1	A	414	ASP	12.1
1	A	415	SER	9.4
1	A	2	PRO	8.2
1	B	414	ASP	7.9

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. The B-factors column lists the minimum, median, 95th 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(\AA^2)	Q<0.9
3	CD	A	5504	1/1	0.98	0.02	45,45,45,45	0
3	CD	A	5506	1/1	0.98	0.08	39,39,39,39	0
3	CD	B	5505	1/1	0.98	0.11	53,53,53,53	0
2	ZN	B	2437	1/1	0.99	0.14	17,17,17,17	0
3	CD	A	5507	1/1	0.99	0.05	29,29,29,29	0
2	ZN	A	1437	1/1	0.99	0.16	18,18,18,18	0
3	CD	B	5503	1/1	0.99	0.04	46,46,46,46	0
3	CD	A	5500	1/1	1.00	0.06	21,21,21,21	0
3	CD	A	5502	1/1	1.00	0.02	32,32,32,32	0

6.5 Other polymers [i](#)

There are no such residues in this entry.