Full wwPDB X-ray Structure Validation Report ⓘ

Aug 16, 2023 – 04:26 AM EDT

PDB ID	:	1YBI
Title	:	Crystal structure of HA33A, a neurotoxin-associated protein from Clostridium botulinum type A
Authors	:	Arndt, J.W.; Gu, J.; Jaroszewski, L.; Schwarzenbacher, R.; Hanson, M.; Lebeda, F.J.; Stevens, R.C.
Deposited on	:	2004-12-20
Resolution	:	1.50 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 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:

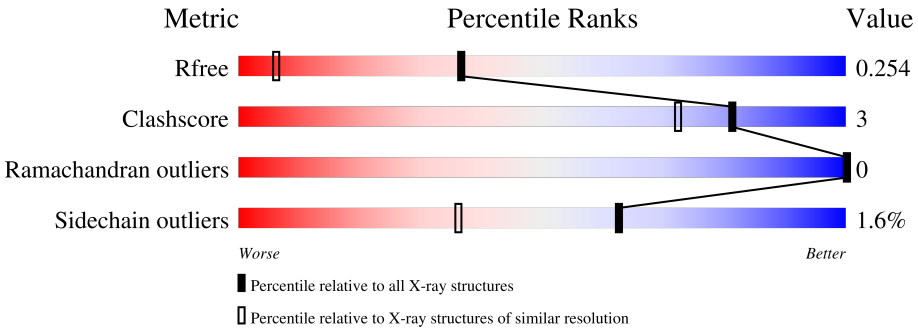
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

1 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 1.50 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	2936 (1.50-1.50)
Clashscore	141614	3144 (1.50-1.50)
Ramachandran outliers	138981	3066 (1.50-1.50)
Sidechain outliers	138945	3064 (1.50-1.50)

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%

Mol	Chain	Length	Quality of chain
1	A	288	90% 8% .
1	B	288	96% . .

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 5241 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 non-toxin haemagglutinin HA34.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	284	Total	C	N	O	S	0	5	0
			2335	1478	399	455	3			
1	B	284	Total	C	N	O	S	0	6	0
			2342	1482	405	452	3			

- Molecule 2 is water.

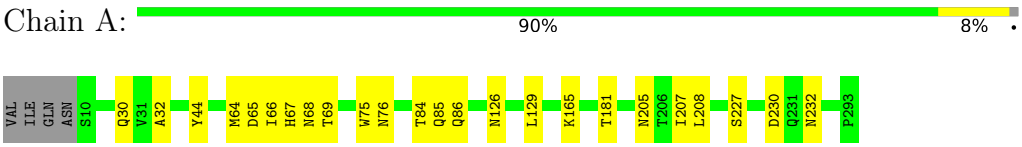
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	238	Total	O	0	0
			238	238		
2	B	326	Total	O	0	0
			326	326		

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. 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. 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.

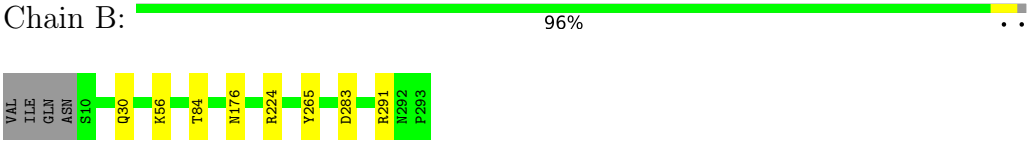
• Molecule 1: non-toxin haemagglutinin HA34

Chain A:



• Molecule 1: non-toxin haemagglutinin HA34

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	104.08Å 146.59Å 35.71Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	21.22 – 1.50 21.22 – 1.50	Depositor EDS
% Data completeness (in resolution range)	87.1 (21.22-1.50) 87.1 (21.22-1.50)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.07	Depositor
$< I/\sigma(I) >$ [1]	4.16 (at 1.50Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.168 , 0.194 0.234 , 0.254	Depositor DCC
R_{free} test set	3820 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å^2)	15.3	Xtriage
Anisotropy	0.093	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 45.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	5241	wwPDB-VP
Average B, all atoms (Å^2)	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.56% 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	0.42	0/2405	0.63	0/3278
1	B	0.47	0/2416	0.69	0/3291
All	All	0.45	0/4821	0.66	0/6569

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	2335	0	2255	22	0
1	B	2342	0	2263	4	0
2	A	238	0	0	0	0
2	B	326	0	0	1	0
All	All	5241	0	4518	24	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 3.

All (24) 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:68:ASN:OD1	1:A:69:THR:HG23	2.03	0.58

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:68:ASN:OD1	1:A:69:THR:CG2	2.52	0.57
1:A:230:ASP:OD2	1:A:232:ASN:HB2	2.06	0.55
1:B:176:ASN:HA	1:B:224[A]:ARG:HG2	1.91	0.52
1:A:68:ASN:OD1	1:A:69:THR:N	2.45	0.50
1:A:67:HIS:O	1:A:68:ASN:HB2	2.12	0.49
1:A:205:ASN:OD1	1:A:207:ILE:HG22	2.13	0.49
1:A:66:ILE:O	1:A:67:HIS:C	2.51	0.48
1:A:75:TRP:NE1	1:A:129:LEU:HD12	2.29	0.48
1:A:126:ASN:HB3	1:B:265:TYR:OH	2.13	0.47
1:B:56:LYS:HE2	2:B:348:HOH:O	2.14	0.46
1:A:230:ASP:OD2	1:A:232:ASN:N	2.39	0.46
1:A:32:ALA:HB1	1:B:283:ASP:HB3	1.99	0.45
1:A:75:TRP:CE2	1:A:129:LEU:HD12	2.52	0.45
1:A:76:ASN:HD21	1:A:85[B]:GLN:HE21	1.64	0.44
1:A:76:ASN:ND2	1:A:85[B]:GLN:NE2	2.66	0.43
1:A:68:ASN:O	1:A:69:THR:OG1	2.34	0.43
1:A:68:ASN:CG	1:A:69:THR:HG23	2.39	0.42
1:A:65:ASP:OD2	1:A:66:ILE:N	2.52	0.42
1:A:64:MET:O	1:A:66:ILE:HD12	2.19	0.41
1:A:165:LYS:NZ	1:A:181:THR:OG1	2.54	0.41
1:A:44:TYR:CD1	1:A:66:ILE:HD13	2.56	0.41
1:A:75:TRP:CD1	1:A:129:LEU:HD12	2.55	0.41
1:A:44:TYR:CE1	1:A:66:ILE:HD13	2.56	0.40

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	287/288 (100%)	276 (96%)	11 (4%)	0	100	100
1	B	288/288 (100%)	282 (98%)	6 (2%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	575/576 (100%)	558 (97%)	17 (3%)	0	100 100

There are no Ramachandran outliers to report.

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	264/264 (100%)	259 (98%)	5 (2%)	57 27
1	B	263/264 (100%)	260 (99%)	3 (1%)	73 53
All	All	527/528 (100%)	519 (98%)	8 (2%)	62 39

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	30	GLN
1	A	84	THR
1	A	86	GLN
1	A	208	LEU
1	A	227	SER
1	B	30	GLN
1	B	84	THR
1	B	291	ARG

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	76	ASN
1	A	268	GLN
1	B	67	HIS
1	B	104	ASN
1	B	268	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

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.