wwPDB X-ray Structure Validation Summary Report (i)

Dec 9, 2023 – 05:11 pm GMT

PDB ID	:	2J42
Title	:	low quality crystal structure of the transport component C2-II of the C2-toxin from Clostridium botulinum
Authors	:	Schleberger, C.; Hochmann, H.; Barth, H.; Aktories, K.; Schulz, G.E.
Deposited on	:	2006-08-24
Resolution	:	3.13 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:

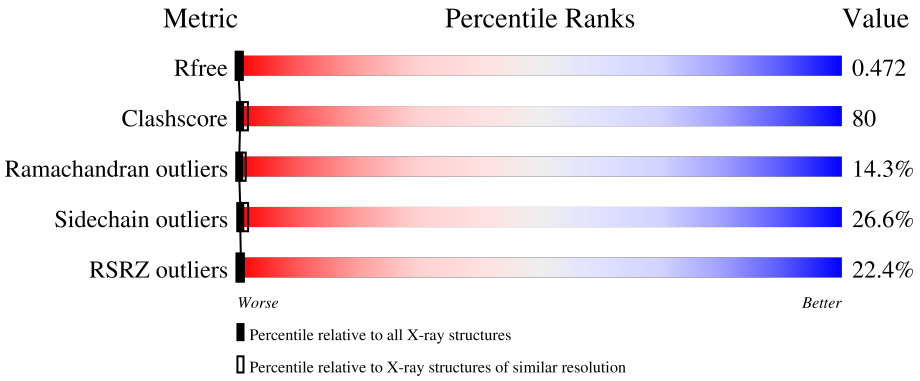
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 3.13 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	1626 (3.18-3.10)
Clashscore	141614	1735 (3.18-3.10)
Ramachandran outliers	138981	1677 (3.18-3.10)
Sidechain outliers	138945	1677 (3.18-3.10)
RSRZ outliers	127900	1588 (3.18-3.10)

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	A	721	16% 15% 36% 17% · 30%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 3948 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 C2 TOXIN COMPONENT-II.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	505	3948	2493	654	786	15	0	0	1

There are 11 discrepancies between the modelled and reference sequences:

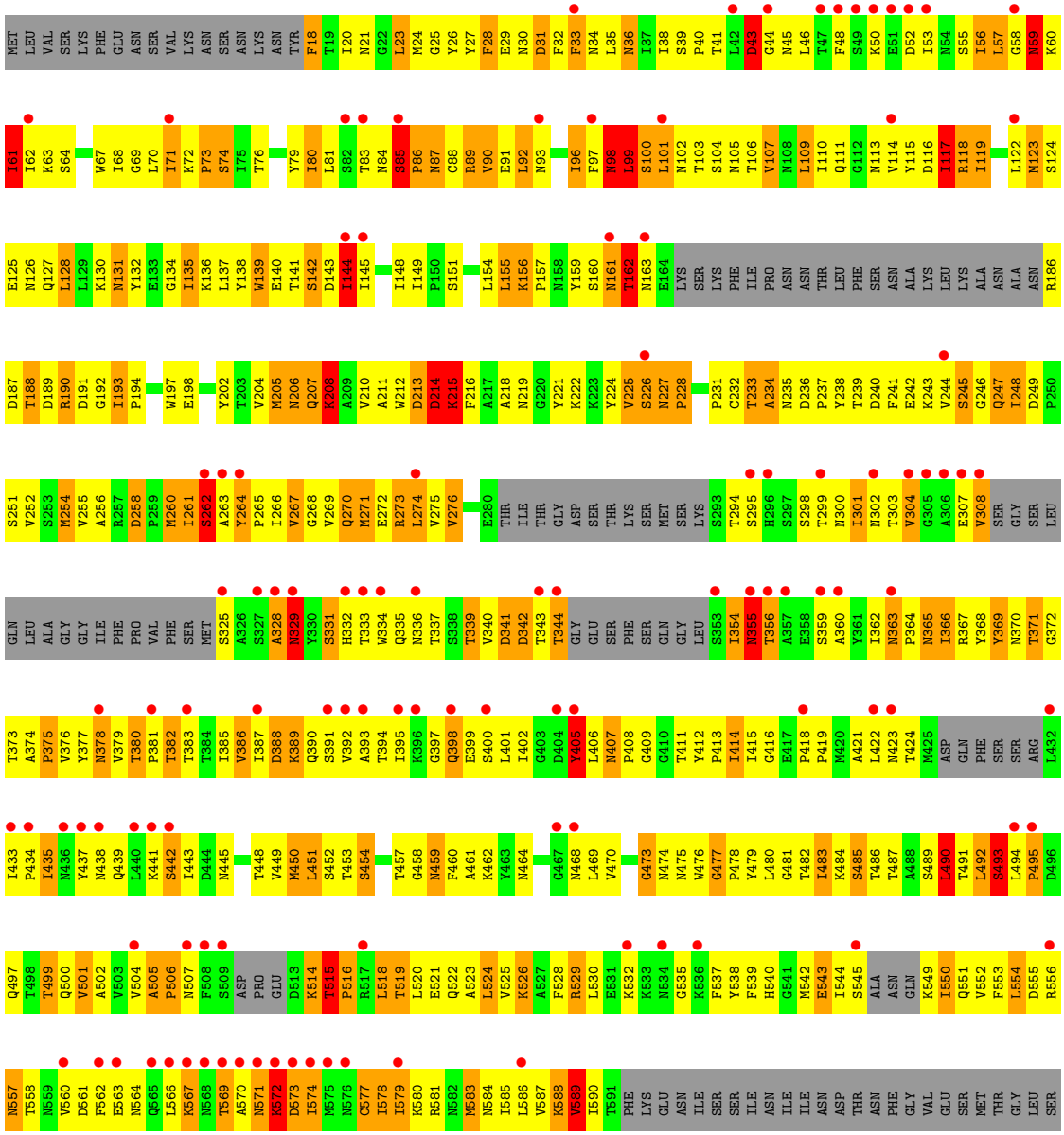
Chain	Residue	Modelled	Actual	Comment	Reference
A	357	ALA	GLY	conflict	UNP O86171
A	494	LEU	PHE	conflict	UNP O86171
A	495	PRO	SER	conflict	UNP O86171
A	496	ASP	GLY	conflict	UNP O86171
A	517	ARG	LYS	conflict	UNP O86171
A	529	ARG	ALA	conflict	UNP O86171
A	542	MET	LEU	conflict	UNP O86171
A	556	ARG	SER	conflict	UNP O86171
A	560	VAL	ASN	conflict	UNP O86171
A	571	ASN	ASP	conflict	UNP O86171
A	576	ASN	HIS	conflict	UNP O86171

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: C2 TOXIN COMPONENT-II

Chain A:



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

ILE GLY TYR TYR ASP GLU PHE PHE TYR PHE TYR CYS ASN GLY SER LYS SER PHE PHE ASP ILE SER CYS ASP ILE ILE ASN SER ILE ASN ARG LEU SER GLY VAL PHE LEU ILE GLU LEU ASP LYS LEU ILE ILE

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	104.40Å 104.40Å 153.20Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 3.13 61.76 – 3.13	Depositor EDS
% Data completeness (in resolution range)	(Not available) (30.00-3.13) 99.6 (61.76-3.13)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.99 (at 3.13Å)	Xtriage
Refinement program	TNT BUSTER/TNT	Depositor
R, R_{free}	0.413 , 0.433 0.446 , 0.472	Depositor DCC
R_{free} test set	775 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	84.0	Xtriage
Anisotropy	0.159	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 91.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.63	EDS
Total number of atoms	3948	wwPDB-VP
Average B, all atoms (Å^2)	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.68% 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	0.26	0/4019	0.68	4/5461 (0.1%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	85	SER	C-N-CD	-33.67	46.52	120.60
1	A	374	ALA	C-N-CD	-7.90	103.23	120.60
1	A	227	ASN	C-N-CD	-7.04	105.11	120.60
1	A	515	THR	C-N-CD	-6.37	106.58	120.60

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	3948	0	3888	629	0
All	All	3948	0	3888	629	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 80.

The worst 5 of 629 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:83:THR:HG23	1:A:98:ASN:HD21	1.10	1.16
1:A:586:LEU:HD21	1:A:588:LYS:HE2	1.31	1.11
1:A:507:ASN:HB2	1:A:581:ARG:HD2	1.36	1.08
1:A:461:ALA:HB1	1:A:470:VAL:HG13	1.36	1.07
1:A:528:PHE:HB2	1:A:530:LEU:HD11	1.38	1.04

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	489/721 (68%)	311 (64%)	108 (22%)	70 (14%)	0 1

5 of 70 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	33	PHE
1	A	57	LEU
1	A	85	SER
1	A	86	PRO
1	A	213	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.

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	451/648 (70%)	331 (73%)	120 (27%)	0 1

5 of 120 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	299	THR
1	A	554	LEU
1	A	363	ASN
1	A	550	ILE
1	A	583	MET

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

Mol	Chain	Res	Type
1	A	407	ASN
1	A	459	ASN
1	A	455	GLN
1	A	468	ASN
1	A	126	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)

There are no ligands in this entry.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

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, 95^{th} 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	505/721 (70%)	1.31	113 (22%)	0	0	22, 43, 63, 71	0

The worst 5 of 113 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	344	THR	8.1
1	A	328	ALA	6.8
1	A	48	PHE	6.8
1	A	359	SER	6.8
1	A	305	GLY	6.5

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.