Full wwPDB X-ray Structure Validation Report (i)

Jun 23, 2024 – 01:30 AM EDT

PDB ID	:	6ES1
Title	:	Crystal structure of the binding domain from botulinum neurotoxin A2 bound to extracellular domain of human receptor SV2C
Authors	:	Gustafsson, R.; Masuyer, G.; Stenmark, P.
Deposited on	:	2017-10-19
Resolution	:	2.00 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

|		
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
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)	:	2.37.1

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 2.00 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	Percentile Ranks	Value
Rfree		0.234
Clashscore		2
Ramachandran outliers		0
Sidechain outliers		1.3%
RSRZ outliers		1.2%

Worse ← → *Better*

▌ Percentile relative to all X-ray structures

◻ Percentile relative to X-ray structures of similar resolution
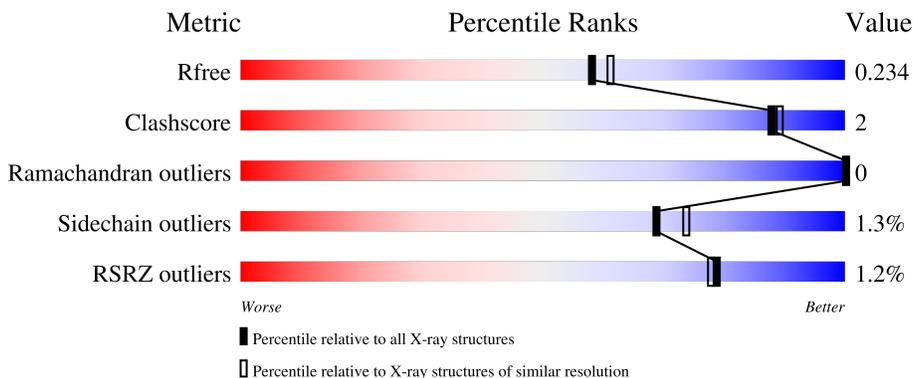
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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	445	% 89% ·· 6%
2	B	117	2% 74% 6% · 19%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	ACT	A	1301	-	-	X	-

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 4647 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 A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	420	3471	2219	587	653	12	0	4	0

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	852	MET	-	initiating methionine	UNP Q45894
A	853	GLY	-	expression tag	UNP Q45894
A	854	SER	-	expression tag	UNP Q45894
A	855	SER	-	expression tag	UNP Q45894
A	856	HIS	-	expression tag	UNP Q45894
A	857	HIS	-	expression tag	UNP Q45894
A	858	HIS	-	expression tag	UNP Q45894
A	859	HIS	-	expression tag	UNP Q45894
A	860	HIS	-	expression tag	UNP Q45894
A	861	HIS	-	expression tag	UNP Q45894
A	862	SER	-	expression tag	UNP Q45894
A	863	SER	-	expression tag	UNP Q45894
A	864	GLY	-	expression tag	UNP Q45894
A	865	GLU	-	expression tag	UNP Q45894
A	866	ASN	-	expression tag	UNP Q45894
A	867	LEU	-	expression tag	UNP Q45894
A	868	TYR	-	expression tag	UNP Q45894
A	869	PHE	-	expression tag	UNP Q45894
A	870	GLN	-	expression tag	UNP Q45894
A	871	GLY	-	expression tag	UNP Q45894
A	872	HIS	-	expression tag	UNP Q45894
A	873	MET	-	expression tag	UNP Q45894

- Molecule 2 is a protein called Synaptic vesicle glycoprotein 2C.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	95	813	520	128	159	6	0	2	0

There are 23 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	451	MET	-	initiating methionine	UNP Q496J9
B	452	HIS	-	expression tag	UNP Q496J9
B	453	HIS	-	expression tag	UNP Q496J9
B	454	HIS	-	expression tag	UNP Q496J9
B	455	HIS	-	expression tag	UNP Q496J9
B	456	HIS	-	expression tag	UNP Q496J9
B	457	HIS	-	expression tag	UNP Q496J9
B	458	SER	-	expression tag	UNP Q496J9
B	459	SER	-	expression tag	UNP Q496J9
B	460	GLY	-	expression tag	UNP Q496J9
B	461	VAL	-	expression tag	UNP Q496J9
B	462	ASP	-	expression tag	UNP Q496J9
B	463	LEU	-	expression tag	UNP Q496J9
B	464	GLY	-	expression tag	UNP Q496J9
B	465	THR	-	expression tag	UNP Q496J9
B	466	GLU	-	expression tag	UNP Q496J9
B	467	ASN	-	expression tag	UNP Q496J9
B	468	LEU	-	expression tag	UNP Q496J9
B	469	TYR	-	expression tag	UNP Q496J9
B	470	PHE	-	expression tag	UNP Q496J9
B	471	GLN	-	expression tag	UNP Q496J9
B	472	SER	-	expression tag	UNP Q496J9
B	473	MET	-	expression tag	UNP Q496J9

- Molecule 3 is ACETATE ION (three-letter code: ACT) (formula: $C_2H_3O_2$).

W O R L D W I D E
PDB
PROTEIN DATA BANK

ACT

OXT

O^-

C

CH3

O

O
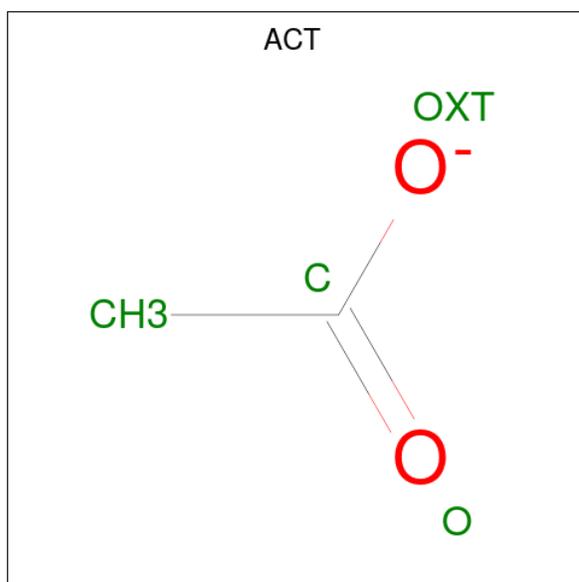
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total 4	C 2	O 2	0	0

- Molecule 4 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	299	Total 299	O 299	0	0
4	B	60	Total 60	O 60	0	0

3 Residue-property plots ⓘ

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: Botulinum neurotoxin type A
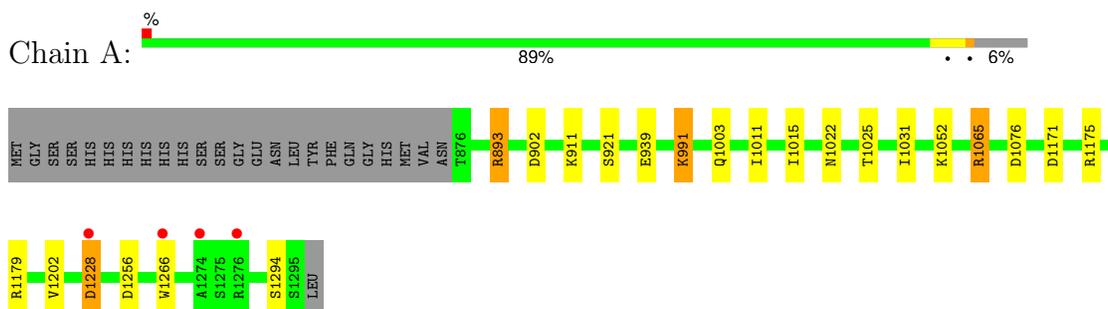
• Molecule 2: Synaptic vesicle glycoprotein 2C
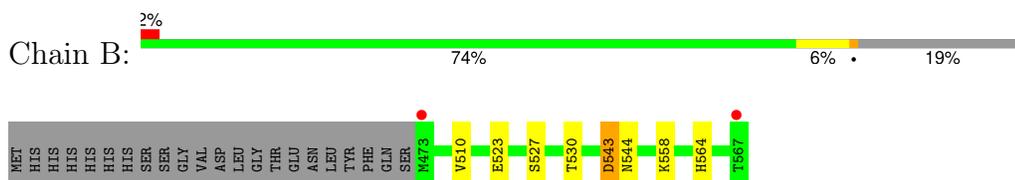
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	100.71Å 122.11Å 47.66Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.66 – 2.00 47.66 – 2.00	Depositor EDS
% Data completeness (in resolution range)	99.8 (47.66-2.00) 99.8 (47.66-2.00)	Depositor EDS
R_{merge}	0.24	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.53 (at 2.00Å)	Xtriage
Refinement program	REFMAC 5.8.0158	Depositor
R, R_{free}	0.181 , 0.230 0.190 , 0.234	Depositor DCC
R_{free} test set	2010 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å^2)	22.0	Xtriage
Anisotropy	0.262	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 48.3	EDS
L-test for twinning[2]	$<|L|> = 0.47, <L^2> = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	4647	wwPDB-VP
Average B, all atoms (Å^2)	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.74% 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 ⓘ

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

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	0.75	1/3540 (0.0%)	0.90	7/4782 (0.1%)
2	B	0.77	0/834	0.88	1/1120 (0.1%)
All	All	0.75	1/4374 (0.0%)	0.90	8/5902 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	939	GLU	CD-OE1	5.36	1.31	1.25

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	1171	ASP	CB-CG-OD1	7.74	125.26	118.30
1	A	1076	ASP	CB-CG-OD2	-7.35	111.68	118.30
1	A	1175	ARG	NE-CZ-NH1	7.34	123.97	120.30
1	A	1076	ASP	CB-CG-OD1	6.05	123.74	118.30
1	A	893	ARG	NE-CZ-NH2	6.01	123.31	120.30
2	B	543	ASP	CB-CG-OD1	5.92	123.63	118.30
1	A	1179	ARG	NE-CZ-NH1	5.60	123.10	120.30
1	A	1065	ARG	NE-CZ-NH2	5.58	123.09	120.30

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	3471	0	3455	12	0
2	B	813	0	746	4	0
3	A	4	0	3	2	0
4	A	299	0	0	2	0
4	B	60	0	0	0	0
All	All	4647	0	4204	17	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 2.

All (17) 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:911[B]:LYS:H	1:A:911[B]:LYS:HD3	1.18	1.06
1:A:911[B]:LYS:HD3	1:A:911[B]:LYS:N	1.89	0.85
1:A:911[B]:LYS:H	1:A:911[B]:LYS:CD	1.97	0.76
1:A:902:ASP:HB2	1:A:921:SER:HB2	1.90	0.53
1:A:1202:VAL:HG21	1:A:1266:TRP:CZ2	2.51	0.46
1:A:1022:ASN:HB3	1:A:1025:THR:OG1	2.17	0.45
2:B:523:GLU:HA	2:B:543:ASP:O	2.17	0.44
1:A:1015:ILE:HG23	1:A:1031:ILE:HG23	2.00	0.44
1:A:1228:ASP:N	1:A:1228:ASP:OD1	2.51	0.44
1:A:1294:SER:O	3:A:1301:ACT:H1	2.18	0.43
1:A:1003:GLN:HA	1:A:1011:ILE:HD11	2.00	0.43
2:B:527:SER:HB2	2:B:530:THR:HB	2.01	0.43
2:B:510:VAL:O	2:B:530:THR:HA	2.19	0.42
2:B:544:ASN:HA	2:B:564:HIS:O	2.20	0.42
1:A:991:LYS:N	1:A:991:LYS:HD2	2.35	0.42
3:A:1301:ACT:H2	4:A:1607:HOH:O	2.19	0.41
1:A:1052:LYS:HE2	4:A:1544:HOH:O	2.20	0.41

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	422/445 (95%)	404 (96%)	18 (4%)	0	100	100
2	B	95/117 (81%)	93 (98%)	2 (2%)	0	100	100
All	All	517/562 (92%)	497 (96%)	20 (4%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains ⓘ

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	388/406 (96%)	383 (99%)	5 (1%)	69	74
2	B	93/111 (84%)	92 (99%)	1 (1%)	73	78
All	All	481/517 (93%)	475 (99%)	6 (1%)	69	76

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	893	ARG
1	A	991	LYS
1	A	1065	ARG
1	A	1228	ASP
1	A	1256	ASP
2	B	558	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 ⓘ

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#|Z| > 2$	Counts	RMSZ	$\#|Z| > 2$
3	ACT	A	1301	-	3,3,3	0.86	0	3,3,3	0.61	0

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.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1301	ACT	2	0

5.7 Other polymers ⓘ

There are no such residues in this entry.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

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, 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	420/445 (94%)	-0.38	4 (0%)	82	81	15, 26, 45, 78	0
2	B	95/117 (81%)	-0.33	2 (2%)	63	62	16, 25, 44, 64	0
All	All	515/562 (91%)	-0.37	6 (1%)	79	78	15, 26, 45, 78	0

All (6) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1228	ASP	4.5
1	A	1276	ARG	3.1
2	B	567	THR	2.3
2	B	473	MET	2.1
1	A	1274	ALA	2.0
1	A	1266	TRP	2.0

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 ⓘ

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(Å^2)	Q<0.9
3	ACT	A	1301	4/4	0.92	0.12	37,37,40,43	0

6.5 Other polymers ⓘ

There are no such residues in this entry.