



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 19, 2016 – 09:49 PM GMT

PDB ID : 5BQM  
Title : Crystal structure of SXN101959, a Clostridium botulinum neurotoxin type D derivative and targeted secretion inhibitor  
Authors : Masuyer, G.; Davies, J.R.; Moore, K.; Chaddock, J.A.; Acharya, K.R.  
Deposited on : 2015-05-29  
Resolution : 3.10 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](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/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  
Mogul : unknown  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026982  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20026982

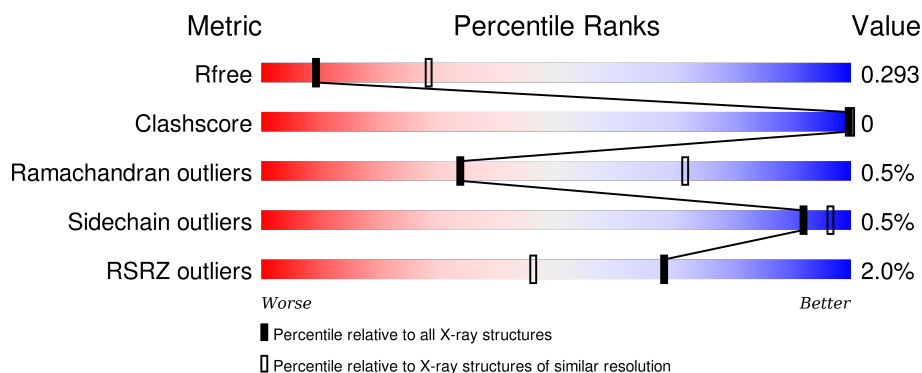
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.10 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))
$R_{free}$	91344	1114 (3.14-3.06)
Clashscore	102246	1222 (3.14-3.06)
Ramachandran outliers	100387	1174 (3.14-3.06)
Sidechain outliers	100360	1174 (3.14-3.06)
RSRZ outliers	91569	1119 (3.14-3.06)

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  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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	453	<div> <div> <div></div> <div>97%</div> <div>..</div> </div> </div>
1	C	453	<div> <div> <div>4%</div> <div>97%</div> <div>..</div> </div> </div>
2	B	474	<div> <div>81%</div> <div>18%</div> </div>
2	D	474	<div> <div> <div>2%</div> <div>80%</div> <div>18%</div> </div> </div>

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 13263 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 D.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	442	Total	C	N	O	S	0	0	0
			3555	2279	586	683	7			
1	C	443	Total	C	N	O	S	0	0	0
			3550	2274	586	683	7			

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	438	VAL	-	expression tag	UNP P19321
A	439	ASP	-	expression tag	UNP P19321
A	440	GLY	-	expression tag	UNP P19321
A	441	ILE	-	expression tag	UNP P19321
A	442	ILE	-	expression tag	UNP P19321
A	443	THR	-	expression tag	UNP P19321
A	444	SER	-	expression tag	UNP P19321
A	445	LYS	-	expression tag	UNP P19321
A	446	THR	-	expression tag	UNP P19321
A	447	LYS	-	expression tag	UNP P19321
A	448	SER	-	expression tag	UNP P19321
A	449	LEU	-	expression tag	UNP P19321
A	450	ILE	-	expression tag	UNP P19321
A	451	GLU	-	expression tag	UNP P19321
A	452	GLY	-	expression tag	UNP P19321
A	453	ARG	-	expression tag	UNP P19321
C	438	VAL	-	expression tag	UNP P19321
C	439	ASP	-	expression tag	UNP P19321
C	440	GLY	-	expression tag	UNP P19321
C	441	ILE	-	expression tag	UNP P19321
C	442	ILE	-	expression tag	UNP P19321
C	443	THR	-	expression tag	UNP P19321
C	444	SER	-	expression tag	UNP P19321
C	445	LYS	-	expression tag	UNP P19321
C	446	THR	-	expression tag	UNP P19321

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	447	LYS	-	expression tag	UNP P19321
C	448	SER	-	expression tag	UNP P19321
C	449	LEU	-	expression tag	UNP P19321
C	450	ILE	-	expression tag	UNP P19321
C	451	GLU	-	expression tag	UNP P19321
C	452	GLY	-	expression tag	UNP P19321
C	453	ARG	-	expression tag	UNP P19321

- Molecule 2 is a protein called Somatoliberin,Botulinum neurotoxin type D.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	387	Total	C	N	O	S	0	1	0
			3082	1983	488	600	11			
2	D	388	Total	C	N	O	S	0	0	0
			3067	1968	489	600	10			

There are 58 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	454	HIS	TYR	conflict	UNP P01286
B	455	VAL	ALA	conflict	UNP P01286
B	461	GLN	ASN	conflict	UNP P01286
B	468	ALA	GLY	conflict	UNP P01286
B	480	LEU	MET	conflict	UNP P01286
B	481	ASN	SER	conflict	UNP P01286
B	487	ARG	SER	conflict	UNP P01286
B	491	GLN	ARG	conflict	UNP P01286
B	494	LEU	ARG	conflict	UNP P01286
B	496	GLY	-	linker	UNP P01286
B	497	GLY	-	linker	UNP P01286
B	498	GLY	-	linker	UNP P01286
B	499	GLY	-	linker	UNP P01286
B	500	SER	-	linker	UNP P01286
B	501	GLY	-	linker	UNP P01286
B	502	GLY	-	linker	UNP P01286
B	503	GLY	-	linker	UNP P01286
B	504	GLY	-	linker	UNP P01286
B	505	SER	-	linker	UNP P01286
B	506	GLY	-	linker	UNP P01286
B	507	GLY	-	linker	UNP P01286
B	508	GLY	-	linker	UNP P01286
B	509	GLY	-	linker	UNP P01286

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	510	SER	-	linker	UNP P01286
B	511	ALA	-	linker	UNP P01286
B	512	LEU	-	linker	UNP P01286
B	513	VAL	-	linker	UNP P01286
B	514	LEU	-	linker	UNP P01286
B	515	GLN	-	linker	UNP P01286
D	454	HIS	TYR	conflict	UNP P01286
D	455	VAL	ALA	conflict	UNP P01286
D	461	GLN	ASN	conflict	UNP P01286
D	468	ALA	GLY	conflict	UNP P01286
D	480	LEU	MET	conflict	UNP P01286
D	481	ASN	SER	conflict	UNP P01286
D	487	ARG	SER	conflict	UNP P01286
D	491	GLN	ARG	conflict	UNP P01286
D	494	LEU	ARG	conflict	UNP P01286
D	496	GLY	-	linker	UNP P01286
D	497	GLY	-	linker	UNP P01286
D	498	GLY	-	linker	UNP P01286
D	499	GLY	-	linker	UNP P01286
D	500	SER	-	linker	UNP P01286
D	501	GLY	-	linker	UNP P01286
D	502	GLY	-	linker	UNP P01286
D	503	GLY	-	linker	UNP P01286
D	504	GLY	-	linker	UNP P01286
D	505	SER	-	linker	UNP P01286
D	506	GLY	-	linker	UNP P01286
D	507	GLY	-	linker	UNP P01286
D	508	GLY	-	linker	UNP P01286
D	509	GLY	-	linker	UNP P01286
D	510	SER	-	linker	UNP P01286
D	511	ALA	-	linker	UNP P01286
D	512	LEU	-	linker	UNP P01286
D	513	VAL	-	linker	UNP P01286
D	514	LEU	-	linker	UNP P01286
D	515	GLN	-	linker	UNP P01286

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total Zn 1 1	0	0
3	C	1	Total Zn 1 1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	5	Total 5	O 5	0	0
4	B	1	Total 1	O 1	0	0
4	D	1	Total 1	O 1	0	0

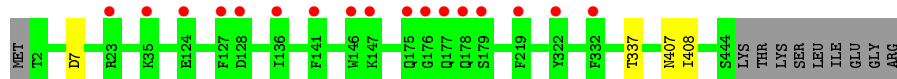
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

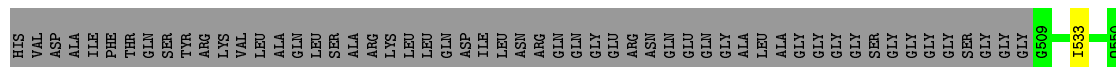
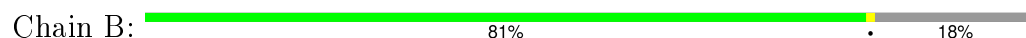
- Molecule 1: Botulinum neurotoxin type D



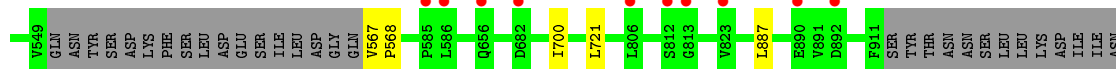
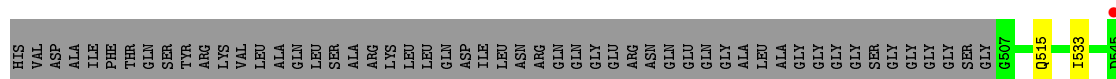
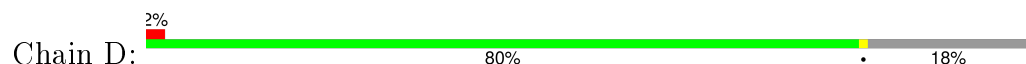
- Molecule 1: Botulinum neurotoxin type D



- Molecule 2: Somatoliberin,Botulinum neurotoxin type D



- Molecule 2: Somatoliberin,Botulinum neurotoxin type D



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	88.20Å 143.92Å 172.76Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	110.00 – 3.10 110.58 – 3.10	Depositor EDS
% Data completeness (in resolution range)	87.9 (110.00-3.10) 87.9 (110.58-3.10)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.35 (at 3.13Å)	Xtriage
Refinement program	REFMAC 5.8.0107	Depositor
R, $R_{free}$	0.250 , 0.295 0.247 , 0.293	Depositor DCC
$R_{free}$ test set	1739 reflections (5.11%)	DCC
Wilson B-factor (Å <sup>2</sup> )	84.1	Xtriage
Anisotropy	0.444	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 43.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Outliers	1 of 35752 reflections (0.003%)	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	13263	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	102.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.14% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.36	0/3641	0.51	0/4947
1	C	0.37	0/3636	0.50	0/4942
2	B	0.37	0/3144	0.49	0/4271
2	D	0.37	0/3126	0.47	0/4246
All	All	0.37	0/13547	0.49	0/18406

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 [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	3555	0	3499	0	0
1	C	3550	0	3480	0	0
2	B	3082	0	3036	0	0
2	D	3067	0	2998	3	0
3	A	1	0	0	0	0
3	C	1	0	0	0	0
4	A	5	0	0	0	0
4	B	1	0	0	0	0
4	D	1	0	0	0	0
All	All	13263	0	13013	3	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 0.

All (3) 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)
2:D:567:VAL:N	2:D:568:PRO:CD	2.76	0.49
2:D:700:ILE:HG21	2:D:721:LEU:HD23	2.02	0.40
2:D:567:VAL:N	2:D:568:PRO:HD3	2.37	0.40

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	440/453 (97%)	417 (95%)	21 (5%)	2 (0%)	34	72
1	C	441/453 (97%)	408 (92%)	31 (7%)	2 (0%)	34	72
2	B	384/474 (81%)	360 (94%)	22 (6%)	2 (0%)	34	72
2	D	384/474 (81%)	359 (94%)	22 (6%)	3 (1%)	24	63
All	All	1649/1854 (89%)	1544 (94%)	96 (6%)	9 (0%)	34	72

All (9) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	442	ILE
2	B	569	ILE
1	C	408	ILE
1	C	407	ASN
2	D	515	GLN
2	D	887	LEU
2	B	533	ILE
1	A	438	VAL

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
2	D	533	ILE

### 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	409/421 (97%)	407 (100%)	2 (0%)	92	96
1	C	407/421 (97%)	405 (100%)	2 (0%)	92	96
2	B	346/424 (82%)	343 (99%)	3 (1%)	84	94
2	D	341/424 (80%)	341 (100%)	0	100	100
All	All	1503/1690 (89%)	1496 (100%)	7 (0%)	92	96

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	333	ASP
1	A	439	ASP
2	B	569	ILE
2	B	746	GLU
2	B	884	ASN
1	C	7	ASP
1	C	337	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	884	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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 2 ligands modelled in this entry, 2 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 [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<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	442/453 (97%)	-0.11	4 (0%) 85 72	61, 90, 118, 140	0
1	C	443/453 (97%)	0.32	17 (3%) 44 21	76, 109, 165, 185	0
2	B	387/474 (81%)	-0.05	2 (0%) 91 83	60, 94, 151, 167	0
2	D	388/474 (81%)	0.12	11 (2%) 56 32	58, 100, 161, 170	0
All	All	1660/1854 (89%)	0.07	34 (2%) 68 46	58, 98, 157, 185	0

All (34) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	176	GLY	7.3
1	C	177	GLN	6.7
1	C	175	GLN	4.8
2	D	586	LEU	3.6
1	C	146	TRP	3.4
1	C	127	PHE	3.1
1	C	124	GLU	3.1
2	D	585	PRO	2.9
1	C	332	PHE	2.8
2	D	823	VAL	2.8
1	A	142	GLU	2.8
1	C	136	ILE	2.7
2	D	892	ASP	2.7
1	C	178	GLN	2.6
2	D	682	ASP	2.6
1	C	147	LYS	2.5
2	D	656	GLN	2.5
1	C	219	PHE	2.4
2	D	545	ASP	2.4
1	C	141	PHE	2.4
2	D	890	GLU	2.3

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	B	743	GLN	2.3
1	A	336	ASN	2.3
1	C	322	TYR	2.2
1	C	128	ASP	2.2
1	A	2	THR	2.2
2	B	657	ALA	2.2
2	D	813	GLY	2.1
1	C	35	LYS	2.1
1	C	23	ARG	2.1
2	D	812	SER	2.1
2	D	806	LEU	2.0
1	C	179	SER	2.0
1	A	175	GLN	2.0

## 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
3	ZN	C	1000	1/1	0.98	0.13	-1.60	85,85,85,85	0
3	ZN	A	1000	1/1	0.99	0.16	-2.11	64,64,64,64	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.