



Full wwPDB X-ray Structure Validation Report ⓘ

Nov 12, 2024 – 12:20 pm GMT

PDB ID : 8S83
Title : Domains 3 and 6 of Sap S-layer protein from Bacillus anthracis
Authors : Sogues, A.; Remaut, H.
Deposited on : 2024-03-05
Resolution : 2.07 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 : 3.0
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4 : 9.0.003 (Gargrove)
Density-Fitness : 1.0.11
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.39

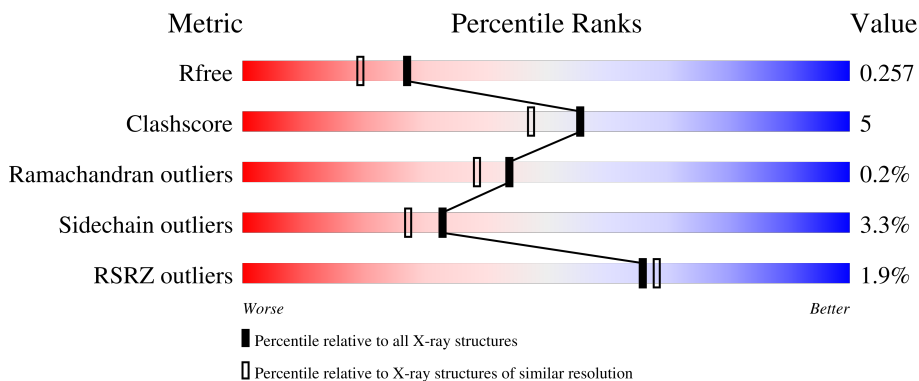
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

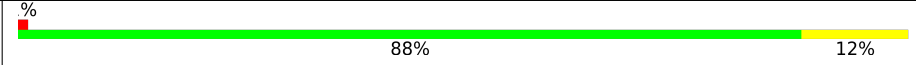
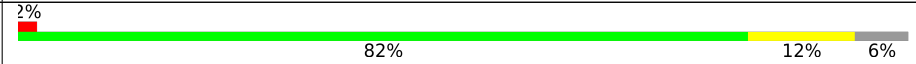
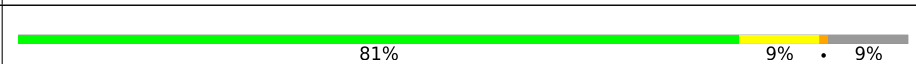

The reported resolution of this entry is 2.07 Å.

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}	164625	3436 (2.08-2.04)
Clashscore	180529	3661 (2.08-2.04)
Ramachandran outliers	177936	3649 (2.08-2.04)
Sidechain outliers	177891	3649 (2.08-2.04)
RSRZ outliers	164620	3436 (2.08-2.04)

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 $\leq 5\%$. 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	113	 88% 12%
1	B	113	 82% 12% 6%
2	C	109	 81% 9% • 9%
2	D	109	 82% 9% • 7%

2 Entry composition i

There are 3 unique types of molecules in this entry. The entry contains 3424 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 S-layer protein sap.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
1	A	113	869	545	150	174	0	0	0
1	B	106	809	511	133	165	0	1	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	379	ALA	-	expression tag	UNP P49051
A	380	HIS	-	expression tag	UNP P49051
A	381	HIS	-	expression tag	UNP P49051
A	382	HIS	-	expression tag	UNP P49051
A	383	HIS	-	expression tag	UNP P49051
A	384	HIS	-	expression tag	UNP P49051
B	379	ALA	-	expression tag	UNP P49051
B	380	HIS	-	expression tag	UNP P49051
B	381	HIS	-	expression tag	UNP P49051
B	382	HIS	-	expression tag	UNP P49051
B	383	HIS	-	expression tag	UNP P49051
B	384	HIS	-	expression tag	UNP P49051

- Molecule 2 is a protein called S-layer protein sap.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	C	99	713	450	117	146	0	0	0
2	D	101	724	456	120	148	0	0	0

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	703	MET	-	initiating methionine	UNP P49051
C	704	HIS	-	expression tag	UNP P49051
C	705	HIS	-	expression tag	UNP P49051
C	706	HIS	-	expression tag	UNP P49051
C	707	HIS	-	expression tag	UNP P49051
C	708	HIS	-	expression tag	UNP P49051
C	709	HIS	-	expression tag	UNP P49051
D	703	MET	-	initiating methionine	UNP P49051
D	704	HIS	-	expression tag	UNP P49051
D	705	HIS	-	expression tag	UNP P49051
D	706	HIS	-	expression tag	UNP P49051
D	707	HIS	-	expression tag	UNP P49051
D	708	HIS	-	expression tag	UNP P49051
D	709	HIS	-	expression tag	UNP P49051

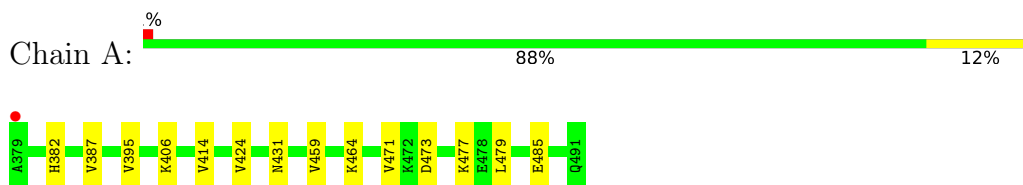
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	90	Total O 90 90	0	0
3	B	86	Total O 86 86	0	0
3	C	86	Total O 86 86	0	0
3	D	47	Total O 47 47	0	0

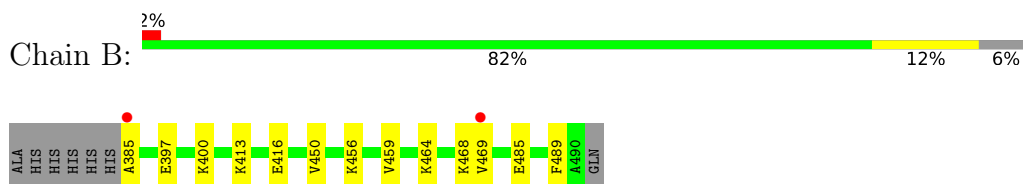
3 Residue-property plots [i](#)

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.

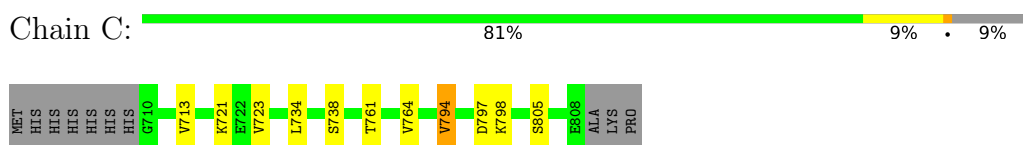
- Molecule 1: S-layer protein sap



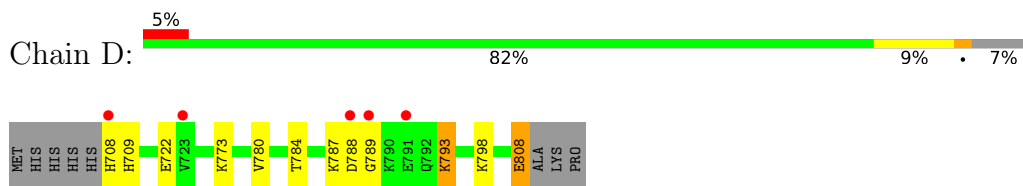
- Molecule 1: S-layer protein sap



- Molecule 2: S-layer protein sap



- Molecule 2: S-layer protein sap



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	40.15Å 61.01Å 97.66Å 90.00° 91.50° 90.00°	Depositor
Resolution (Å)	51.74 – 2.07 51.74 – 2.07	Depositor EDS
% Data completeness (in resolution range)	99.7 (51.74-2.07) 99.7 (51.74-2.07)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.21 (at 2.07Å)	Xtrriage
Refinement program	PHENIX 1.20_4459	Depositor
R, R_{free}	0.201 , 0.258 0.202 , 0.257	Depositor DCC
R_{free} test set	1347 reflections (4.64%)	wwPDB-VP
Wilson B-factor (Å ²)	39.5	Xtrriage
Anisotropy	0.309	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 40.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.032 for h,-k,-l	Xtrriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	3424	wwPDB-VP
Average B, all atoms (Å ²)	46.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 35.44 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 5.8329e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

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 [i](#)

5.1 Standard geometry [i](#)

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	0.41	0/883	0.60	0/1195
1	B	0.43	0/818	0.60	0/1107
2	C	0.44	0/718	0.61	0/976
2	D	0.46	0/730	0.66	0/994
All	All	0.44	0/3149	0.62	0/4272

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	869	0	858	7	0
1	B	809	0	816	8	0
2	C	713	0	740	8	0
2	D	724	0	738	6	0
3	A	90	0	0	2	1
3	B	86	0	0	2	0
3	C	86	0	0	2	0
3	D	47	0	0	3	1
All	All	3424	0	3152	29	1

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 5.

All (29) 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:C:713:VAL:HG21	2:C:794:VAL:HG11	1.63	0.78
2:C:798:LYS:NZ	3:C:901:HOH:O	2.13	0.72
2:D:808:GLU:O	3:D:901:HOH:O	2.13	0.66
1:A:382:HIS:H	1:A:382:HIS:CD2	2.15	0.63
2:D:798:LYS:NZ	3:D:902:HOH:O	2.24	0.61
1:B:385:ALA:N	3:B:502:HOH:O	2.39	0.56
1:A:395:VAL:HG11	1:A:414:VAL:HG22	1.91	0.52
2:C:723:VAL:HG12	2:C:805:SER:O	2.10	0.52
1:B:397:GLU:HB2	1:B:400:LYS:HD3	1.92	0.50
1:A:473:ASP:HB3	1:A:479:LEU:HD21	1.94	0.48
1:A:387:VAL:HG11	1:A:471:VAL:HG11	1.96	0.48
2:C:721:LYS:HE3	2:C:721:LYS:HB2	1.64	0.47
2:D:787:LYS:O	2:D:789:GLY:N	2.47	0.47
2:C:734:LEU:O	3:C:902:HOH:O	2.21	0.46
1:A:431:ASN:ND2	3:A:501:HOH:O	2.29	0.46
2:D:784:THR:HG23	2:D:793:LYS:HA	1.97	0.46
2:D:708:HIS:N	3:D:903:HOH:O	2.49	0.45
2:D:784:THR:HG23	2:D:793:LYS:HB2	1.99	0.45
1:A:477:LYS:NZ	3:A:508:HOH:O	2.49	0.45
1:B:416:GLU:HG3	1:B:459:VAL:HG22	2.01	0.43
1:B:450:VAL:HA	1:B:456:LYS:O	2.19	0.43
1:B:413:LYS:HB3	1:B:489:PHE:HE2	1.83	0.43
2:C:761:THR:HA	2:C:764:VAL:O	2.18	0.42
1:B:468:LYS:NZ	3:B:503:HOH:O	2.44	0.42
2:C:713:VAL:HA	2:C:738:SER:O	2.20	0.42
1:B:464:LYS:HD3	1:B:485:GLU:OE1	2.19	0.41
2:C:797:ASP:N	2:C:797:ASP:OD1	2.54	0.41
1:B:400:LYS:HE2	1:B:400:LYS:HB3	1.69	0.41
1:A:464:LYS:HD2	1:A:485:GLU:OE1	2.21	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:A:550:HOH:O	3:D:937:HOH:O[1_545]	2.06	0.14

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	111/113 (98%)	109 (98%)	2 (2%)	0	100	100
1	B	105/113 (93%)	104 (99%)	1 (1%)	0	100	100
2	C	97/109 (89%)	96 (99%)	1 (1%)	0	100	100
2	D	99/109 (91%)	96 (97%)	2 (2%)	1 (1%)	13	5
All	All	412/444 (93%)	405 (98%)	6 (2%)	1 (0%)	44	38

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	D	788	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	94/94 (100%)	91 (97%)	3 (3%)	34	28
1	B	88/94 (94%)	86 (98%)	2 (2%)	45	41
2	C	78/88 (89%)	77 (99%)	1 (1%)	65	65
2	D	78/88 (89%)	72 (92%)	6 (8%)	10	4
All	All	338/364 (93%)	326 (96%)	12 (4%)	33	24

All (12) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	406	LYS
1	A	424	VAL
1	A	459	VAL
1	B	469[A]	VAL
1	B	469[B]	VAL
2	C	794	VAL
2	D	709	HIS
2	D	722	GLU
2	D	773	LYS
2	D	780	VAL
2	D	793	LYS
2	D	808	GLU

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

Mol	Chain	Res	Type
1	A	382	HIS

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 oligosaccharides 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

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	113/113 (100%)	-0.02	1 (0%) 81 83	33, 43, 61, 82	0
1	B	106/113 (93%)	-0.08	2 (1%) 66 68	16, 41, 56, 72	1 (0%)
2	C	99/109 (90%)	-0.05	0 100 100	31, 41, 56, 70	0
2	D	101/109 (92%)	0.41	5 (4%) 35 37	35, 49, 75, 86	0
All	All	419/444 (94%)	0.06	8 (1%) 66 68	16, 43, 66, 86	1 (0%)

All (8) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	379	ALA	3.9
2	D	788	ASP	3.2
2	D	791	GLU	2.6
1	B	469[A]	VAL	2.5
2	D	708	HIS	2.4
1	B	385	ALA	2.3
2	D	789	GLY	2.1
2	D	723	VAL	2.1

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 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.