W O R L D W I D E
wwPDB
P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 12, 2024 – 12:21 pm GMT

PDB ID	:	8RX2
Title	:	Domains 1 and 2 of Sap S-layer protein from Bacillus anthracis
Authors	:	Sogues, A.; Remaut, H.
Deposited on	:	2024-02-06
Resolution	:	2.01 Å(reported)

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

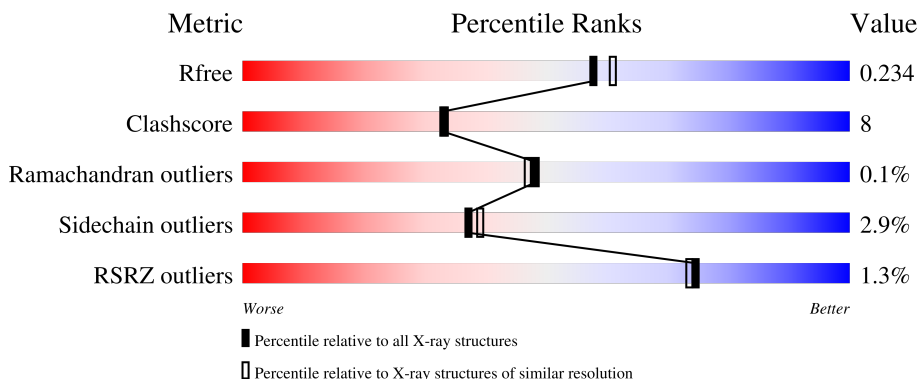|		
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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:
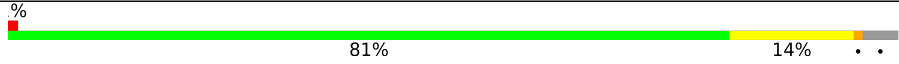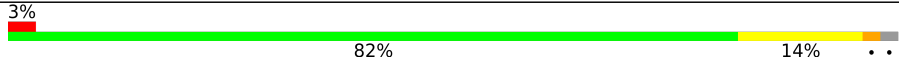*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.01 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		8
Ramachandran outliers		0.1%
Sidechain outliers		2.9%
RSRZ outliers		1.3%

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}	164625	9409 (2.00-2.00)
Clashscore	180529	10737 (2.00-2.00)
Ramachandran outliers	177936	10628 (2.00-2.00)
Sidechain outliers	177891	10627 (2.00-2.00)
RSRZ outliers	164620	9409 (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	177	84% 11% • •
1	B	177	% 86% 9% • •
1	C	177	% 81% 14% • •
1	D	177	3% 82% 14% • •

2   Entity composition (i)

There are 2 unique types of molecules in this entry. The entry contains 5657 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
1	A	171	Total	C	N	O	Se	0	2	0
			1280	800	210	269	1			
1	B	170	Total	C	N	O	Se	0	0	0
			1255	787	206	261	1			
1	C	170	Total	C	N	O	Se	0	0	0
			1240	779	204	256	1			
1	D	174	Total	C	N	O	Se	0	0	0
			1289	807	216	265	1			

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	208	MSE	-	initiating methionine	UNP P49051
A	209	HIS	-	expression tag	UNP P49051
A	210	HIS	-	expression tag	UNP P49051
A	211	HIS	-	expression tag	UNP P49051
A	212	HIS	-	expression tag	UNP P49051
A	213	HIS	-	expression tag	UNP P49051
A	214	HIS	-	expression tag	UNP P49051
B	208	MSE	-	initiating methionine	UNP P49051
B	209	HIS	-	expression tag	UNP P49051
B	210	HIS	-	expression tag	UNP P49051
B	211	HIS	-	expression tag	UNP P49051
B	212	HIS	-	expression tag	UNP P49051
B	213	HIS	-	expression tag	UNP P49051
B	214	HIS	-	expression tag	UNP P49051
C	208	MSE	-	initiating methionine	UNP P49051
C	209	HIS	-	expression tag	UNP P49051
C	210	HIS	-	expression tag	UNP P49051
C	211	HIS	-	expression tag	UNP P49051
C	212	HIS	-	expression tag	UNP P49051
C	213	HIS	-	expression tag	UNP P49051
C	214	HIS	-	expression tag	UNP P49051

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	208	MSE	-	initiating methionine	UNP P49051
D	209	HIS	-	expression tag	UNP P49051
D	210	HIS	-	expression tag	UNP P49051
D	211	HIS	-	expression tag	UNP P49051
D	212	HIS	-	expression tag	UNP P49051
D	213	HIS	-	expression tag	UNP P49051
D	214	HIS	-	expression tag	UNP P49051

• Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	156	Total 156	O 156	0	0
2	B	152	Total 152	O 152	0	0
2	C	127	Total 127	O 127	0	0
2	D	158	Total 158	O 158	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: S-layer protein sap

Chain A:

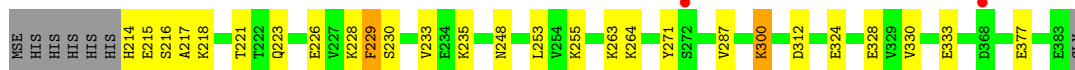- Molecule 1: S-layer protein sap

Chain B:

- Molecule 1: S-layer protein sap

Chain C:

- Molecule 1: S-layer protein sap

Chain D:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	48.44Å 128.61Å 55.13Å 90.00° 95.76° 90.00°	Depositor
Resolution (Å)	41.73 – 2.01 41.73 – 2.01	Depositor EDS
% Data completeness (in resolution range)	99.9 (41.73-2.01) 99.9 (41.73-2.01)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.24 (at 2.01Å)	Xtriage
Refinement program	PHENIX (1.20_4459: ???)	Depositor
R, R_{free}	0.201 , 0.237 0.201 , 0.234	Depositor DCC
R_{free} test set	2254 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å^2)	33.5	Xtriage
Anisotropy	0.421	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 38.8	EDS
L-test for twinning [2]	$< |L| > = 0.51, < L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	5657	wwPDB-VP
Average B, all atoms (Å^2)	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 14.78% 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 (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	0.39	0/1292	0.57	0/1748
1	B	0.41	0/1262	0.59	0/1705
1	C	0.38	0/1247	0.59	0/1687
1	D	0.41	0/1299	0.60	0/1755
All	All	0.40	0/5100	0.59	0/6895

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	1280	0	1307	19	0
1	B	1255	0	1298	16	0
1	C	1240	0	1275	20	0
1	D	1289	0	1322	24	0
2	A	156	0	0	9	0
2	B	152	0	0	6	2
2	C	127	0	0	7	2
2	D	158	0	0	10	0
All	All	5657	0	5202	79	2

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 8.

The worst 5 of 79 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:250:ASP:OD1	2:A:401:HOH:O	1.83	0.96
1:C:312:ASP:OD1	2:C:401:HOH:O	1.84	0.96
1:C:221:THR:HG22	1:C:223:GLN:H	1.30	0.94
1:C:226:GLU:OE1	2:C:402:HOH:O	1.91	0.87
1:C:333:GLU:N	2:C:405:HOH:O	2.03	0.86

All (2) 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)
2:B:445:HOH:O	2:C:511:HOH:O[2_655]	2.06	0.14
2:B:550:HOH:O	2:C:515:HOH:O[2_555]	2.15	0.05

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	171/177 (97%)	164 (96%)	7 (4%)	0	100	100
1	B	168/177 (95%)	163 (97%)	5 (3%)	0	100	100
1	C	168/177 (95%)	164 (98%)	4 (2%)	0	100	100
1	D	172/177 (97%)	165 (96%)	6 (4%)	1 (1%)	22	17
All	All	679/708 (96%)	656 (97%)	22 (3%)	1 (0%)	48	47

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	214	HIS

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	142/148 (96%)	137 (96%)	5 (4%)	31	31
1	B	138/148 (93%)	135 (98%)	3 (2%)	47	51
1	C	134/148 (90%)	129 (96%)	5 (4%)	29	29
1	D	141/148 (95%)	138 (98%)	3 (2%)	48	53
All	All	555/592 (94%)	539 (97%)	16 (3%)	37	39

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	281	THR
1	D	242	LYS
1	C	215	GLU
1	C	300	LYS
1	B	252	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	285	ASN
1	D	223	GLN
1	D	273	ASN

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.

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

5.5 Carbohydrates ⓘ

There are no oligosaccharides 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 ⓘ

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(Å²)	Q<0.9
1	A	170/177 (96%)	-0.11	0	100	100	14, 32, 48, 58	2 (1%)
1	B	169/177 (95%)	-0.15	1 (0%)	85	85	21, 33, 46, 59	0
1	C	169/177 (95%)	0.02	2 (1%)	76	75	22, 34, 51, 64	0
1	D	173/177 (97%)	-0.01	6 (3%)	47	45	23, 32, 47, 59	0
All	All	681/708 (96%)	-0.06	9 (1%)	74	73	14, 33, 50, 64	2 (0%)

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	211	HIS	4.2
1	D	214	HIS	4.1
1	B	214	HIS	3.2
1	D	384	GLY	3.1
1	C	272	SER	2.9

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 (i)

There are no such residues in this entry.