wwPDB X-ray Structure Validation Summary Report (i)

Sep 20, 2023 – 02:02 PM EDT

PDB ID	:	5J44
Title	:	Crystal structure of the Secreted Extracellular protein A (SepA) from Shigella flexneri
Authors	:	Birtley, J.R.; Stern, L.J.; McCormick, B.; Maldonado-Contreras, A.
Deposited on	:	2016-03-31
Resolution	:	2.91 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:

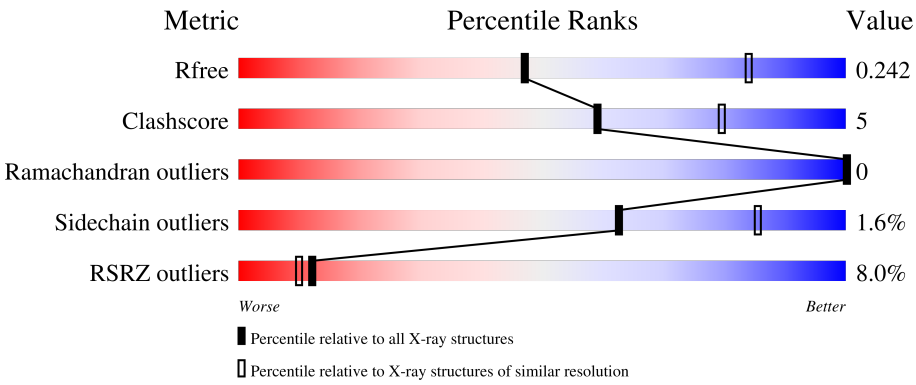
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 2.91 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	2307 (2.94-2.90)
Clashscore	141614	2531 (2.94-2.90)
Ramachandran outliers	138981	2462 (2.94-2.90)
Sidechain outliers	138945	2464 (2.94-2.90)
RSRZ outliers	127900	2248 (2.94-2.90)

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	1033	3% 84% 13% •
1	B	1033	13% 86% 11% •

2 Entry composition (i)

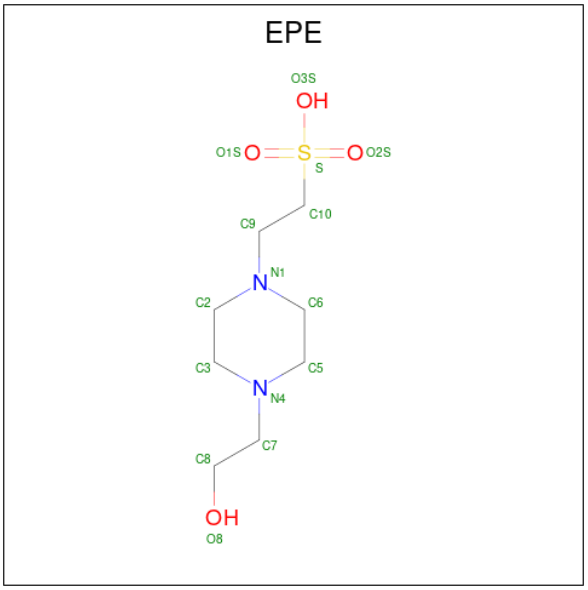
There are 3 unique types of molecules in this entry. The entry contains 15034 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Serine protease SepA autotransporter.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	1006	7500	4650	1290	1543	17	0	0	0
1	B	1006	7500	4650	1290	1543	17	0	0	0

- Molecule 2 is 4-(2-HYDROXYETHYL)-1-PIPERAZINE ETHANESULFONIC ACID (three-letter code: EPE) (formula: $C_8H_{18}N_2O_4S$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
2	A	1	15	8	2	4	1	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	12	12	12	0	0

Continued on next page...

Continued from previous page...

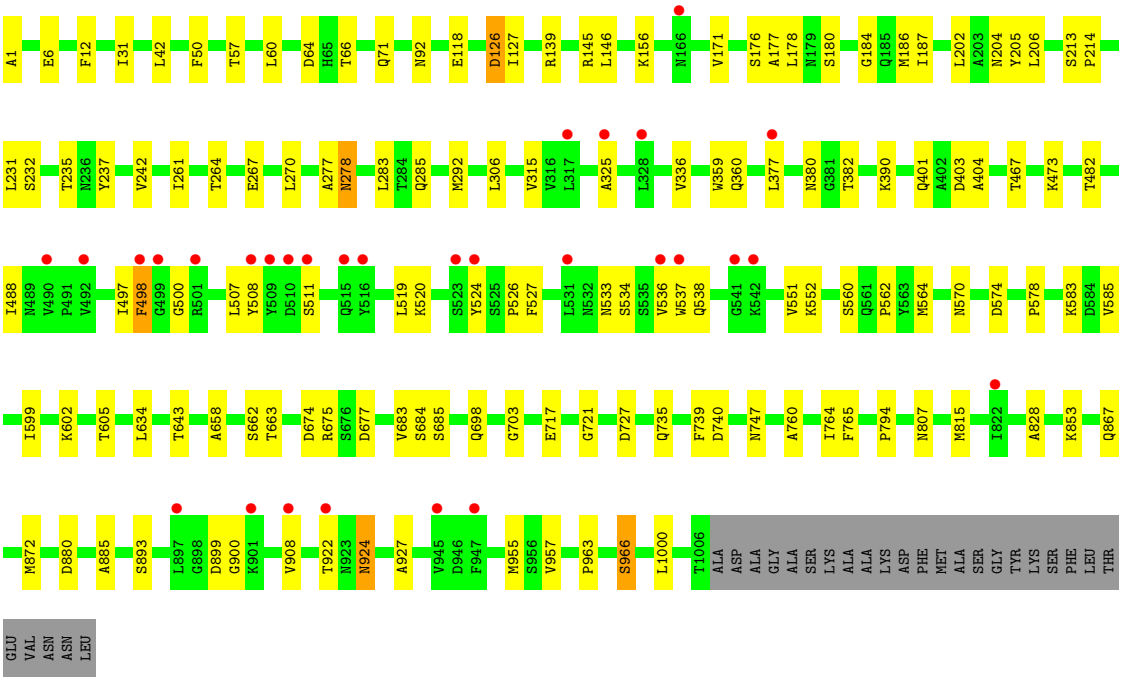
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	B	7	7	7	0	0

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.

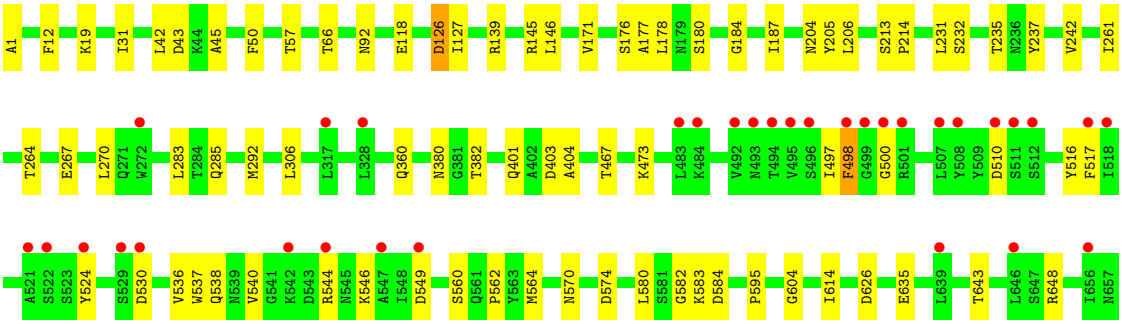
• Molecule 1: Serine protease SepA autotransporter

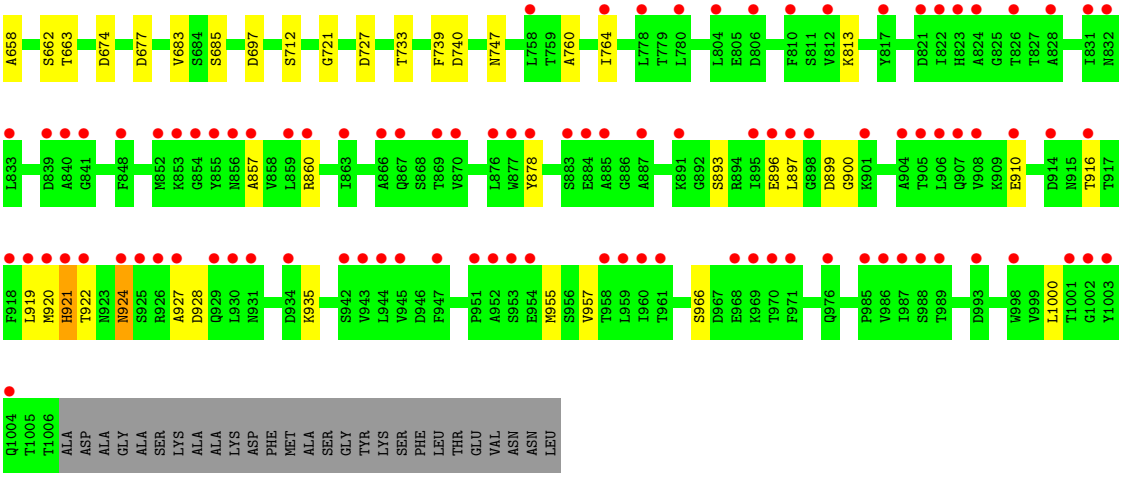
Chain A:



• Molecule 1: Serine protease SepA autotransporter

Chain B:




A658 S662 T663 D674 D677 V683 S684 S685 D697 S712 G721 D727 T733 F739 D740 N747 L758 T759 A760 I764 L778 T779 L780 L804 E805 D806 F810 S811 V812 K813 Y817 D821 I822 H823 A824 G825 T826 T827 A828 I831 N832
L833 D839 A840 G841 F848 M852 K853 G854 Y855 N856 A857 V858 L859 R860 I863 A866 Q867 S868 T869 V870 L876 W877 Y878 S883 E884 A885 G886 A887 K891 G892 S893 R894 I895 E896 L897 G898 D899 G900 K901 A904 T905 L906 Q907 V908 K909 E910 D914 N915 T916 T917
F918 L919 M920 H921 T922 N923 N924 S925 R926 A927 D928 Q929 L930 N931 D934 K935 S942 V943 L944 V945 D946 F947 P951 A952 S953 E954 M955 S956 V957 T958 L959 I960 T961 S966 D967 E968 K969 T970 F971 Q976 P985 V986 I987 S988 T989 D993 W998 V999 L1000 T1001 G1002 Y1003
Q1004 T1005 T1006 ALA ASP ALA GLY ALA SER LYS ALA ALA LYS ASP PHE MET ALA SER GLY TYR LYS SER PHE LEU THR GLU VAL ASN ASN LEU

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, α, β, γ	143.52Å 143.52Å 269.25Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	72.76 – 2.91 91.32 – 2.91	Depositor EDS
% Data completeness (in resolution range)	99.9 (72.76-2.91) 100.0 (91.32-2.91)	Depositor EDS
R_{merge}	0.32	Depositor
R_{sym}	0.30	Depositor
$< I/\sigma(I) >$ [1]	1.79 (at 2.91Å)	Xtriage
Refinement program	PHENIX (1.10_2155: ???)	Depositor
R, R_{free}	0.214 , 0.239 0.217 , 0.242	Depositor DCC
R_{free} test set	3584 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	45.3	Xtriage
Anisotropy	0.334	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 41.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	0.036 for -h,-k,l	Xtriage
F_o,F_c correlation	0.87	EDS
Total number of atoms	15034	wwPDB-VP
Average B, all atoms (Å^2)	70.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.40% 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)

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

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	0.24	0/7628	0.46	0/10353
1	B	0.24	0/7628	0.46	0/10353
All	All	0.24	0/15256	0.46	0/20706

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	7500	0	7282	75	1
1	B	7500	0	7282	65	0
2	A	15	0	17	1	0
3	A	12	0	0	1	0
3	B	7	0	0	0	0
All	All	15034	0	14581	140	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.

The worst 5 of 140 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:B:584:ASP:OD2	1:B:604:GLY:HA3	1.79	0.82
1:A:277:ALA:O	1:A:278:ASN:ND2	2.12	0.80
1:A:482:THR:HG21	1:A:560:SER:HA	1.65	0.76
1:A:498:PHE:HA	1:A:526:PRO:HB3	1.70	0.74
1:A:178:LEU:HD13	1:A:187:ILE:HD12	1.69	0.73

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)
1:A:727:ASP:OD2	1:A:853:LYS:NZ[4_556]	1.88	0.32

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	1004/1033 (97%)	965 (96%)	39 (4%)	0	100	100
1	B	1004/1033 (97%)	961 (96%)	43 (4%)	0	100	100
All	All	2008/2066 (97%)	1926 (96%)	82 (4%)	0	100	100

There are no Ramachandran outliers to report.

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	816/835 (98%)	804 (98%)	12 (2%)	65 86
1	B	816/835 (98%)	802 (98%)	14 (2%)	60 84
All	All	1632/1670 (98%)	1606 (98%)	26 (2%)	62 85

5 of 26 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	145	ARG
1	B	538	GLN
1	B	955	MET
1	B	498	PHE
1	B	540	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	624	GLN
1	B	380	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)

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	$\#\|Z\| > 2$
2	EPE	A	1101	-	15,15,15	0.82	1 (6%)	18,20,20	1.97	5 (27%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	EPE	A	1101	-	-	3/9/19/19	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1101	EPE	C10-S	2.64	1.81	1.77

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	1101	EPE	C7-N4-C3	4.08	121.67	111.23
2	A	1101	EPE	C5-N4-C3	3.95	117.73	108.83
2	A	1101	EPE	C7-N4-C5	3.62	120.50	111.23
2	A	1101	EPE	O3S-S-C10	2.64	110.04	105.77
2	A	1101	EPE	C6-N1-C2	2.50	114.47	108.83

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	1101	EPE	C10-C9-N1-C2
2	A	1101	EPE	C10-C9-N1-C6
2	A	1101	EPE	C8-C7-N4-C5

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1101	EPE	1	0

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 (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	1006/1033 (97%)	0.34	30 (2%) 50 46	28, 51, 106, 161	0
1	B	1006/1033 (97%)	0.80	131 (13%) 3 2	25, 65, 164, 212	0
All	All	2012/2066 (97%)	0.57	161 (8%) 12 10	25, 56, 150, 212	0

The worst 5 of 161 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	959	LEU	10.2
1	B	927	ALA	9.4
1	B	901	LYS	8.9
1	B	945	VAL	8.6
1	B	906	LEU	6.6

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 (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. 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
2	EPE	A	1101	15/15	0.78	0.37	49,73,151,154	0

6.5 Other polymers (i)

There are no such residues in this entry.