



# Full wwPDB X-ray Structure Validation Report ⓘ

Jul 18, 2022 – 04:09 pm BST

PDB ID : 7P4A  
Title : Non-canonical Staphylococcus aureus pathogenicity island repression.  
Authors : Miguel-Romero, L.; Alqasmi, M.; Bacarizo, J.; Tan, J.A.; Cogdell, R.J.; Chen, J.; Byron, O.; Christie, G.E.; Marina, A.; Penades, J.R.  
Deposited on : 2021-07-10  
Resolution : 2.90 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

<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  
Mogul : 1.8.4, CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.29  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0267  
CCP4 : 7.1.010 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.29

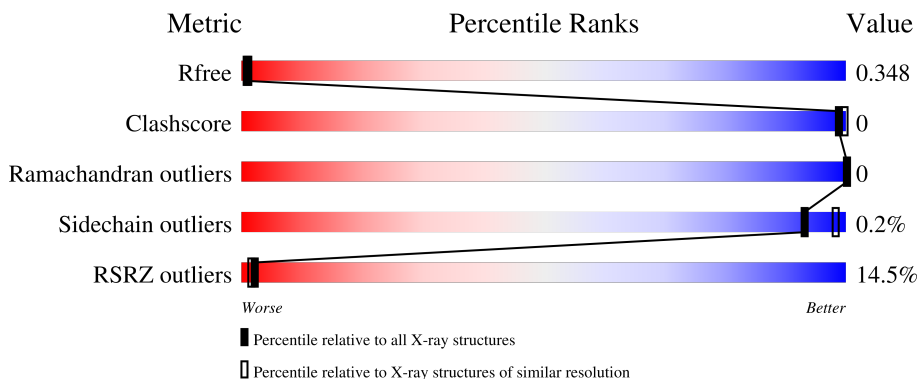
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.90 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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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  $\geq 3$ , 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	247	
1	B	247	
2	D	52	
2	E	52	

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 8342 atoms, of which 4130 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 Stl.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	Se			
1	A	246	4051	1316	2018	320	392	5	61	0	0
1	B	207	3224	1061	1581	265	316	1	74	0	0

- Molecule 2 is a protein called Sri.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace	
			Total	C	H	N	O	S				Se
2	E	49	817	258	412	67	78	1	1	8	0	0
2	D	19	242	79	119	21	23		10	0	0	

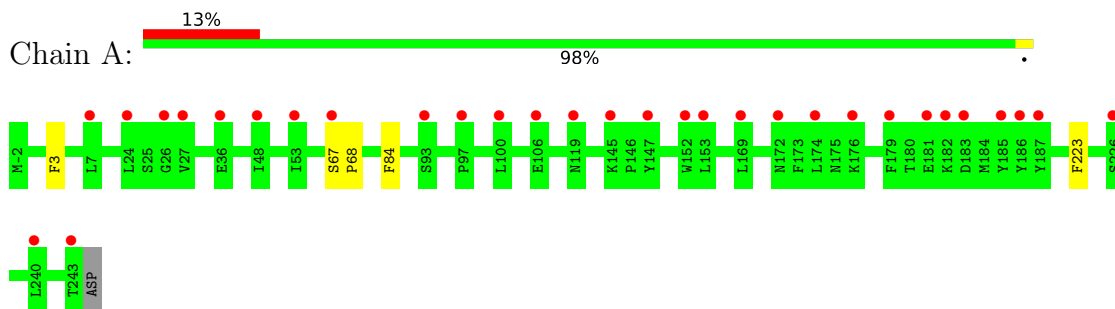
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	6	Total O 6 6	0	0
3	B	2	Total O 2 2	0	0

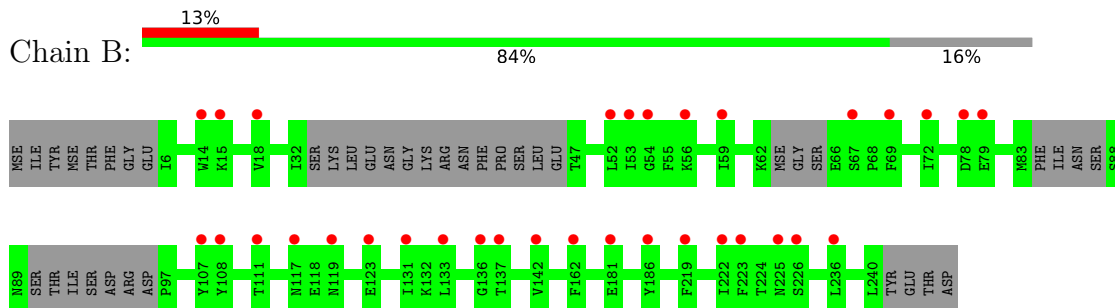
### 3 Residue-property plots [i](#)

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.

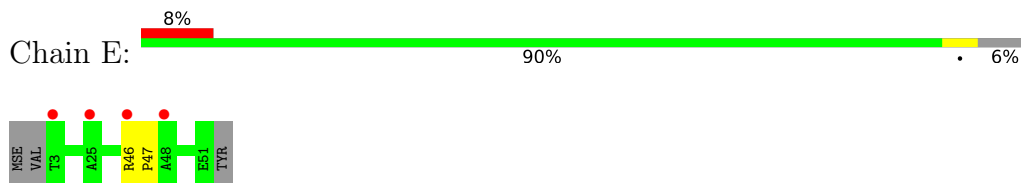
- Molecule 1: Stl



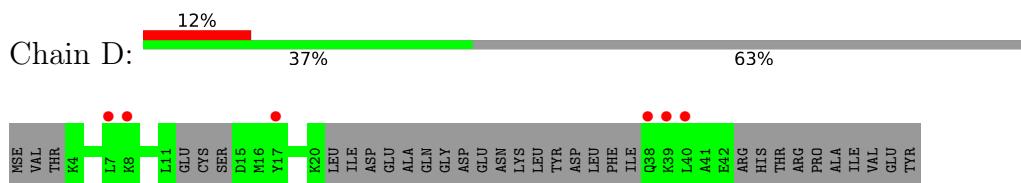
- Molecule 1: Stl



- Molecule 2: Sri



- Molecule 2: Sri



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 61 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	100.21Å 100.21Å 298.69Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	83.48 – 2.90 83.34 – 2.90	Depositor EDS
% Data completeness (in resolution range)	99.4 (83.48-2.90) 99.4 (83.34-2.90)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.54 (at 2.91Å)	Xtrriage
Refinement program	REFMAC 5.8.0267	Depositor
R, $R_{free}$	0.298 , 0.343 0.305 , 0.348	Depositor DCC
$R_{free}$ test set	1063 reflections (5.19%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	88.8	Xtrriage
Anisotropy	0.031	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	(Not available) , (Not available)	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.88	EDS
Total number of atoms	8342	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	104.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.39% 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, 0.333 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	0.69	0/2069	0.78	0/2783
1	B	0.72	0/1671	0.76	0/2250
2	D	0.75	0/120	0.74	0/156
2	E	0.67	0/409	0.77	0/545
All	All	0.70	0/4269	0.77	0/5734

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	2033	2018	1996	2	0
1	B	1643	1581	1514	0	0
2	D	123	119	96	0	0
2	E	405	412	410	1	0
3	A	6	0	0	0	0
3	B	2	0	0	0	0
All	All	4212	4130	4016	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:E:46:ARG:N	2:E:47:PRO:CD	2.76	0.49
1:A:67:SER:N	1:A:68:PRO:HD2	2.31	0.45
1:A:3:PHE:CD2	1:A:84:PHE:CD1	3.07	0.43

There are no symmetry-related clashes.

### 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	244/247 (99%)	206 (84%)	38 (16%)	0	100	100
1	B	197/247 (80%)	173 (88%)	24 (12%)	0	100	100
2	D	13/52 (25%)	11 (85%)	2 (15%)	0	100	100
2	E	47/52 (90%)	41 (87%)	6 (13%)	0	100	100
All	All	501/598 (84%)	431 (86%)	70 (14%)	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	228/235 (97%)	227 (100%)	1 (0%)	91	97
1	B	168/235 (72%)	168 (100%)	0	100	100
2	D	7/45 (16%)	7 (100%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	E	44/45 (98%)	44 (100%)	0	100	100
All	All	447/560 (80%)	446 (100%)	1 (0%)	93	98

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	223	PHE

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

### 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](#)

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 [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	241/247 (97%)	0.93	31 (12%) <b>3</b> <b>2</b>	50, 89, 130, 164	0
1	B	205/247 (82%)	0.96	33 (16%) <b>1</b> <b>1</b>	52, 124, 168, 182	0
2	D	18/52 (34%)	1.22	6 (33%) <b>0</b> <b>0</b>	111, 143, 179, 190	0
2	E	48/52 (92%)	0.95	4 (8%) <b>11</b> <b>8</b>	64, 101, 153, 168	0
All	All	512/598 (85%)	0.95	74 (14%) <b>2</b> <b>1</b>	50, 99, 163, 190	0

All (74) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	26	GLY	5.1
1	B	223	PHE	4.4
1	A	185	TYR	4.2
1	A	183	ASP	4.0
1	A	243	THR	3.8
1	B	14	TRP	3.7
1	A	179	PHE	3.5
1	B	69	PHE	3.5
1	A	186	TYR	3.4
1	A	36	GLU	3.3
1	B	54	GLY	3.3
1	A	152	TRP	3.3
1	B	236	LEU	3.2
1	B	59	ILE	3.2
1	A	93	SER	3.2
1	B	133	LEU	3.2
1	A	119	ASN	3.1
1	A	169	LEU	3.0
1	A	48	ILE	3.0
1	B	72	ILE	3.0
1	A	153	LEU	3.0

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>	<b>RSRZ</b>
1	B	56	LYS	3.0
1	B	225	ASN	3.0
2	D	39	LYS	2.9
1	B	117	ASN	2.9
2	D	38	GLN	2.9
1	A	181	GLU	2.9
1	B	123	GLU	2.9
2	E	3	THR	2.9
1	B	136	GLY	2.8
1	B	219	PHE	2.8
1	A	67	SER	2.8
1	B	222	ILE	2.7
1	B	18	VAL	2.7
1	B	108	TYR	2.7
1	B	111	THR	2.7
1	A	176	LYS	2.7
1	B	119	ASN	2.7
1	A	145	LYS	2.6
1	A	174	LEU	2.6
1	B	226	SER	2.6
1	A	147	TYR	2.6
1	A	7	LEU	2.6
1	A	240	LEU	2.6
1	A	97	PRO	2.6
1	B	67	SER	2.5
2	E	46	ARG	2.5
2	D	8	LYS	2.5
1	B	131	ILE	2.4
1	A	187	TYR	2.4
1	B	137	THR	2.4
1	B	142	VAL	2.4
1	B	181	GLU	2.4
2	E	25	ALA	2.4
1	B	78	ASP	2.4
1	B	162	PHE	2.3
1	B	107	TYR	2.3
1	A	27	VAL	2.3
2	D	40	LEU	2.3
1	A	24	LEU	2.2
1	A	182	LYS	2.2
1	B	186	TYR	2.2
1	A	172	ASN	2.2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	53	ILE	2.2
1	A	100	LEU	2.2
2	D	7	LEU	2.2
2	E	48	ALA	2.2
1	A	226	SER	2.2
1	A	106	GLU	2.1
2	D	17	TYR	2.1
1	B	52	LEU	2.1
1	B	79	GLU	2.1
1	B	53	ILE	2.0
1	B	15	LYS	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 monosaccharides in this entry.

## 6.4 Ligands [i](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.