



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 16, 2017 – 10:48 AM EDT

PDB ID : 3EE1
Title : Novel fold of VirA, a type III secretion system effector protein from *Shigella flexneri*
Authors : Davis, J.S.
Deposited on : unknown
Resolution : 3.01 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

<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
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030345
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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-20030345

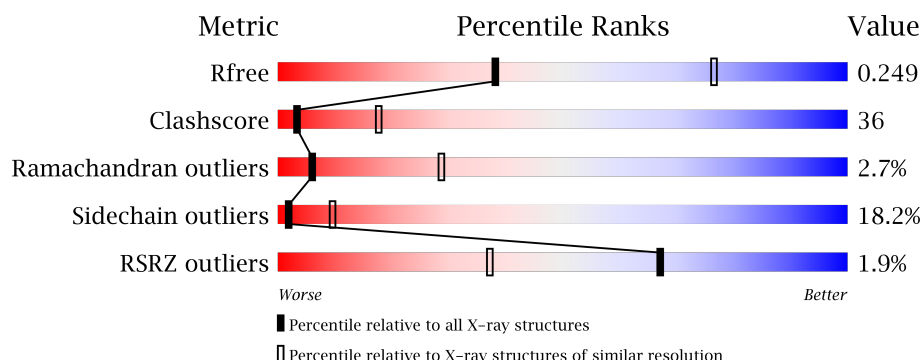
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.01 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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	100719	1924 (3.04-3.00)
Clashscore	112137	2279 (3.04-3.00)
Ramachandran outliers	110173	2207 (3.04-3.00)
Sidechain outliers	110143	2210 (3.04-3.00)
RSRZ outliers	101464	1948 (3.04-3.00)

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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 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	400	<div> <div>2%</div> <div> <div></div> <div>41%</div> <div>36%</div> <div>8%</div> <div>15%</div> </div> </div>
1	B	400	<div> <div>2%</div> <div> <div></div> <div>40%</div> <div>36%</div> <div>9%</div> <div>15%</div> </div> </div>

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 5244 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 Effector protein virA.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	341	Total	C	N	O	S	Se	0	0	0
			2622	1660	439	511	7	5			
1	B	341	Total	C	N	O	S	Se	0	0	0
			2622	1660	439	511	7	5			

There are 34 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLU	-	EXPRESSION TAG	UNP Q7BU69
A	2	VAL	-	EXPRESSION TAG	UNP Q7BU69
A	3	ALA	-	EXPRESSION TAG	UNP Q7BU69
A	4	SER	-	EXPRESSION TAG	UNP Q7BU69
A	5	TYR	-	EXPRESSION TAG	UNP Q7BU69
A	6	CYS	-	EXPRESSION TAG	UNP Q7BU69
A	7	ASP	-	EXPRESSION TAG	UNP Q7BU69
A	8	ARG	-	EXPRESSION TAG	UNP Q7BU69
A	9	VAL	-	EXPRESSION TAG	UNP Q7BU69
A	10	VAL	-	EXPRESSION TAG	UNP Q7BU69
A	11	ALA	-	EXPRESSION TAG	UNP Q7BU69
A	12	ALA	-	EXPRESSION TAG	UNP Q7BU69
A	13	VAL	-	EXPRESSION TAG	UNP Q7BU69
A	36	VAL	ILE	VARIANT	UNP Q7BU69
A	56	TYR	HIS	VARIANT	UNP Q7BU69
A	173	PRO	HIS	VARIANT	UNP Q7BU69
A	239	LEU	VAL	VARIANT	UNP Q7BU69
B	1	GLU	-	EXPRESSION TAG	UNP Q7BU69
B	2	VAL	-	EXPRESSION TAG	UNP Q7BU69
B	3	ALA	-	EXPRESSION TAG	UNP Q7BU69
B	4	SER	-	EXPRESSION TAG	UNP Q7BU69
B	5	TYR	-	EXPRESSION TAG	UNP Q7BU69
B	6	CYS	-	EXPRESSION TAG	UNP Q7BU69
B	7	ASP	-	EXPRESSION TAG	UNP Q7BU69
B	8	ARG	-	EXPRESSION TAG	UNP Q7BU69

Continued on next page...

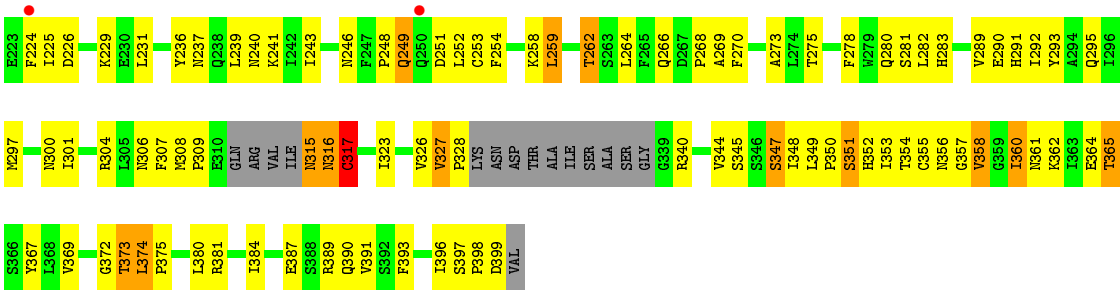
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	9	VAL	-	EXPRESSION TAG	UNP Q7BU69
B	10	VAL	-	EXPRESSION TAG	UNP Q7BU69
B	11	ALA	-	EXPRESSION TAG	UNP Q7BU69
B	12	ALA	-	EXPRESSION TAG	UNP Q7BU69
B	13	VAL	-	EXPRESSION TAG	UNP Q7BU69
B	36	VAL	ILE	VARIANT	UNP Q7BU69
B	56	TYR	HIS	VARIANT	UNP Q7BU69
B	173	PRO	HIS	VARIANT	UNP Q7BU69
B	239	LEU	VAL	VARIANT	UNP Q7BU69

These plots are drawn for all protein, RNA and DNA 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 ($\text{RSRZ} > 2$). 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.

- Chain A:
-
- 2% 41% 36% 8% 15%
- E1 S4 Y5 H8 Y13 F78 A81 N82 I83 D84 G85 L86 S87 T88 G89 C91 L92 I96 C97 S98 V99 F100 Y101 D102 G104 L105 I106 I107 T108 PHE GLY ILE TVR SER PRO HIS GLU THR LEU ALA GLU LYS TYR SER E58 L61 M62 G63 D64 F65 E223 F224 I225 D226 K229 Y236 Q237 Q238 L239 N240 K241 I242 I243 N246 F247 P248 Q249 Q250 D251 L252 C253 F254 K258 L259 T262 S263 L264 T265 Q266 F270 A273 L274 T275 F278 W279 Q280 S281 L282 H283 T284 V289 E290 H291 I292 Y293 A294 K295 L296 M297 N298 R301 R304 L305 N306 F307 K308 P309 E310 G311 ARG VAL ILE N315 K316 C317 L323 V326 V327 P328 LYS ASN ASP THR ALA ILE SER ALA SER GLY R340 V344 S345 S346 S347 I348 L349 P350 S351 K352 L353 T354 C355 K356 G357 V358 G359 I360 N361 K362 L363 E364 T365 S366 V367 L368

- Chain B:
-
- 2% 40% 36% 9% 15%
- S68 L69 S70 S74 Y5 K73 W74 D75 F78 A81 N82 R83 D84 G85 R86 L90 G91 R92 N93 R94 G95 T96 G97 S98 V99 F100 Y101 L102 D103 G104 D105 K106 I107 Q108 S109 T110 Q111 L112 S113 S114 R115 Y116 Y117 M118 N119 L120 L121 S122 S123 L124 P125 M130 L131 G132 K133 Y134 H135 T138 A139 P140 Y141 S142 G143 T144 F146 A151 P152 E153 V154 T157 A158 C161 L166 P167 N168 D169 D170 Y171 D176 T177 D178 F179 V182 Y183 I186 Y187 R188 D189 L190 R191 D194 T198 F202 H203 P208 L209 L212 A213 D214 L215 L216 L221 H222



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	150.25Å 170.91Å 46.16Å 90.00° 104.89° 90.00°	Depositor
Resolution (Å)	28.63 – 3.01 28.63 – 3.01	Depositor EDS
% Data completeness (in resolution range)	96.2 (28.63-3.01) 96.3 (28.63-3.01)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.77 (at 3.00Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.214 , 0.256 0.204 , 0.249	Depositor DCC
R_{free} test set	1096 reflections (5.11%)	DCC
Wilson B-factor (Å ²)	98.9	Xtriage
Anisotropy	0.038	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 85.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.032 for -h-2*k,l	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	5244	wwPDB-VP
Average B, all atoms (Å ²)	120.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.70% of the height of the origin peak. No significant pseudotranslation is detected.*

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

5.1 Standard geometry

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	0.56	0/2670	0.74	2/3618 (0.1%)
1	B	0.53	0/2670	0.72	2/3618 (0.1%)
All	All	0.54	0/5340	0.73	4/7236 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1
1	B	0	1
All	All	0	2

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	104	GLY	N-CA-C	-6.59	96.63	113.10
1	B	104	GLY	N-CA-C	-6.23	97.52	113.10
1	A	102	LEU	CA-CB-CG	5.45	127.83	115.30
1	B	102	LEU	CA-CB-CG	5.19	127.23	115.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	103	ASP	Peptide
1	B	103	ASP	Peptide

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	2622	0	2537	190	0
1	B	2622	0	2537	187	0
All	All	5244	0	5074	369	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 36.

The worst 5 of 369 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:102:LEU:HA	1:A:108:GLN:HA	1.24	1.13
1:B:102:LEU:HA	1:B:108:GLN:HA	1.28	1.12
1:A:102:LEU:HD13	1:A:103:ASP:OD1	1.59	1.00
1:B:102:LEU:HD13	1:B:103:ASP:OD1	1.61	0.98
1:A:102:LEU:HB2	1:A:108:GLN:HG3	1.54	0.90

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	333/400 (83%)	283 (85%)	41 (12%)	9 (3%)	6	29
1	B	333/400 (83%)	285 (86%)	39 (12%)	9 (3%)	6	29
All	All	666/800 (83%)	568 (85%)	80 (12%)	18 (3%)	6	29

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	102	LEU
1	A	316	ASN
1	A	317	CYS
1	A	373	THR
1	B	4	SER

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	294/357 (82%)	242 (82%)	52 (18%)	2	10
1	B	294/357 (82%)	239 (81%)	55 (19%)	2	9
All	All	588/714 (82%)	481 (82%)	107 (18%)	2	9

5 of 107 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	365	THR
1	B	102	LEU
1	B	345	SER
1	A	374	LEU
1	B	75	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	306	ASN
1	B	119	ASN
1	B	306	ASN
1	A	352	HIS
1	B	111	GLN

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](#)

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 [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	336/400 (84%)	-0.31	6 (1%) 69 39	68, 103, 171, 489	0
1	B	336/400 (84%)	-0.22	7 (2%) 64 34	71, 115, 212, 465	0
All	All	672/800 (84%)	-0.26	13 (1%) 67 37	68, 110, 192, 489	0

The worst 5 of 13 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	339	GLY	2.8
1	A	1	GLU	2.7
1	A	171	TYR	2.5
1	A	71	GLN	2.4
1	B	140	PRO	2.4

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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.