wwPDB X-ray Structure Validation Summary Report (i)

Oct 29, 2024 – 08:22 PM EDT

PDB ID	:	4E2H
Title	:	Crystal structure of the periplasmic domain of Shigella flexneri WzzB
Authors	:	Kalynych, S.; Yao, D.; Magee, J.D.; Cygler, M.
Deposited on	:	2012-03-08
Resolution	:	2.36 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 (1)) were used in the production of this report:

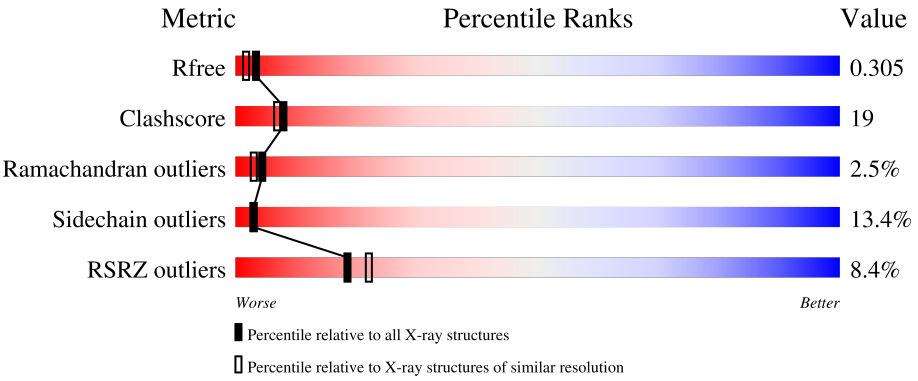
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.36 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	1460 (2.36-2.36)
Clashscore	180529	1571 (2.36-2.36)
Ramachandran outliers	177936	1559 (2.36-2.36)
Sidechain outliers	177891	1559 (2.36-2.36)
RSRZ outliers	164620	1460 (2.36-2.36)

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	240	7% 55% 30% 5% 10%
1	B	240	7% 63% 25% • • 8%
1	C	240	9% 53% 26% 5% • 13%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 5007 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 Chain length determinant protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	216	Total	C	N	O	Se	0	0	0
			1647	1035	285	324	3			
1	C	208	Total	C	N	O	Se	0	0	0
			1578	993	265	317	3			
1	B	222	Total	C	N	O	Se	0	0	0
			1704	1063	296	342	3			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	46	HIS	-	expression tag	UNP P37792
A	47	HIS	-	expression tag	UNP P37792
A	48	HIS	-	expression tag	UNP P37792
A	49	HIS	-	expression tag	UNP P37792
A	50	HIS	-	expression tag	UNP P37792
A	51	HIS	-	expression tag	UNP P37792
A	52	GLY	-	expression tag	UNP P37792
A	53	SER	-	expression tag	UNP P37792
C	46	HIS	-	expression tag	UNP P37792
C	47	HIS	-	expression tag	UNP P37792
C	48	HIS	-	expression tag	UNP P37792
C	49	HIS	-	expression tag	UNP P37792
C	50	HIS	-	expression tag	UNP P37792
C	51	HIS	-	expression tag	UNP P37792
C	52	GLY	-	expression tag	UNP P37792
C	53	SER	-	expression tag	UNP P37792
B	46	HIS	-	expression tag	UNP P37792
B	47	HIS	-	expression tag	UNP P37792
B	48	HIS	-	expression tag	UNP P37792
B	49	HIS	-	expression tag	UNP P37792
B	50	HIS	-	expression tag	UNP P37792
B	51	HIS	-	expression tag	UNP P37792
B	52	GLY	-	expression tag	UNP P37792

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	53	SER	-	expression tag	UNP P37792

- Molecule 2 is water.

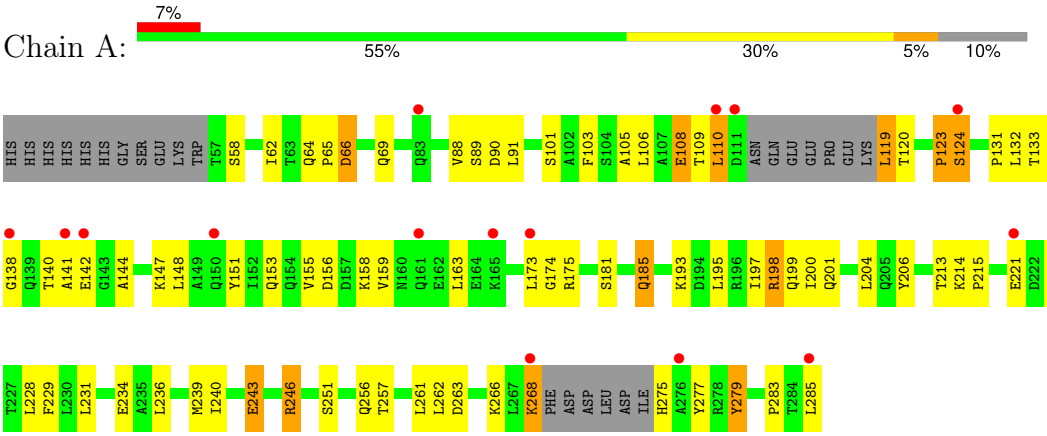
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	26	Total O 26 26	0	0
2	C	31	Total O 31 31	0	0
2	B	21	Total O 21 21	0	0

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.

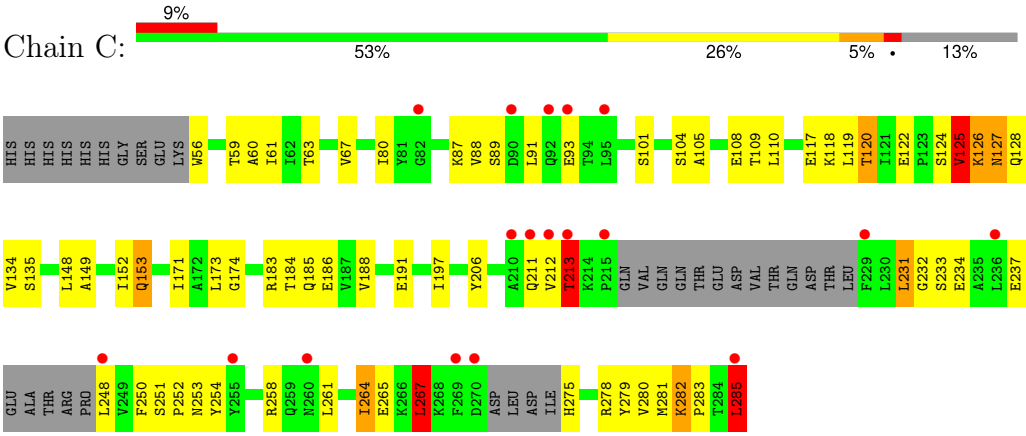
• Molecule 1: Chain length determinant protein

Chain A:



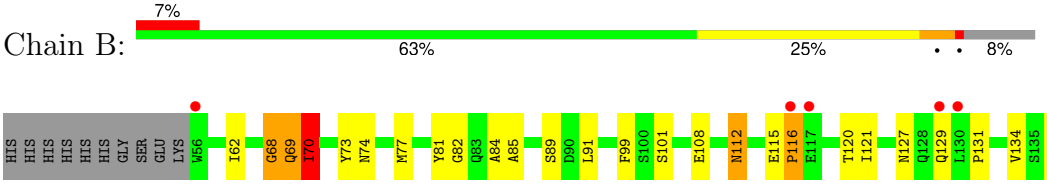
• Molecule 1: Chain length determinant protein

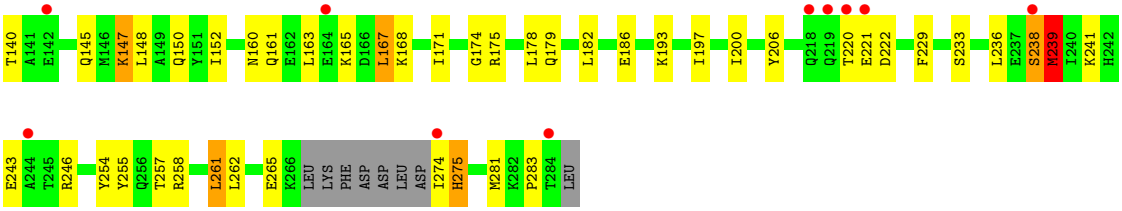
Chain C:



• Molecule 1: Chain length determinant protein

Chain B:




T140 A141 E142 Q145 M146 K147 L148 A149 Q150 Y151 I152 N160 Q161 E162 L163 E164 K165 D166 L167 K168 I171 G174 R175 L178 Q179 L182 E186 K193 I197 I200 Y206 Q218 Q219 T220 E221 D222 F229 S233 L236 E237 S238 M239 I240 K241 H242
E243 A244 T245 R246 Y254 Y255 Q256 T257 R258 L261 L262 E265 K266 LEU LYS PHE ASP ASP LEU ASP I274 H275 M281 K282 P283 T284 LEU

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	81.51Å 62.62Å 90.74Å 90.00° 94.67° 90.00°	Depositor
Resolution (Å)	49.60 – 2.36 49.60 – 2.36	Depositor EDS
% Data completeness (in resolution range)	78.0 (49.60-2.36) 78.0 (49.60-2.36)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.63 (at 2.37Å)	Xtriage
Refinement program	PHENIX 1.7_650	Depositor
R, R_{free}	0.245 , 0.302 0.246 , 0.305	Depositor DCC
R_{free} test set	1508 reflections (5.11%)	wwPDB-VP
Wilson B-factor (Å^2)	33.2	Xtriage
Anisotropy	0.710	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 49.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	5007	wwPDB-VP
Average B, all atoms (Å^2)	52.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.70% 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	0.49	0/1663	0.66	0/2254
1	B	0.48	0/1722	0.66	0/2336
1	C	0.51	0/1595	0.70	4/2161 (0.2%)
All	All	0.49	0/4980	0.67	4/6751 (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	B	0	2
1	C	0	4
All	All	0	6

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	130	LEU	N-CA-C	-6.33	93.92	111.00
1	C	285	LEU	CA-CB-CG	5.47	127.88	115.30
1	C	129	GLN	N-CA-C	-5.41	96.38	111.00
1	C	285	LEU	CB-CG-CD1	5.20	119.85	111.00

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	68	GLY	Peptide
1	C	129	GLN	Peptide
1	C	130	LEU	Peptide
1	C	213	THR	Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
1	C	239	MSE	Peptide

5.2 Too-close contacts

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	1647	0	1618	65	0
1	B	1704	0	1652	55	0
1	C	1578	0	1498	79	0
2	A	26	0	0	9	0
2	B	21	0	0	1	0
2	C	31	0	0	8	0
All	All	5007	0	4768	188	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 19.

The worst 5 of 188 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:C:285:LEU:HD13	1:C:285:LEU:O	1.58	1.02
1:B:148:LEU:HD23	1:B:283:PRO:HB3	1.43	0.99
1:C:238:SER:O	1:C:239:MSE:HB3	1.77	0.82
1:C:285:LEU:O	1:C:285:LEU:CD1	2.28	0.81
1:C:241:LYS:HA	2:C:313:HOH:O	1.80	0.80

There are no symmetry-related clashes.

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	210/240 (88%)	193 (92%)	14 (7%)	3 (1%)	9 7
1	B	218/240 (91%)	193 (88%)	19 (9%)	6 (3%)	4 2
1	C	200/240 (83%)	171 (86%)	22 (11%)	7 (4%)	3 1
All	All	628/720 (87%)	557 (89%)	55 (9%)	16 (2%)	4 3

5 of 16 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	124	SER
1	C	129	GLN
1	C	130	LEU
1	C	131	PRO
1	C	238	SER

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	173/208 (83%)	148 (86%)	25 (14%)	2 2
1	B	180/208 (86%)	160 (89%)	20 (11%)	5 4
1	C	161/208 (77%)	137 (85%)	24 (15%)	2 2
All	All	514/624 (82%)	445 (87%)	69 (13%)	3 3

5 of 69 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	161	GLN
1	B	167	LEU
1	B	243	GLU
1	A	279	TYR
1	A	268	LYS

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

Mol	Chain	Res	Type
1	C	253	ASN
1	C	275	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 (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^{th} 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	212/240 (88%)	0.49	16 (7%)	22	25	27, 50, 78, 104	0
1	B	218/240 (90%)	0.57	16 (7%)	22	26	27, 50, 76, 113	0
1	C	204/240 (85%)	0.68	21 (10%)	13	16	18, 51, 85, 110	0
All	All	634/720 (88%)	0.58	53 (8%)	18	22	18, 50, 81, 113	0

The worst 5 of 53 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	130	LEU	5.8
1	B	221	GLU	5.7
1	B	56	TRP	5.7
1	A	141	ALA	5.4
1	C	269	PHE	5.3

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.