WORLDWIDE PDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 17, 2024 – 02:29 AM EDT

PDB ID	:	3EFB
Title	:	Crystal Structure of Probable sor Operon Regulator from Shigella flexneri
Authors	:	Kim, Y.; Evdokimova, E.; Kudritska, M.; Savchenko, A.; Edwards, A.; Anderson, W.F.; Joachimiak, A.; Center for Structural Genomics of Infectious Diseases (CSGID)
Deposited on	:	2008-09-08
Resolution	:	2.00 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 ⓘ 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:

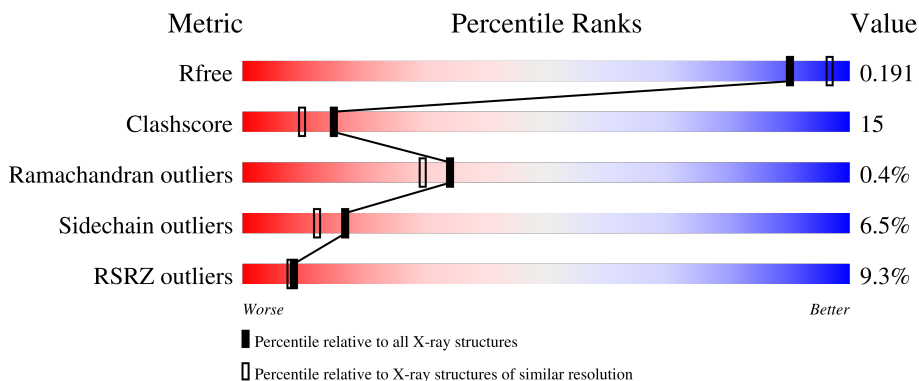|		
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.00 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		0.191
Clashscore		15
Ramachandran outliers		0.4%
Sidechain outliers		6.5%
RSRZ outliers		9.3%

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}	130704	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (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	266	8% / 69% 22% • 8%
1	B	266	6% / 65% 21% • 12%
1	C	266	6% / 68% 17% • 12%
1	D	266	13% / 59% 30% • 7%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard

residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	ACY	B	266	-	-	X	-

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 8130 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 Probable sor-operon regulator.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	246	Total	C	N	O	S	Se	0	14	0
			2005	1256	348	390	3	8			
1	B	235	Total	C	N	O	S	Se	0	3	0
			1845	1170	318	347	3	7			
1	C	233	Total	C	N	O	S	Se	9	6	0
			1841	1162	319	350	3	7			
1	D	248	Total	C	N	O	S	Se	0	3	0
			1936	1221	332	373	3	7			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	GLY	-	EXPRESSION TAG	UNP Q83PA8
B	0	GLY	-	EXPRESSION TAG	UNP Q83PA8
C	0	GLY	-	EXPRESSION TAG	UNP Q83PA8
D	0	GLY	-	EXPRESSION TAG	UNP Q83PA8

- Molecule 2 is ACETIC ACID (three-letter code: ACY) (formula: $C_2H_4O_2$).

ACY

OXT

OH

C

CH3

O

O

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	B	1	Total 4	C 2	O 2	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	154	Total 154	O 154	0	0
3	B	118	Total 118	O 118	0	0
3	C	129	Total 129	O 129	0	0
3	D	98	Total 98	O 98	0	0

W O R L D W I D E

PDB

PROTEIN DATA BANK

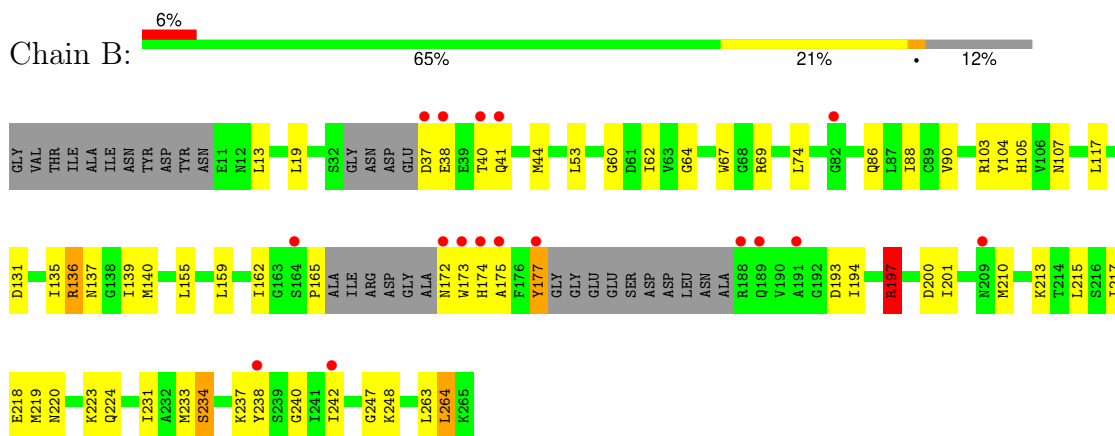# 3 Residue-property plots ⓘ

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.
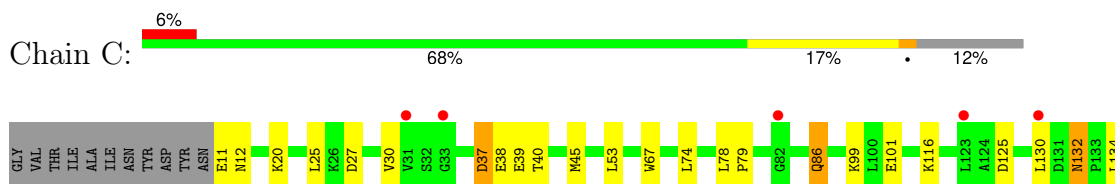
- Molecule 1: Probable sor-operon regulator

- Molecule 1: Probable sor-operon regulator

- Molecule 1: Probable sor-operon regulator

I135 R136 N137 G138 I139 M140 L159 P165 ALA ILE ARG ASP GLY ALA ASN TRP HIS ALA PHE TYR GLY GLU GLU GLU SER ASP ASP LEU ASN A187 R188 Q189 V190 D193 D200 I201 A204 M205 V206 E207 T208 N209 M210 S211 L215 M219 A232 M233 S234 E235 E236 K237

Y238 I241 C252 N256 T259 L264 LYS

• Molecule 1: Probable sor-operon regulator

Chain D:

13%

59%   30%   •   7%

GLY VAL THR ILE ALA ILE ASN TYR ASP TYR N10 E11 L15 K26 G33 ASN ASP E36 D37 E38 E39 T40 Q41 M44 L47 Q51 L52 L53 D54 R55 L56 W67 G68 R69 L74 L78 G82 Q83 S84 R85 Q86 C89 V90 P91 K99 L100 R103

Y104 H105 V106 N107 K116 S121 H122 L123 A124 L129 L130 D131 N132 P133 L134 I135 R136 N137 M140 Q141 S142 Q143 H144 F145 K146 L159 V160 G161 I162 G163 S164 P165 ALA ILE ARG ASP GLY ALA N172 W173 H174 A175 F176 Y177 G178 G179 E180 E181 S182 D183 D184 L185 N186 A187 R188

Q189 V190 A191 G192 D193 I194 R197 F198 F199 D200 G203 V206 E207 T208 N209 M210 S211 E212 K213 S216 I217 E218 M219 N220 K221 L222 K223 Y227 A232 M233 S234 E235 E236 K237 Y238 Y249 S257 S258 T259 A260 E261 L262 L263 L264 K265

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	69.54Å 238.46Å 53.76Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	33.38 – 2.00 33.38 – 2.01	Depositor EDS
% Data completeness (in resolution range)	99.4 (33.38-2.00) 99.0 (33.38-2.01)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.55 (at 2.01Å)	Xtriage
Refinement program	PHENIX (phenix.refine)/refmac5.5	Depositor
R, R_{free}	0.200 , 0.249 0.194 , 0.191	Depositor DCC
R_{free} test set	3049 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	36.7	Xtriage
Anisotropy	0.515	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 52.7	EDS
L-test for twinning[2]	$< |L| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	8130	wwPDB-VP
Average B, all atoms (Å^2)	47.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 64.03 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 8.8222e-06. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[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 ⓘ

5.1 Standard geometry ⓘ

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

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	0.48	0/2027	0.61	0/2719
1	B	0.44	0/1869	0.63	2/2509 (0.1%)
1	C	0.46	0/1862	0.60	0/2500
1	D	0.41	0/1961	0.55	0/2635
All	All	0.45	0/7719	0.60	2/10363 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	197	ARG	NE-CZ-NH1	6.23	123.42	120.30
1	B	197	ARG	NE-CZ-NH2	-5.29	117.65	120.30

There are no chirality outliers.

There are no planarity outliers.

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	2005	0	1982	52	0
1	B	1845	0	1846	56	0
1	C	1841	0	1855	48	0
1	D	1936	0	1917	80	0
2	B	4	0	3	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	154	0	0	7	0
3	B	118	0	0	4	0
3	C	129	0	0	8	0
3	D	98	0	0	7	0
All	All	8130	0	7603	231	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 15.

The worst 5 of 231 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:45:MSE:SE	1:C:233:MSE:HG2	1.94	1.17
1:B:37:ASP:HB3	1:B:40:THR:OG1	1.44	1.16
1:B:194:ILE:HB	1:B:219:MSE:HE1	1.25	1.13
1:A:194:ILE:HB	1:A:219:MSE:HE1	1.34	1.03
1:D:194:ILE:HB	1:D:219:MSE:HE1	1.36	1.02

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	256/266 (96%)	251 (98%)	5 (2%)	0	100	100
1	B	230/266 (86%)	227 (99%)	2 (1%)	1 (0%)	34	30
1	C	235/266 (88%)	230 (98%)	4 (2%)	1 (0%)	34	30
1	D	245/266 (92%)	239 (98%)	4 (2%)	2 (1%)	19	13
All	All	966/1064 (91%)	947 (98%)	15 (2%)	4 (0%)	34	30

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	234	SER
1	C	234	SER
1	D	234	SER
1	D	209	ASN

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	217/211 (103%)	204 (94%)	13 (6%)	19	14
1	B	199/211 (94%)	190 (96%)	9 (4%)	27	24
1	C	199/211 (94%)	187 (94%)	12 (6%)	19	14
1	D	208/211 (99%)	188 (90%)	20 (10%)	8	5
All	All	823/844 (98%)	769 (93%)	54 (7%)	17	12

5 of 54 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	132	ASN
1	D	15	LEU
1	D	206	VAL
1	C	193	ASP
1	C	236	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 21 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	21	GLN
1	D	86	GLN
1	D	224	GLN
1	D	105	HIS
1	D	80	GLN

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.

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

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	ACY	B	266	-	3,3,3	1.09	0	3,3,3	0.61	0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	266	ACY	2	0

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(Å^2)	Q<0.9
1	A	239/266 (89%)	0.32	20 (8%)	11	10	25, 39, 79, 108	3 (1%)
1	B	228/266 (85%)	0.31	17 (7%)	14	13	25, 38, 71, 111	4 (1%)
1	C	226/266 (84%)	0.34	15 (6%)	18	17	24, 39, 77, 100	11 (4%)
1	D	241/266 (90%)	0.72	35 (14%)	2	2	28, 47, 91, 153	24 (9%)
All	All	934/1064 (87%)	0.43	87 (9%)	8	8	24, 40, 82, 153	42 (4%)

The worst 5 of 87 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	10	ASN	11.9
1	D	165	PRO	10.9
1	D	172	ASN	9.5
1	D	177	TYR	8.8
1	B	172	ASN	8.4

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 ⓘ

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(Å²)	Q<0.9
2	ACY	B	266	4/4	0.89	0.25	45,56,68,69	0

6.5 Other polymers ⓘ

There are no such residues in this entry.