WORLDWIDE PDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Nov 12, 2024 – 06:38 AM JST

PDB ID	:	8XTA
Title	:	Clostridioides difficile MarR (WP_003434724)
Authors	:	Park, S.; Kwon, N.
Deposited on	:	2024-01-10
Resolution	:	2.50 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

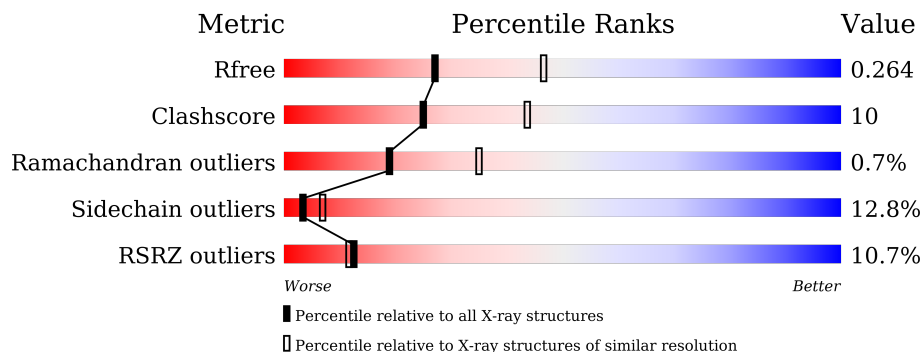|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

1　Overall quality at a glance　ⓘ

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

The reported resolution of this entry is 2.50 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		0.264
Clashscore		10
Ramachandran outliers		0.7%
Sidechain outliers		12.8%
RSRZ outliers		10.7%

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$_{free}$	164625	5504 (2.50-2.50)
Clashscore	180529	6282 (2.50-2.50)
Ramachandran outliers	177936	6191 (2.50-2.50)
Sidechain outliers	177891	6193 (2.50-2.50)
RSRZ outliers	164620	5504 (2.50-2.50)

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	161	24% / 52% 32% 7% 10%
2	B	144	6% / 74% 19% · ·
2	C	144	6% / 74% 18% · ·
2	D	144	3% / 85% 9% · ·

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 4726 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 MarR family transcriptional regulator.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	145	Total	C	N	O	S	0	0	0
			1185	752	199	229	5			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-19	MET	-	initiating methionine	UNP A0A9X9QLH1
A	-18	GLY	-	expression tag	UNP A0A9X9QLH1
A	-17	SER	-	expression tag	UNP A0A9X9QLH1
A	-16	SER	-	expression tag	UNP A0A9X9QLH1
A	-15	HIS	-	expression tag	UNP A0A9X9QLH1
A	-14	HIS	-	expression tag	UNP A0A9X9QLH1
A	-13	HIS	-	expression tag	UNP A0A9X9QLH1
A	-12	HIS	-	expression tag	UNP A0A9X9QLH1
A	-11	HIS	-	expression tag	UNP A0A9X9QLH1
A	-10	HIS	-	expression tag	UNP A0A9X9QLH1
A	-9	SER	-	expression tag	UNP A0A9X9QLH1
A	-8	SER	-	expression tag	UNP A0A9X9QLH1
A	-7	GLY	-	expression tag	UNP A0A9X9QLH1
A	-6	LEU	-	expression tag	UNP A0A9X9QLH1
A	-5	VAL	-	expression tag	UNP A0A9X9QLH1
A	-4	PRO	-	expression tag	UNP A0A9X9QLH1
A	-3	ARG	-	expression tag	UNP A0A9X9QLH1
A	-2	GLY	-	expression tag	UNP A0A9X9QLH1
A	-1	SER	-	expression tag	UNP A0A9X9QLH1
A	0	HIS	-	expression tag	UNP A0A9X9QLH1

- Molecule 2 is a protein called MarR family transcriptional regulator.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	139	Total	C	N	O	S	0	0	0
			1139	725	188	222	4			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	139	Total	C	N	O	S	0	0	0
			1139	725	188	222	4			
2	D	139	Total	C	N	O	S	0	0	0
			1139	725	188	222	4			

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-2	GLY	-	expression tag	UNP A0A9X9QLH1
B	-1	SER	-	expression tag	UNP A0A9X9QLH1
B	0	HIS	-	expression tag	UNP A0A9X9QLH1
C	-2	GLY	-	expression tag	UNP A0A9X9QLH1
C	-1	SER	-	expression tag	UNP A0A9X9QLH1
C	0	HIS	-	expression tag	UNP A0A9X9QLH1
D	-2	GLY	-	expression tag	UNP A0A9X9QLH1
D	-1	SER	-	expression tag	UNP A0A9X9QLH1
D	0	HIS	-	expression tag	UNP A0A9X9QLH1

- Molecule 3 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	35	Total	O	0	0
			35	35		
3	B	53	Total	O	0	0
			53	53		
3	C	23	Total	O	0	0
			23	23		
3	D	13	Total	O	0	0
			13	13		

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.
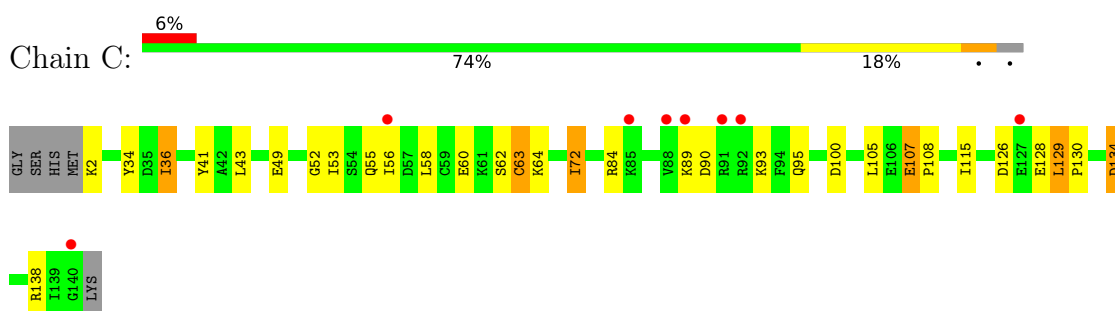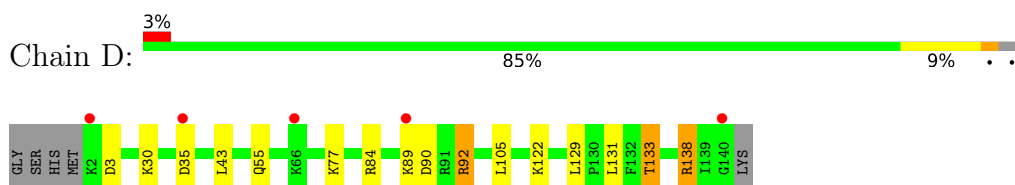
• Molecule 1: MarR family transcriptional regulator
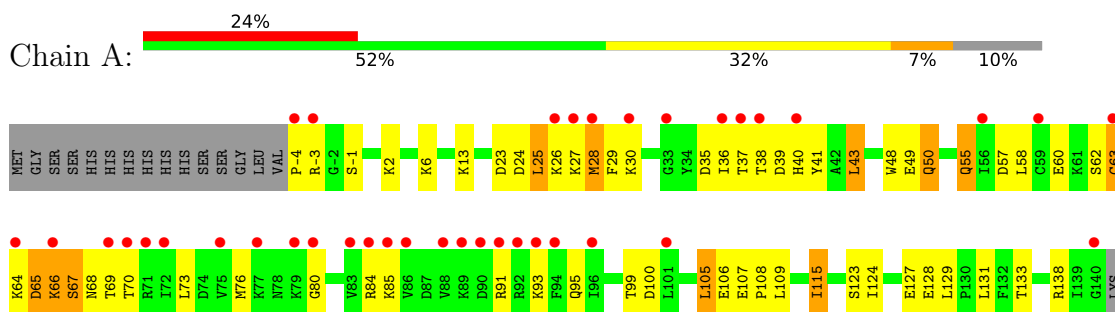
• Molecule 2: MarR family transcriptional regulator
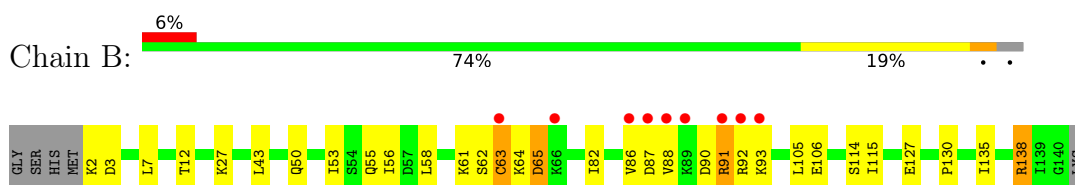
• Molecule 2: MarR family transcriptional regulator

• Molecule 2: MarR family transcriptional regulator

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	176.48Å 69.40Å 95.01Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	95.01 – 2.50 95.01 – 2.50	Depositor EDS
% Data completeness (in resolution range)	99.8 (95.01-2.50) 99.8 (95.01-2.50)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.91 (at 2.51Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, R_{free}	0.210 , 0.260 0.216 , 0.264	Depositor DCC
R_{free} test set	2065 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	51.0	Xtriage
Anisotropy	0.342	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 47.8	EDS
L-test for twinning[2]	$< |L| > = 0.48, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	4726	wwPDB-VP
Average B, all atoms (Å^2)	60.0	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 19.87 % 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 9.8432e-03. 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 ⓘ

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	0.95	0/1202	1.09	4/1611 (0.2%)
2	B	0.96	1/1154 (0.1%)	0.98	1/1548 (0.1%)
2	C	0.91	1/1154 (0.1%)	0.99	2/1548 (0.1%)
2	D	0.90	1/1154 (0.1%)	1.05	4/1548 (0.3%)
All	All	0.93	3/4664 (0.1%)	1.03	11/6255 (0.2%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	35	ASP	CB-CG	5.37	1.63	1.51
2	B	106	GLU	CD-OE1	-5.36	1.19	1.25
2	C	34	TYR	CZ-OH	-5.33	1.28	1.37

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	138	ARG	NE-CZ-NH1	7.85	124.22	120.30
1	A	25	LEU	CB-CG-CD2	7.51	123.76	111.00
1	A	138	ARG	NE-CZ-NH2	-6.86	116.87	120.30
1	A	24	ASP	CB-CG-OD1	-6.59	112.37	118.30
2	B	138	ARG	NE-CZ-NH1	6.49	123.55	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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1185	0	1212	57	0
2	B	1139	0	1165	16	0
2	C	1139	0	1165	20	0
2	D	1139	0	1165	6	0
3	A	35	0	0	4	0
3	B	53	0	0	2	0
3	C	23	0	0	2	0
3	D	13	0	0	0	0
All	All	4726	0	4707	89	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 10.

The worst 5 of 89 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:65:ASP:HA	1:A:66:LYS:HE3	1.46	0.97
1:A:115:ILE:HD11	2:C:115:ILE:HG21	1.50	0.93
2:D:90:ASP:OD1	2:D:92:ARG:HD3	1.73	0.88
1:A:39:ASP:HA	3:A:205:HOH:O	1.76	0.84
1:A:65:ASP:OD2	1:A:68:ASN:ND2	2.15	0.79

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	143/161 (89%)	125 (87%)	16 (11%)	2 (1%)	9	17
2	B	137/144 (95%)	131 (96%)	5 (4%)	1 (1%)	19	35
2	C	137/144 (95%)	129 (94%)	7 (5%)	1 (1%)	19	35
2	D	137/144 (95%)	135 (98%)	2 (2%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
All	All	554/593 (93%)	520 (94%)	30 (5%)	4 (1%)	19	35

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	28	MET
2	B	63	CYS
2	C	53	ILE
1	A	55	GLN

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	135/149 (91%)	105 (78%)	30 (22%)	1	1
2	B	130/134 (97%)	114 (88%)	16 (12%)	4	8
2	C	130/134 (97%)	117 (90%)	13 (10%)	6	13
2	D	130/134 (97%)	122 (94%)	8 (6%)	15	31
All	All	525/551 (95%)	458 (87%)	67 (13%)	3	7

5 of 67 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	126	ASP
2	D	3	ASP
2	D	131	LEU
1	A	115	ILE
1	A	105	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	40	HIS

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Res	Type
2	B	55	GLN
2	C	55	GLN
2	D	55	GLN
2	D	95	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 oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

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	145/161 (90%)	1.11	38 (26%)	2	2	34, 65, 122, 151	0
2	B	139/144 (96%)	0.23	9 (6%)	26	24	33, 51, 108, 150	0
2	C	139/144 (96%)	0.10	8 (5%)	30	28	30, 50, 91, 138	0
2	D	139/144 (96%)	-0.06	5 (3%)	46	43	34, 50, 76, 92	0
All	All	562/593 (94%)	0.35	60 (10%)	12	12	30, 53, 115, 151	0

The worst 5 of 60 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	86	VAL	6.6
2	B	88	VAL	6.6
1	A	85	LYS	6.1
1	A	88	VAL	5.3
1	A	89	LYS	5.2

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 ⓘ

There are no such residues in this entry.