WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 17, 2023 – 09:36 AM EST

PDB ID	:	2IT9
Title	:	Crystal structure of a protein with unknown function from DUF155 family (YP_292156.1) from Prochlorococcus sp. NATL2A at 1.80 A resolution
Authors	:	Joint Center for Structural Genomics (JCSG)
Deposited on	:	2006-10-19
Resolution	:	1.80 Å(reported)

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

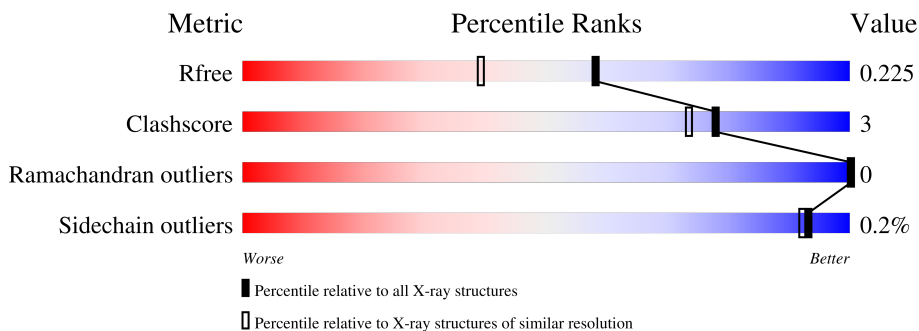|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.31.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.2

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 1.80 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		0.225
Clashscore		3
Ramachandran outliers		0
Sidechain outliers		0.2%

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	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)

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%

Mol	Chain	Length	Quality of chain
1	A	127	87% 7% • 6%
1	B	127	86% 8% • 6%
1	C	127	88% 6% 6%
1	D	127	90% 6% •

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

2   Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 4474 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 Hypothetical protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	120	Total 1014	C 655	N 168	O 183	S 1	Se 7	0	8	0
1	B	120	Total 979	C 629	N 166	O 178	S 1	Se 5	0	3	0
1	C	120	Total 994	C 637	N 171	O 178	S 2	Se 6	0	6	0
1	D	122	Total 1050	C 670	N 180	O 192	S 2	Se 6	1	8	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	GLY	-	expression tag	UNP Q46J75
A	1	MSE	MET	modified residue	UNP Q46J75
A	46	MSE	MET	modified residue	UNP Q46J75
A	60	MSE	MET	modified residue	UNP Q46J75
A	117	MSE	MET	modified residue	UNP Q46J75
A	120	MSE	MET	modified residue	UNP Q46J75
B	0	GLY	-	expression tag	UNP Q46J75
B	1	MSE	MET	modified residue	UNP Q46J75
B	46	MSE	MET	modified residue	UNP Q46J75
B	60	MSE	MET	modified residue	UNP Q46J75
B	117	MSE	MET	modified residue	UNP Q46J75
B	120	MSE	MET	modified residue	UNP Q46J75
C	0	GLY	-	expression tag	UNP Q46J75
C	1	MSE	MET	modified residue	UNP Q46J75
C	46	MSE	MET	modified residue	UNP Q46J75
C	60	MSE	MET	modified residue	UNP Q46J75
C	117	MSE	MET	modified residue	UNP Q46J75
C	120	MSE	MET	modified residue	UNP Q46J75
D	0	GLY	-	expression tag	UNP Q46J75
D	1	MSE	MET	modified residue	UNP Q46J75
D	46	MSE	MET	modified residue	UNP Q46J75

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

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	60	MSE	MET	modified residue	UNP Q46J75
D	117	MSE	MET	modified residue	UNP Q46J75
D	120	MSE	MET	modified residue	UNP Q46J75

- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Cl		
2	A	1	1	1	0	0

- Molecule 3 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula: $C_6H_{14}O_4$).
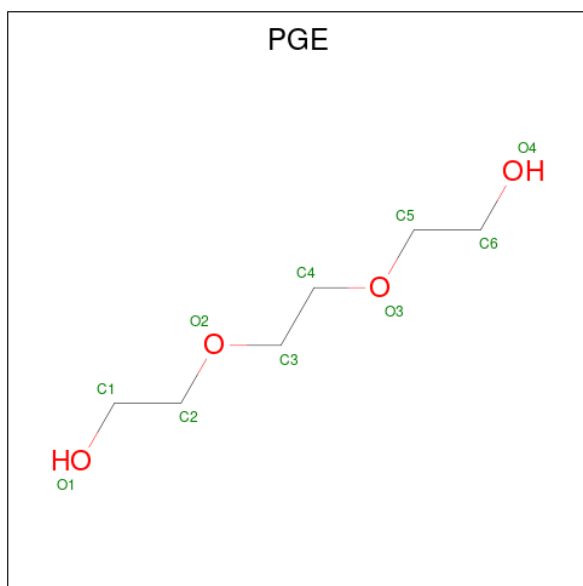
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total 10	C 6	O 4	0	0
3	B	1	Total 10	C 6	O 4	0	0
3	D	1	Total 6	C 4	O 2	0	0

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).

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

- Molecule 5 is 2-AMINO-2-HYDROXYMETHYL-PROPANE-1,3-DIOL (three-letter code: TRS) (formula: $C_4H_{12}NO_3$).
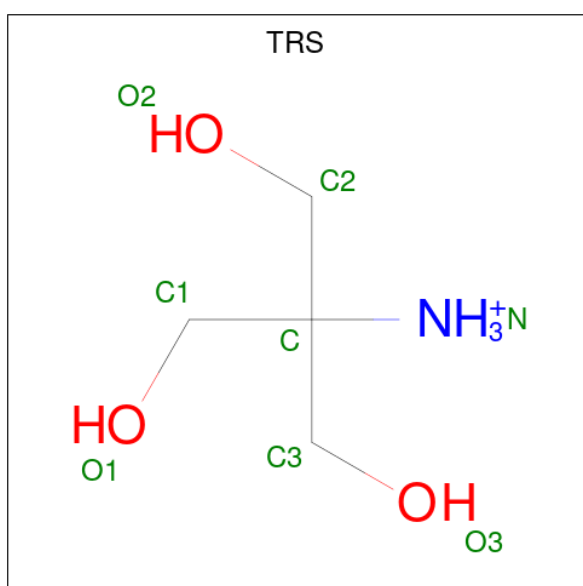
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	D	1	Total	C	N	O	0	0
			8	4	1	3		

• Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	105	Total	O	0	0
			105	105		
6	B	106	Total	O	0	0
			106	106		
6	C	105	Total	O	0	0
			105	105		
6	D	70	Total	O	0	0
			70	70		

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. 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. 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: Hypothetical protein

Chain A:

- Molecule 1: Hypothetical protein

Chain B:

- Molecule 1: Hypothetical protein

Chain C:

- Molecule 1: Hypothetical protein

Chain D:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	59.72Å 81.02Å 124.35Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	67.88 – 1.80 67.88 – 1.80	Depositor EDS
% Data completeness (in resolution range)	99.3 (67.88-1.80) 99.3 (67.88-1.80)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	0.06	Depositor
$< I/\sigma(I) >$ [1]	1.71 (at 1.80Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.179 , 0.213 0.188 , 0.225	Depositor DCC
R_{free} test set	2857 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å^2)	26.6	Xtriage
Anisotropy	0.298	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 44.1	EDS
L-test for twinning[2]	$< |L| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	4474	wwPDB-VP
Average B, all atoms (Å^2)	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.16% 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 ⓘ

5.1   Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: CL, PGE, TRS, EDO

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	0.76	0/1053	0.85	1/1412 (0.1%)
1	B	0.61	0/1005	0.81	1/1357 (0.1%)
1	C	0.70	1/1029 (0.1%)	0.86	2/1385 (0.1%)
1	D	1.07	2/1082 (0.2%)	0.82	0/1458
All	All	0.81	3/4169 (0.1%)	0.84	4/5612 (0.1%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	49[A]	CYS	CB-SG	-19.40	1.49	1.82
1	D	49[B]	CYS	CB-SG	-19.40	1.49	1.82
1	C	107	ARG	CB-CG	-5.94	1.36	1.52

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	107	ARG	NE-CZ-NH2	-8.13	116.24	120.30
1	A	107	ARG	NE-CZ-NH2	-5.46	117.57	120.30
1	B	99	ARG	NE-CZ-NH1	5.23	122.92	120.30
1	C	107	ARG	CG-CD-NE	-5.09	101.11	111.80

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	1014	0	1017	10	0
1	B	979	0	952	6	0
1	C	994	0	990	4	0
1	D	1050	0	1030	5	0
2	A	1	0	0	1	0
3	A	10	0	14	0	0
3	B	10	0	14	0	0
3	D	6	0	7	0	0
4	B	4	0	6	0	0
4	C	8	0	12	0	0
4	D	4	0	6	0	0
5	D	8	0	12	0	0
6	A	105	0	0	3	0
6	B	106	0	0	0	0
6	C	105	0	0	1	0
6	D	70	0	0	1	0
All	All	4474	0	4060	24	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 3.

The worst 5 of 24 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:D:46[A]:MSE:HE1	1:D:114:VAL:HG13	1.74	0.70
1:A:107:ARG:HD2	6:A:215:HOH:O	1.94	0.66
1:D:46[B]:MSE:HE3	1:D:118:ARG:HG2	1.78	0.64
1:A:93:ALA:HA	1:A:98:ASN:O	2.06	0.55
1:A:60[A]:MSE:CE	6:D:144:HOH:O	2.57	0.53

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	125/127 (98%)	121 (97%)	4 (3%)	0	100	100
1	B	121/127 (95%)	118 (98%)	3 (2%)	0	100	100
1	C	124/127 (98%)	120 (97%)	4 (3%)	0	100	100
1	D	128/127 (101%)	124 (97%)	4 (3%)	0	100	100
All	All	498/508 (98%)	483 (97%)	15 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	109/107 (102%)	109 (100%)	0	100	100
1	B	99/107 (92%)	98 (99%)	1 (1%)	76	71
1	C	105/107 (98%)	105 (100%)	0	100	100
1	D	114/107 (106%)	114 (100%)	0	100	100
All	All	427/428 (100%)	426 (100%)	1 (0%)	93	92

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

Mol	Chain	Res	Type
1	B	55	VAL

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

Mol	Chain	Res	Type
1	A	112	ASN
1	C	58	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 ⓘ

Of 9 ligands modelled in this entry, 1 is monoatomic - leaving 8 for Mogul analysis.

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	$\#|Z| > 2$
3	PGE	B	127	-	9,9,9	0.44	0	8,8,8	0.47	0
4	EDO	B	128	-	3,3,3	0.43	0	2,2,2	0.38	0
4	EDO	C	127	-	3,3,3	0.50	0	2,2,2	0.39	0
3	PGE	A	128	-	9,9,9	0.38	0	8,8,8	0.65	0
3	PGE	D	127	-	5,5,9	0.53	0	4,4,8	0.43	0
4	EDO	D	128	-	3,3,3	0.49	0	2,2,2	0.21	0
4	EDO	C	128	-	3,3,3	0.55	0	2,2,2	0.16	0
5	TRS	D	129	-	7,7,7	0.19	0	9,9,9	0.64	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	PGE	B	127	-	-	4/7/7/7	-
4	EDO	B	128	-	-	1/1/1/1	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	EDO	C	127	-	-	1/1/1/1	-
3	PGE	A	128	-	-	1/7/7/7	-
3	PGE	D	127	-	-	<mark>2/3/3/7</mark>	-
4	EDO	D	128	-	-	0/1/1/1	-
4	EDO	C	128	-	-	1/1/1/1	-
5	TRS	D	129	-	-	<mark>5/9/9/9</mark>	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 15 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	D	127	PGE	O1-C1-C2-O2
4	B	128	EDO	O1-C1-C2-O2
4	C	127	EDO	O1-C1-C2-O2
4	C	128	EDO	O1-C1-C2-O2
3	B	127	PGE	C6-C5-O3-C4

There are no ring outliers.

No monomer is involved in short contacts.

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 ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2   Non-standard residues in protein, DNA, RNA chains ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3   Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4   Ligands ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5   Other polymers ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.