Full wwPDB X-ray Structure Validation Report (i)

Jan 22, 2024 – 07:08 pm GMT

PDB ID	:	8C7I
Title	:	Crystal structure of the PS2 assembly factor Psb32 from the cyanobactium Thermosyncechococcus vestitus (formerly elongatus)
Authors	:	Liauw, P.; Gasper, R.; Nowaczyk, M.M.; Hofmann, E.
Deposited on	:	2023-01-16
Resolution	:	2.12 Å(reported)

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

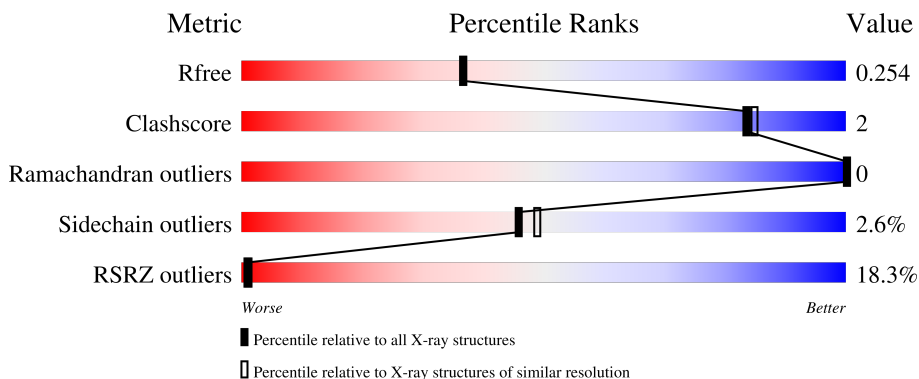|		
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.12 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		0.254
Clashscore		2
Ramachandran outliers		0
Sidechain outliers		2.6%
RSRZ outliers		18.3%

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	6241 (2.14-2.10)
Clashscore	141614	6778 (2.14-2.10)
Ramachandran outliers	138981	6705 (2.14-2.10)
Sidechain outliers	138945	6706 (2.14-2.10)
RSRZ outliers	127900	6112 (2.14-2.10)

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	437	17% — 84% — 9% — 7%

2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 6285 atoms, of which 3088 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 Green fluorescent protein,Tll0404 protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	405	6269	2004	3088	544	625	8	0	0	0

There are 43 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-267	MET	-	initiating methionine	UNP P42212
A	-266	ALA	-	expression tag	UNP P42212
A	-265	SER	-	expression tag	UNP P42212
A	-264	TRP	-	expression tag	UNP P42212
A	-263	SER	-	expression tag	UNP P42212
A	-262	HIS	-	expression tag	UNP P42212
A	-261	PRO	-	expression tag	UNP P42212
A	-260	GLN	-	expression tag	UNP P42212
A	-259	PHE	-	expression tag	UNP P42212
A	-258	GLU	-	expression tag	UNP P42212
A	-257	LYS	-	expression tag	UNP P42212
A	-256	ALA	-	expression tag	UNP P42212
A	-255	SER	-	expression tag	UNP P42212
A	-225	ARG	SER	engineered mutation	UNP P42212
A	-216	ASN	TYR	engineered mutation	UNP P42212
A	-191	LEU	PHE	engineered mutation	UNP P42212
A	-189	CRO	SER	chromophore	UNP P42212
A	-189	CRO	TYR	chromophore	UNP P42212
A	-189	CRO	GLY	chromophore	UNP P42212
A	-175	ARG	GLN	engineered mutation	UNP P42212
A	-156	SER	PHE	engineered mutation	UNP P42212
A	-150	THR	ASN	engineered mutation	UNP P42212
A	-110	PHE	TYR	engineered mutation	UNP P42212
A	-102	THR	MET	engineered mutation	UNP P42212
A	-92	ALA	VAL	engineered mutation	UNP P42212
A	-84	VAL	ILE	engineered mutation	UNP P42212
A	-49	VAL	ALA	engineered mutation	UNP P42212

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	-16	TYR	-	linker	UNP P42212
A	-15	ARG	-	linker	UNP P42212
A	-14	ILE	-	linker	UNP P42212
A	-13	ARG	-	linker	UNP P42212
A	-12	SER	-	linker	UNP P42212
A	-11	GLY	-	linker	UNP P42212
A	-10	GLY	-	linker	UNP P42212
A	-9	GLY	-	linker	UNP P42212
A	-8	GLY	-	linker	UNP P42212
A	-7	GLU	-	linker	UNP P42212
A	-6	ASN	-	linker	UNP P42212
A	-5	LEU	-	linker	UNP P42212
A	-4	TYR	-	linker	UNP P42212
A	-3	PHE	-	linker	UNP P42212
A	-2	GLN	-	linker	UNP P42212
A	-1	GLY	-	linker	UNP P42212

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	16	Total 16	O 16	0	0

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: Green fluorescent protein,Tll0404 protein
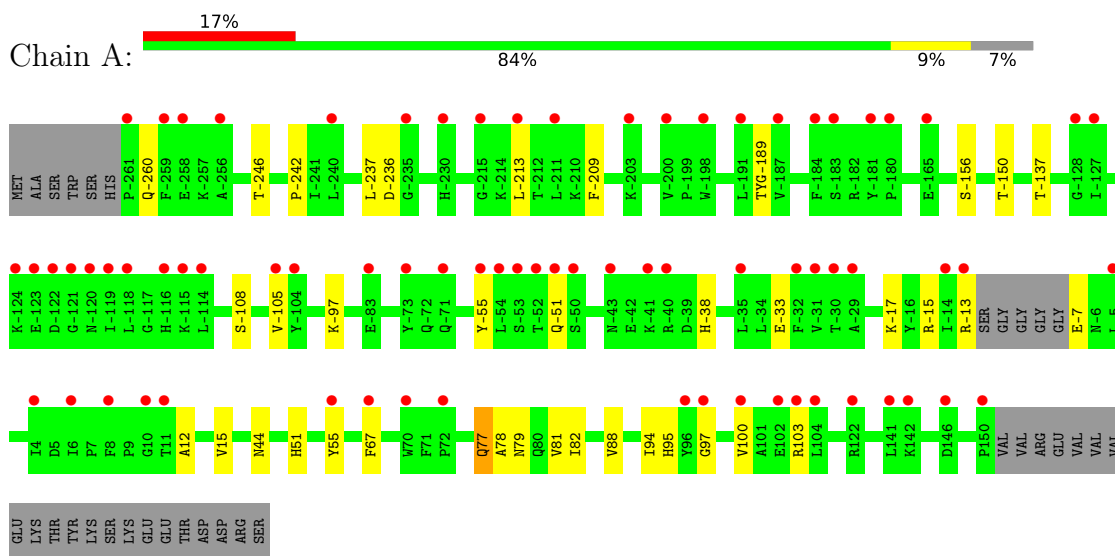
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 61 2 2	Depositor
Cell constants a, b, c, α, β, γ	145.65Å 145.65Å 91.81Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	47.68 – 2.12	Depositor
	47.68 – 2.12	EDS
% Data completeness	99.2 (47.68-2.12)	Depositor
(in resolution range)	88.8 (47.68-2.12)	EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.96 (at 2.12Å)	Xtriage
Refinement program	PHENIX dev_4788	Depositor
R, R_{free}	0.227 , 0.258	Depositor
	0.224 , 0.254	DCC
R_{free} test set	1639 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	49.4	Xtriage
Anisotropy	0.301	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.42 , 74.9	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	6285	wwPDB-VP
Average B, all atoms (Å^2)	94.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.52% 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: CRO

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	0.33	0/3226	0.56	0/4376

There are no bond length outliers.

There are no bond angle outliers.

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	3181	3088	3092	15	1
2	A	16	0	0	0	0
All	All	3197	3088	3092	15	1

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 2.

All (15) 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:78:ALA:HA	1:A:97:GLY:HA3	1.91	0.52
1:A:-108:SER:OG	1:A:-51:GLN:NE2	2.44	0.50

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:77:GLN:O	1:A:79:ASN:N	2.46	0.46
1:A:81:VAL:HG22	1:A:100:VAL:HG21	1.98	0.46
1:A:94:ILE:HG12	1:A:95:HIS:N	2.32	0.45
1:A:51:HIS:O	1:A:88:VAL:HG23	2.16	0.44
1:A:-105:VAL:O	1:A:-55:TYR:HA	2.17	0.43
1:A:67:PHE:CE2	1:A:77:GLN:CG	3.01	0.43
1:A:-213:LEU:HB3	1:A:-33:GLU:HB2	2.01	0.42
1:A:-156:SER:HA	1:A:-150:THR:HG22	2.01	0.42
1:A:-242:PRO:HG2	1:A:-137:THR:HG22	2.03	0.41
1:A:81:VAL:HG12	1:A:82:ILE:N	2.36	0.41
1:A:-209:PHE:O	1:A:-38:HIS:HB2	2.20	0.41
1:A:-237:LEU:HD23	1:A:-236:ASP:N	2.36	0.41
1:A:12:ALA:O	1:A:44:ASN:HB3	2.21	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:-260:GLN:O	1:A:-15:ARG:NH2[10_554]	2.19	0.01

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	398/437 (91%)	382 (96%)	16 (4%)	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	348/375 (93%)	339 (97%)	9 (3%)	46 49

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	-246	THR
1	A	-97	LYS
1	A	-17	LYS
1	A	-13	ARG
1	A	-7	GLU
1	A	15	VAL
1	A	55	TYR
1	A	77	GLN
1	A	103	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

1 non-standard protein/DNA/RNA residue 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	#$|Z| > 2$
1	CRO	A	-189	1	23,23,24	3.09	7 (30%)	30,32,34	3.05	7 (23%)

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
1	CRO	A	-189	1	-	2/12/31/32	0/2/2/2

All (7) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	-189	CRO	C1-N3	6.98	1.49	1.37
1	A	-189	CRO	C1-N2	6.75	1.42	1.32
1	A	-189	CRO	CA2-C2	6.18	1.54	1.48
1	A	-189	CRO	CG2-CB2	5.62	1.57	1.46
1	A	-189	CRO	OH-CZ	4.45	1.47	1.37
1	A	-189	CRO	CB2-CA2	3.26	1.37	1.35
1	A	-189	CRO	C2-N3	2.31	1.45	1.39

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	-189	CRO	CA2-C2-N3	10.87	108.51	103.37
1	A	-189	CRO	C2-N3-C1	-7.86	103.99	107.97
1	A	-189	CRO	O2-C2-CA2	-6.19	127.48	130.96
1	A	-189	CRO	CA3-N3-C1	4.01	131.97	127.16
1	A	-189	CRO	CA2-N2-C1	2.96	107.96	105.77
1	A	-189	CRO	C2-CA2-N2	-2.93	106.88	108.93
1	A	-189	CRO	CG2-CB2-CA2	-2.66	126.69	129.94

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	-189	CRO	C2-CA2-CB2-CG2
1	A	-189	CRO	N2-CA2-CB2-CG2

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates ⓘ

There are no monosaccharides 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	404/437 (92%)	1.11	74 (18%)	1	1	55, 84, 118, 145	0

All (74) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	10	GLY	7.4
1	A	-203	LYS	5.0
1	A	70	TRP	4.5
1	A	67	PHE	4.5
1	A	72	PRO	4.5
1	A	-41	LYS	4.4
1	A	-118	LEU	4.3
1	A	104	LEU	4.2
1	A	-230	HIS	4.1
1	A	-127	ILE	4.1
1	A	97	GLY	4.0
1	A	-191	LEU	4.0
1	A	-124	LYS	3.9
1	A	-256	ALA	3.7
1	A	-115	LYS	3.6
1	A	-235	GLY	3.5
1	A	-128	GLY	3.4
1	A	-51	GLN	3.3
1	A	141	LEU	3.1
1	A	-119	ILE	3.0
1	A	-121	GLY	3.0
1	A	102	GLU	2.9
1	A	-43	ASN	2.8
1	A	-200	VAL	2.8
1	A	-122	ASP	2.8
1	A	-116	HIS	2.7
1	A	-13	ARG	2.7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	-258	GLU	2.7
1	A	-5	LEU	2.7
1	A	-29	ALA	2.7
1	A	-32	PHE	2.7
1	A	-181	TYR	2.6
1	A	8	PHE	2.6
1	A	-187	VAL	2.6
1	A	-55	TYR	2.6
1	A	142	LYS	2.6
1	A	-73	TYR	2.6
1	A	-54	LEU	2.6
1	A	-14	ILE	2.5
1	A	-183	SER	2.5
1	A	-31	VAL	2.5
1	A	103	ARG	2.5
1	A	96	TYR	2.5
1	A	-123	GLU	2.5
1	A	-52	THR	2.5
1	A	11	THR	2.4
1	A	-180	PRO	2.4
1	A	-261	PRO	2.4
1	A	-213	LEU	2.4
1	A	150	PRO	2.4
1	A	122	ARG	2.4
1	A	-215	GLY	2.4
1	A	-83	GLU	2.3
1	A	-240	LEU	2.3
1	A	-211	LEU	2.3
1	A	-198	TRP	2.3
1	A	-30	THR	2.3
1	A	4	ILE	2.3
1	A	-104	TYR	2.3
1	A	-53	SER	2.3
1	A	-120	ASN	2.2
1	A	-165	GLU	2.2
1	A	-40	ARG	2.2
1	A	-105	VAL	2.2
1	A	55	TYR	2.2
1	A	-114	LEU	2.1
1	A	-259	PHE	2.1
1	A	-71	GLN	2.1
1	A	6	ILE	2.1

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	-35	LEU	2.1
1	A	-50	SER	2.1
1	A	-184	PHE	2.1
1	A	100	VAL	2.0
1	A	146	ASP	2.0

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

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($Å^2$)	Q<0.9
1	CRO	A	-189	22/23	0.92	0.24	63,76,90,99	0

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.