Full wwPDB X-ray Structure Validation Report (i)

Jan 24, 2021 – 12:09 PM EST

PDB ID	:	2OZT
Title	:	Crystal structure of O-succinylbenzoate synthase from Thermosynechococcus elongatus BP-1
Authors	:	Malashkevich, V.N.; Bonanno, J.; Toro, R.; Sauder, J.M.; Schwinn, K.D.; Bain, K.T.; Adams, J.M.; Reyes, C.; Rooney, I.; Gheyi, T.; Wasserman, S.R.; Emtage, S.; Burley, S.K.; Almo, S.C.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on	:	2007-02-27
Resolution	:	1.42 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 following versions of software and data (see references (1)) were used in the production of this report:

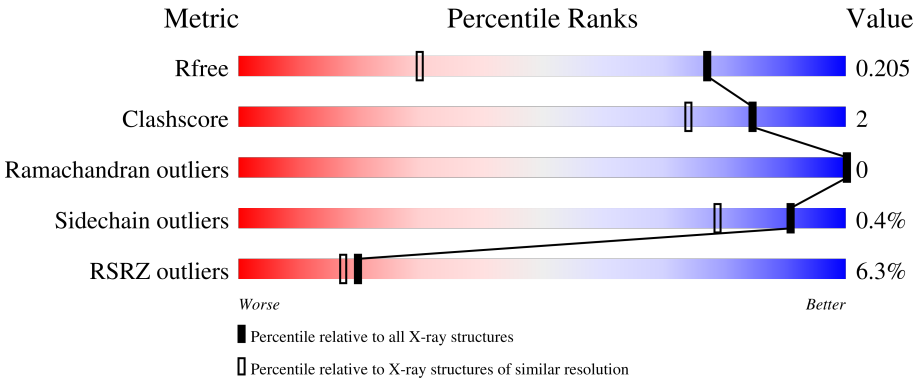
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.16
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.16

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 1.42 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}	130704	2579 (1.44-1.40)
Clashscore	141614	2696 (1.44-1.40)
Ramachandran outliers	138981	2632 (1.44-1.40)
Sidechain outliers	138945	2631 (1.44-1.40)
RSRZ outliers	127900	2528 (1.44-1.40)

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	332	6% 93%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2974 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.

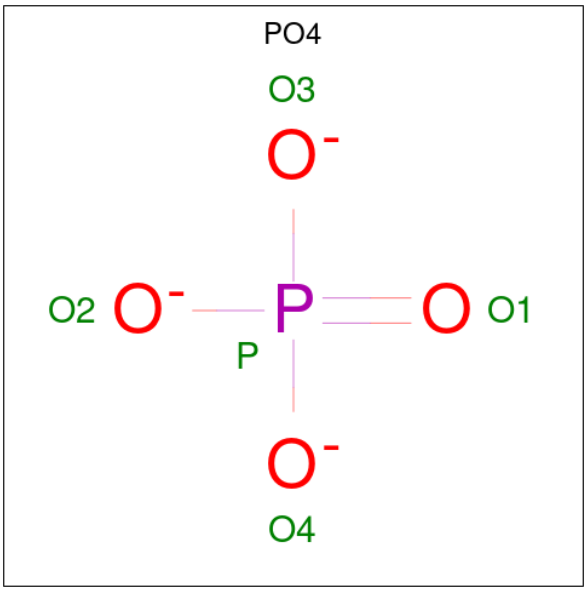
- Molecule 1 is a protein called Tlr1174 protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	321	2581	1665	453	458	3	2	0	3	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MSE	-	cloning artifact	UNP Q8DJP8
A	0	SER	-	cloning artifact	UNP Q8DJP8
A	1	LEU	-	cloning artifact	UNP Q8DJP8
A	75	MSE	MET	modified residue	UNP Q8DJP8
A	140	MSE	MET	modified residue	UNP Q8DJP8
A	323	GLU	-	cloning artifact	UNP Q8DJP8
A	324	GLY	-	cloning artifact	UNP Q8DJP8
A	325	HIS	-	cloning artifact	UNP Q8DJP8
A	326	HIS	-	cloning artifact	UNP Q8DJP8
A	327	HIS	-	cloning artifact	UNP Q8DJP8
A	328	HIS	-	cloning artifact	UNP Q8DJP8
A	329	HIS	-	cloning artifact	UNP Q8DJP8
A	330	HIS	-	cloning artifact	UNP Q8DJP8

- Molecule 2 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total O P 5 4 1	0	0
2	A	1	Total O P 5 4 1	0	0

- Molecule 3 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total Na 1 1	0	0

- Molecule 4 is water.

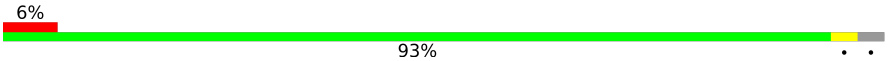
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	382	Total O 382 382	0	0

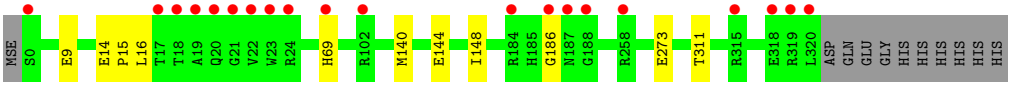
3 Residue-property plots (i)

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: Tlr1174 protein

Chain A:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	39.64Å 113.50Å 148.16Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.70 – 1.42 19.70 – 1.42	Depositor EDS
% Data completeness (in resolution range)	(Not available) (19.70-1.42) 95.6 (19.70-1.42)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.16 (at 1.41Å)	Xtriage
Refinement program	REFMAC 5.3.0028, SHELX	Depositor
R, R_{free}	0.178 , 0.207 0.176 , 0.205	Depositor DCC
R_{free} test set	3115 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å^2)	17.6	Xtriage
Anisotropy	0.209	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.40 , 50.9	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	0.97	EDS
Total number of atoms	2974	wwPDB-VP
Average B, all atoms (Å^2)	22.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.96% 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: NA, PO4

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	0.69	0/2662	0.75	0/3638

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	A	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	186	GLY	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	2581	0	2540	12	0
2	A	10	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	1	0	0	0	0
4	A	382	0	0	5	0
All	All	2974	0	2540	12	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 2.

All (12) 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:9:GLU:O	4:A:1074:HOH:O	1.98	0.80
1:A:9:GLU:C	4:A:1074:HOH:O	2.28	0.70
1:A:140:MSE:HE1	1:A:148:ILE:CD1	2.22	0.70
1:A:140:MSE:HE1	1:A:148:ILE:HD12	1.73	0.69
1:A:9:GLU:OE2	4:A:994:HOH:O	2.16	0.61
1:A:140:MSE:CE	1:A:148:ILE:HD12	2.39	0.51
1:A:16:LEU:HD12	4:A:1067:HOH:O	2.12	0.49
1:A:140:MSE:HE3	1:A:144:GLU:HB3	1.94	0.47
1:A:140:MSE:HE1	1:A:148:ILE:HD11	1.95	0.45
1:A:69:HIS:CG	1:A:69:HIS:O	2.71	0.43
1:A:14:GLU:HB2	1:A:15:PRO:HD2	2.01	0.42
1:A:311:THR:HG23	4:A:870:HOH:O	2.21	0.41

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	322/332 (97%)	316 (98%)	6 (2%)	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	263/267 (98%)	262 (100%)	1 (0%)	91 79

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

Mol	Chain	Res	Type
1	A	273	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	20	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 3 ligands modelled in this entry, 1 is monoatomic - leaving 2 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	$\#\|Z\| > 2$
2	PO4	A	602	-	4,4,4	0.89	0	6,6,6	0.72	0
2	PO4	A	601	-	4,4,4	0.99	0	6,6,6	1.47	1 (16%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	601	PO4	O4-P-O1	2.98	121.81	110.89

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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	319/332 (96%)	0.13	20 (6%)	20	17	12, 18, 35, 49	1 (0%)

All (20) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	19	ALA	10.5
1	A	17	THR	6.2
1	A	22	VAL	6.0
1	A	23[A]	TRP	5.1
1	A	21	GLY	4.9
1	A	18	THR	4.8
1	A	0	SER	4.4
1	A	69	HIS	4.3
1	A	320	LEU	3.7
1	A	20	GLN	3.5
1	A	102	ARG	3.3
1	A	188	GLY	3.1
1	A	319	ARG	2.7
1	A	184	ARG	2.7
1	A	186	GLY	2.5
1	A	258	ARG	2.5
1	A	315	ARG	2.4
1	A	187	ASN	2.1
1	A	318	GLU	2.1
1	A	24	ARG	2.0

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 (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

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, 95^{th} 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
2	PO4	A	602	5/5	0.97	0.17	27,28,31,31	0
2	PO4	A	601	5/5	0.97	0.17	25,25,27,27	0
3	NA	A	701	1/1	0.99	0.04	25,25,25,25	0

6.5 Other polymers (i)

There are no such residues in this entry.