



wwPDB X-ray Structure Validation Summary Report

Mar 1, 2014 – 02:55 AM GMT

PDB ID : 3EPM
Title : Crystal structure of Caulobacter crescentus ThiC
Authors : Li, S.; Chatterjee, A.; Zhang, Y.; Grove, T.L.; Lee, M.; Krebs, C.; Booker, S.J.; Begley, T.P.; Ealick, S.E.
Deposited on : 2008-09-29
Resolution : 2.79 Å(reported)

This is a wwPDB validation summary report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

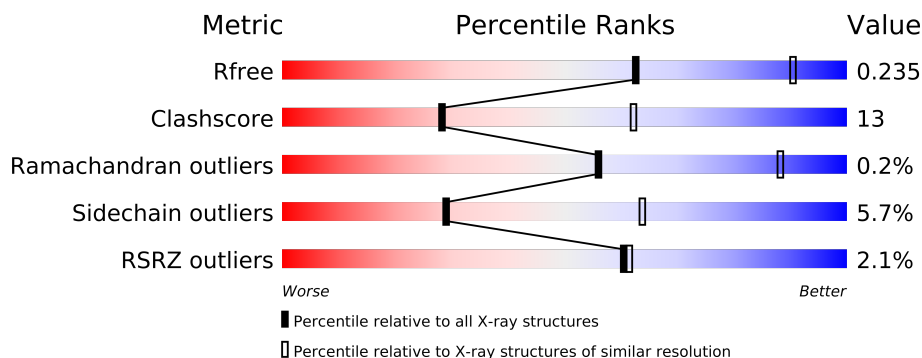
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22683

1 Overall quality at a glance

The reported resolution of this entry is 2.79 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}	66092	1799 (2.80-2.80)
Clashscore	79885	2295 (2.80-2.80)
Ramachandran outliers	78287	2252 (2.80-2.80)
Sidechain outliers	78261	2254 (2.80-2.80)
RSRZ outliers	66119	1802 (2.80-2.80)

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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	612	
1	B	612	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
3	HMH	A	1001	-	X

2 Entry composition i

There are 5 unique types of molecules in this entry. The entry contains 8210 atoms, of which 0 are hydrogen and 0 are deuterium.

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.

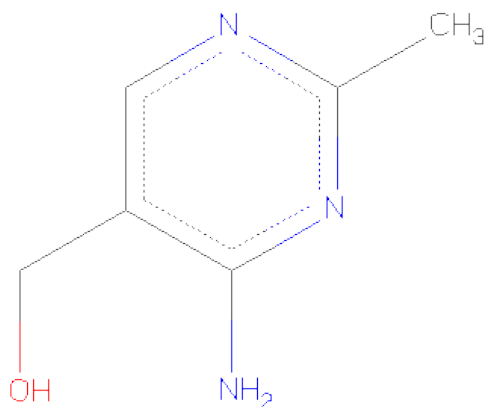
- Molecule 1 is a protein called Thiamine biosynthesis protein thiC.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	515	Total	C	N	O	S	Se	0	0	0
			4042	2558	715	751	6	12			
1	B	512	Total	C	N	O	S	Se	0	0	0
			4018	2543	711	746	6	12			

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total	Zn	0	0
			1	1		
2	A	1	Total	Zn	0	0
			1	1		

- Molecule 3 is 4-AMINO-5-HYDROXYMETHYL-2-METHYLPYRIMIDINE (three-letter code: HMH) (formula: C₆H₉N₃O).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			10	6	3	1		

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



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

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	91	Total	O	0	0
			91	91		
5	B	42	Total	O	0	0
			42	42		

THR	N426	THR
LEU		
PRO	E429	
LYS	Q430	
GLU	L431	
ALA	K432	
HIS	H433	
LYS	C434	
THR	H435	
ALA	E436	
HIS		
PHE	F439	
CYS	V440	
SER	T441	
MSE		
CYS	P444	
GLY	L445	
PRO		
LYS	D448	
PHE		
CYS	T457	
SER	S458	
MSE	L459	
LYS	T460	
TLE	G461	
SER		
GLN	M464	
GLU	I465	
VAL		
ARG	T470	
ASP		
PHE	Y475	
ALA	V476	
ALA	T477	
GLY	P478	
LYS	K479	
ALA	E480	
PRO	H481	
ASN		
SER	L484	
ALA		
GLU	R487	
LEU		
GLY		
MSE	V490	
ALA	K491	
GLU		
GLY	V494	
MSE		
SER	L499	
GLU		
LYS		
PHE	L506	
ARG		
GLU	D517	
GLN		
GLY	I520	
SER		
GLU	H546	
TLE	ASP	
TYR	GLU	

LEU

LYS

THR

GLU

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	63.25Å 103.44Å 95.37Å 90.00° 91.62° 90.00°	Depositor
Resolution (Å)	43.29 – 2.79 47.44 – 2.79	Depositor EDS
% Data completeness (in resolution range)	99.5 (43.29-2.79) 99.5 (47.44-2.79)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	0.14	Depositor
$\langle I/\sigma(I) \rangle$ ¹	5.57 (at 2.81Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.181 , 0.244 0.173 , 0.235	Depositor DCC
R_{free} test set	1537 reflections (5.05%)	DCC
Wilson B-factor (Å ²)	44.4	Xtriage
Anisotropy	0.568	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 18.9	EDS
Estimated twinning fraction	0.032 for h,-k,-l	Xtriage
L-test for twinning	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 30451 reflections	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	8210	wwPDB-VP
Average B, all atoms (Å ²)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.67% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, HMH, SO4

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	0.35	0/4128	0.53	0/5587
1	B	0.34	0/4104	0.52	0/5554
All	All	0.34	0/8232	0.52	0/11141

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	4042	0	3962	94	0
1	B	4018	0	3936	120	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	10	0	9	1	0
4	A	5	0	0	0	0
5	A	91	0	0	2	0
5	B	42	0	0	2	0
All	All	8210	0	7907	202	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 13.

The worst 5 of 202 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:372:LEU:HB3	1:A:396:LEU:HD22	1.52	0.91
1:A:197:ASN:HD22	1:A:199:ASN:H	1.16	0.91
1:B:197:ASN:ND2	1:B:199:ASN:H	1.70	0.89
1:A:318:LEU:HB3	1:A:319:PRO:HD3	1.54	0.88
1:A:457:THR:HG22	1:B:499:LEU:HD13	1.58	0.84

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	509/612 (83%)	483 (95%)	25 (5%)	1 (0%)	56	88
1	B	506/612 (83%)	479 (95%)	26 (5%)	1 (0%)	56	88
All	All	1015/1224 (83%)	962 (95%)	51 (5%)	2 (0%)	56	88

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	160	ASP
1	A	131	VAL

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	423/486 (87%)	398 (94%)	25 (6%)	28	62
1	B	420/486 (86%)	397 (94%)	23 (6%)	30	65
All	All	843/972 (87%)	795 (94%)	48 (6%)	29	64

5 of 48 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	487	ARG
1	B	131	VAL
1	B	499	LEU
1	A	506	LEU
1	A	520	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 34 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	430	GLN
1	B	154	ASN
1	B	430	GLN
1	A	546	HIS
1	A	255	ASN

5.3.3 RNA ⓘ

There are no RNA chains 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 carbohydrates in this entry.

5.6 Ligand geometry ⓘ

Of 4 ligands modelled in this entry, 2 are 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	$\# Z > 2$
3	HMH	A	1001	-	10,10,10	0.89	0	13,13,13	1.99	5 (38%)
4	SO4	A	2001	-	4,4,4	0.13	0	6,6,6	0.09	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	HMH	A	1001	-	-	0/2/2/2	0/1/1/1
4	SO4	A	2001	-	-	0/0/0/0	0/0/0/0

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	1001	HMH	C6A-N1A-C2A	3.43	121.70	115.68
3	A	1001	HMH	N1A-C2A-N3A	-3.34	119.63	125.65
3	A	1001	HMH	N4A-C4A-N3A	2.68	120.77	116.88
3	A	1001	HMH	C5A-C4A-N4A	-2.22	118.92	122.23
3	A	1001	HMH	CM2-C2A-N1A	2.21	119.65	117.02

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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	515/612 (84%)	-0.09	10 (1%) 64 64	21, 39, 67, 121	0
1	B	512/612 (83%)	-0.11	12 (2%) 57 58	22, 41, 66, 119	0
All	All	1027/1224 (83%)	-0.10	22 (2%) 60 61	21, 40, 67, 121	0

The worst 5 of 22 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	13	THR	4.2
1	A	15	SER	4.0
1	B	94	ASP	3.7
1	A	162	PRO	3.7
1	A	164	VAL	3.3

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 carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	RSR	LLDF	B-factors(\AA^2)	Q<0.9
3	HMH	A	1001	10/10	0.42	5.23	55,80,88,94	0
4	SO4	A	2001	5/5	0.15	-0.05	59,68,96,111	0
2	ZN	A	613	1/1	0.12	-1.39	56,56,56,56	0
2	ZN	B	613	1/1	0.08	-6.43	69,69,69,69	0

6.5 Other polymers ⓘ

There are no such residues in this entry.