



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 8, 2026 – 06:14 AM UTC

PDB ID : 3EPM / pdb_00003epm
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 X-ray Structure Validation Summary 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 ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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

MolProbity : 4-5-2 with Phenix2.0
Mogul : 2022.3.0, CSD as543be (2022)
Xtriage (Phenix) : 2.0
EDS : 3.0
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4 : 9.0.010 (Gargrove)
Density-Fitness : 1.0.12
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.49

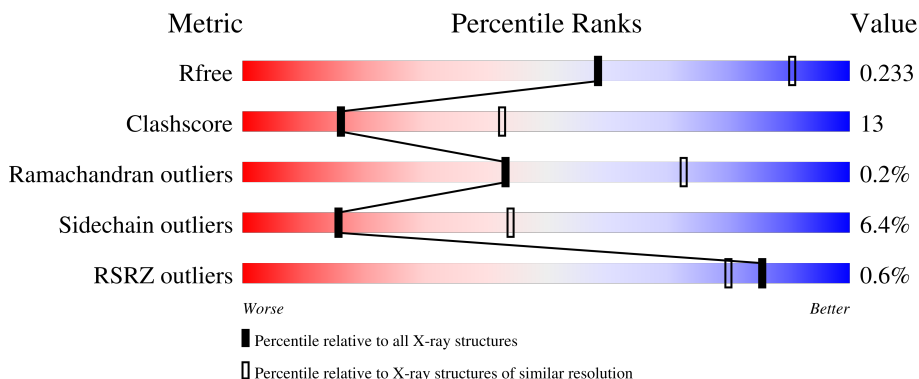
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.79 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}	180053	3866 (2.80-2.80)
Clashscore	190562	4276 (2.80-2.80)
Ramachandran outliers	187476	4196 (2.80-2.80)
Sidechain outliers	187428	4198 (2.80-2.80)
RSRZ outliers	180081	3869 (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 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 $\leq 5\%$. 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	612	 62% 19% • 16%
1	B	612	 57% 24% • 16%

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 8210 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.

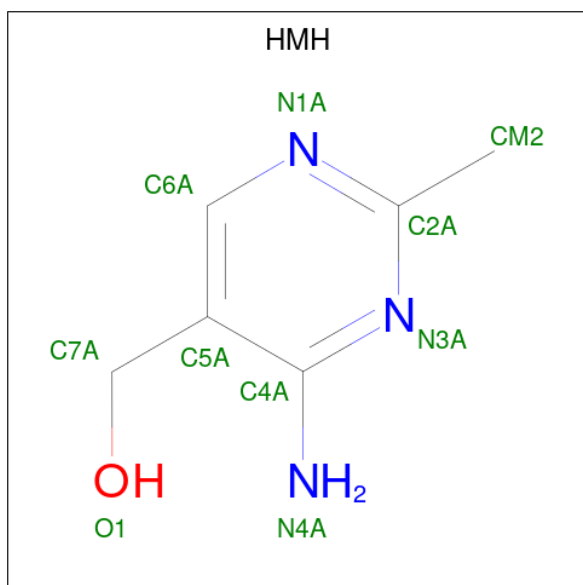
- Molecule 1 is a protein called Thiamine biosynthesis protein thiC.

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

- Molecule 2 is ZINC ION (CCD ID: ZN) (formula: Zn).

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

- Molecule 3 is 4-AMINO-5-HYDROXYMETHYL-2-METHYLPYRIMIDINE (CCD ID: HMH) (formula: C₆H₉N₃O).



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

- Molecule 4 is SULFATE ION (CCD ID: SO4) (formula: O₄S).



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

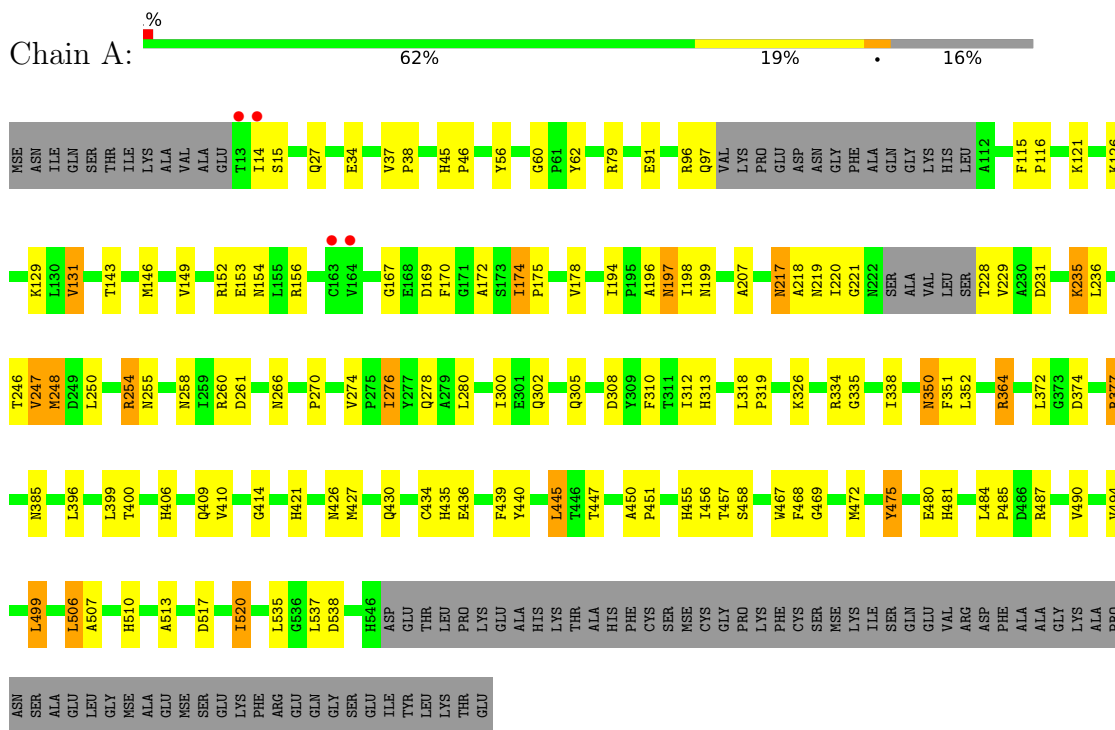
- Molecule 5 is water.

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

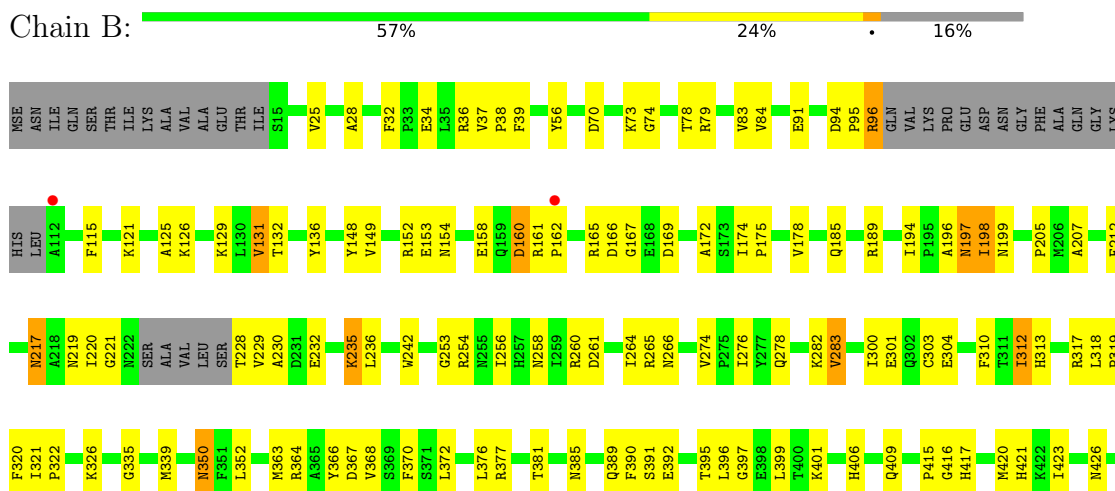
3 Residue-property plots

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: Thiamine biosynthesis protein thiC



- Molecule 1: Thiamine biosynthesis protein thiC



E429	LEU
Q430	PRO
L431	LYS
K432	GLU
H433	ALA
C434	HIS
H435	LYS
E436	THR
T441	ALA
	HIS
	PHE
P444	CYS
L445	SER
	NSE
D448	CYS
	GLY
I456	PRO
T457	LYS
S458	PHE
K459	CYS
I460	SER
G461	NSE
	LYS
M464	ILE
I465	SER
	GLN
T470	GLU
	VAL
Y475	ARG
V476	ASP
T477	PHE
P478	ALA
K479	ALA
E480	GLY
H481	LYS
	ALA
L484	PRO
P485	ASN
D486	SER
R487	ALA
	GLU
V490	LEU
K491	GLY
	NSE
V494	ALA
	GLU
L499	NSE
	SER
L506	GLU
	LYS
H510	PHE
	ARG
D517	GLU
	GLN
I520	GLY
	SER
H546	GLU
ASP	ILE
GLU	TYR
THR	LEU

LYS
THR
GLU

4 Data and refinement statistics i

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 43.29 – 2.79	Depositor EDS
% Data completeness (in resolution range)	99.5 (43.29-2.79) 99.5 (43.29-2.79)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	0.14	Depositor
$\langle I/\sigma(I) \rangle$ ¹	5.57 (at 2.81Å)	Xtrriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.181 , 0.244 0.173 , 0.233	Depositor DCC
R_{free} test set	1538 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å ²)	44.4	Xtrriage
Anisotropy	0.568	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 36.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.032 for h,-k,-l	Xtrriage
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

Xtrriage'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.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

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

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	0.46	0/4128	0.85	12/5587 (0.2%)
1	B	0.43	0/4104	0.82	4/5554 (0.1%)
All	All	0.45	0/8232	0.84	16/11141 (0.1%)

There are no bond length outliers.

The worst 5 of 16 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	32	PHE	CA-C-N	7.49	127.20	119.56
1	B	32	PHE	C-N-CA	7.49	127.20	119.56
1	B	510	HIS	CA-C-N	6.74	127.28	119.47
1	B	510	HIS	C-N-CA	6.74	127.28	119.47
1	A	510	HIS	CA-C-N	6.61	126.55	119.28

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

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	4042	0	3962	98	0
1	B	4018	0	3936	123	0
2	A	1	0	0	0	0
2	B	1	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	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	209	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 13.

The worst 5 of 209 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: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 [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	509/612 (83%)	483 (95%)	25 (5%)	1 (0%)	43	72
1	B	506/612 (83%)	479 (95%)	26 (5%)	1 (0%)	43	72
All	All	1015/1224 (83%)	962 (95%)	51 (5%)	2 (0%)	43	72

All (2) Ramachandran outliers are listed below:

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

5.3.2 Protein sidechains [i](#)

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%)	397 (94%)	26 (6%)	17	46
1	B	420/486 (86%)	392 (93%)	28 (7%)	15	42
All	All	843/972 (87%)	789 (94%)	54 (6%)	16	44

5 of 54 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	96	ARG
1	B	236	LEU
1	B	487	ARG
1	B	131	VAL
1	B	217	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 35 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	385	ASN
1	B	389	GLN
1	B	430	GLN
1	A	389	GLN
1	A	385	ASN

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

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
4	SO4	A	2001	-	4,4,4	0.26	0	6,6,6	0.09	0
3	HMH	A	1001	-	10,10,10	1.11	1 (10%)	12,13,13	2.21	7 (58%)

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	-	-	1/2/2/2	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	1001	HMH	C5A-C4A	-2.26	1.39	1.42

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	1001	HMH	N1A-C2A-N3A	-3.54	119.63	125.53
3	A	1001	HMH	C6A-N1A-C2A	3.43	121.70	116.07

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	1001	HMH	N4A-C4A-N3A	2.78	120.77	117.03
3	A	1001	HMH	CM2-C2A-N3A	2.40	120.72	117.13
3	A	1001	HMH	C5A-C4A-N4A	-2.39	118.92	122.14

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	1001	HMH	C6A-C5A-C7A-O1

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1001	HMH	1	0

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 [i](#)

6.1 Protein, DNA and RNA chains [i](#)

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	503/612 (82%)	-0.19	4 (0%) 82 75	21, 39, 67, 121	0
1	B	500/612 (81%)	-0.23	2 (0%) 88 84	22, 41, 67, 119	0
All	All	1003/1224 (81%)	-0.21	6 (0%) 85 80	21, 40, 67, 121	0

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	164	VAL	2.5
1	A	13	THR	2.4
1	B	112	ALA	2.2
1	A	14	ILE	2.2
1	A	163	CYS	2.1

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no oligosaccharides 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, 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(\AA^2)	Q<0.9
3	HMH	A	1001	10/10	0.64	0.21	55,80,88,94	0
4	SO4	A	2001	5/5	0.85	0.08	59,68,96,111	0
2	ZN	B	613	1/1	0.96	0.05	69,69,69,69	0
2	ZN	A	613	1/1	0.97	0.04	56,56,56,56	0

6.5 Other polymers [i](#)

There are no such residues in this entry.