



Full wwPDB X-ray Structure Validation Report ⓘ

Jun 19, 2024 – 12:15 AM EDT

PDB ID : 4EHS
Title : Crystal structure of Helicobacter pylori DnaG Primase C terminal domain
Authors : Abdul Rehman, S.A.; Gourinath, S.
Deposited on : 2012-04-04
Resolution : 1.78 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 ⓘ 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 : 4.02b-467
Mogul : 2022.3.0, CSD as543be (2022)
Xtriage (Phenix) : 1.20.1
EDS : 2.37.1
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) : 2.37.1

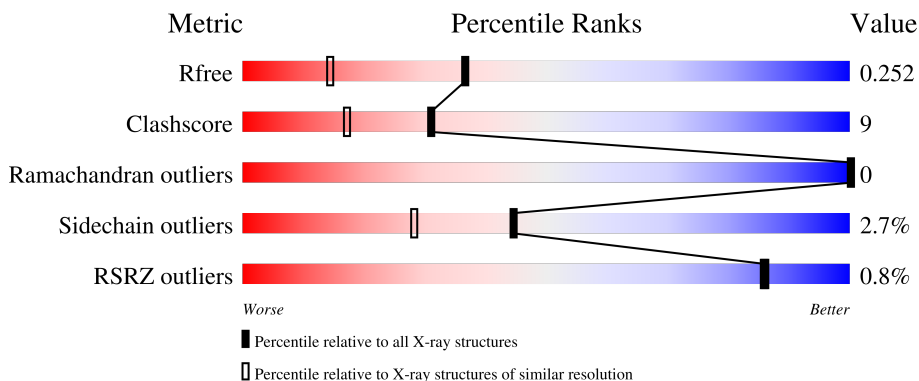
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.78 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	9185 (1.80-1.76)
Clashscore	141614	10184 (1.80-1.76)
Ramachandran outliers	138981	10051 (1.80-1.76)
Sidechain outliers	138945	10050 (1.80-1.76)
RSRZ outliers	127900	9032 (1.80-1.76)

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	155	 68% 10% • 20%
1	B	155	 % 65% 11% • 21%

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 2222 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 DNA primase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	124	1004	657	167	176	2	2	0	0	0
1	B	122	989	649	165	171	2	2	0	0	0

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	560	LEU	-	EXPRESSION TAG	UNP P56064
A	561	GLU	-	EXPRESSION TAG	UNP P56064
A	562	HIS	-	EXPRESSION TAG	UNP P56064
A	563	HIS	-	EXPRESSION TAG	UNP P56064
A	564	HIS	-	EXPRESSION TAG	UNP P56064
A	565	HIS	-	EXPRESSION TAG	UNP P56064
A	566	HIS	-	EXPRESSION TAG	UNP P56064
A	567	HIS	-	EXPRESSION TAG	UNP P56064
B	560	LEU	-	EXPRESSION TAG	UNP P56064
B	561	GLU	-	EXPRESSION TAG	UNP P56064
B	562	HIS	-	EXPRESSION TAG	UNP P56064
B	563	HIS	-	EXPRESSION TAG	UNP P56064
B	564	HIS	-	EXPRESSION TAG	UNP P56064
B	565	HIS	-	EXPRESSION TAG	UNP P56064
B	566	HIS	-	EXPRESSION TAG	UNP P56064
B	567	HIS	-	EXPRESSION TAG	UNP P56064

- Molecule 2 is BETA-MERCAPTOETHANOL (three-letter code: BME) (formula: C₂H₆OS).



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

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	103	Total	O	0	0
			103	103		
3	B	110	Total	O	0	0
			110	110		

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	48.88Å 61.45Å 82.42Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.27 – 1.78 49.27 – 1.78	Depositor EDS
% Data completeness (in resolution range)	99.6 (49.27-1.78) 99.6 (49.27-1.78)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.51 (at 1.77Å)	Xtrriage
Refinement program	REFMAC refmac_5.5.0072	Depositor
R, R_{free}	0.195 , 0.249 0.198 , 0.252	Depositor DCC
R_{free} test set	1250 reflections (5.11%)	wwPDB-VP
Wilson B-factor (Å ²)	22.2	Xtrriage
Anisotropy	0.385	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 38.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	2222	wwPDB-VP
Average B, all atoms (Å ²)	23.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.28% 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: BME

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	0.94	0/1017	1.01	2/1358 (0.1%)
1	B	0.98	1/1002 (0.1%)	1.25	12/1338 (0.9%)
All	All	0.96	1/2019 (0.0%)	1.13	14/2696 (0.5%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	521	GLU	CD-OE1	5.30	1.31	1.25

All (14) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	545	ARG	NE-CZ-NH2	-13.39	113.61	120.30
1	B	545	ARG	NE-CZ-NH1	11.38	125.99	120.30
1	B	513	ARG	NE-CZ-NH1	9.72	125.16	120.30
1	B	513	ARG	NE-CZ-NH2	-8.91	115.84	120.30
1	A	545	ARG	NE-CZ-NH1	-7.82	116.39	120.30
1	B	498	ASP	CB-CG-OD1	7.47	125.02	118.30
1	B	560	LEU	CB-CG-CD2	-7.10	98.92	111.00
1	B	454	LEU	CA-CB-CG	6.65	130.59	115.30
1	B	518	ARG	NE-CZ-NH1	6.02	123.31	120.30
1	B	509	GLU	OE1-CD-OE2	5.71	130.16	123.30
1	B	545	ARG	CG-CD-NE	-5.49	100.28	111.80
1	B	491	LYS	CD-CE-NZ	-5.32	99.46	111.70
1	B	498	ASP	CB-CG-OD2	-5.07	113.74	118.30
1	A	461	ASP	CB-CG-OD1	5.04	122.84	118.30

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	1004	0	1079	26	0
1	B	989	0	1068	15	0
2	A	8	0	10	3	0
2	B	8	0	10	0	0
3	A	103	0	0	12	0
3	B	110	0	0	5	0
All	All	2222	0	2167	38	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 9.

All (38) 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:B:550:LYS:HE2	3:B:777:HOH:O	1.60	1.02
1:A:510:LYS:HD2	3:A:799:HOH:O	1.59	1.00
1:A:455:GLU:CG	3:A:801:HOH:O	2.20	0.89
1:A:455:GLU:HG2	3:A:801:HOH:O	1.73	0.84
1:A:484:GLN:HB3	3:A:732:HOH:O	1.79	0.81
1:B:447:LYS:HG2	1:B:475:LYS:HE3	1.69	0.75
1:B:550:LYS:CE	3:B:777:HOH:O	2.25	0.74
1:A:483:CYS:HB3	2:A:601:BME:O1	1.95	0.67
1:A:561:GLU:HA	3:A:733:HOH:O	1.93	0.67
1:A:534:PHE:CE2	1:A:538:MSE:SE	2.99	0.65
1:A:451:ARG:NE	3:A:801:HOH:O	2.31	0.64
1:B:550:LYS:NZ	3:B:777:HOH:O	2.31	0.61
1:B:440:PRO:O	1:B:518:ARG:NH2	2.26	0.61
1:A:455:GLU:HG3	3:A:801:HOH:O	1.96	0.58
1:A:558:ALA:HB1	1:B:560:LEU:HD11	1.85	0.58
1:A:454:LEU:HB3	1:A:496:LEU:HD23	1.85	0.58
1:A:496:LEU:O	1:A:496:LEU:HD13	2.06	0.56
1:A:451:ARG:CD	3:A:801:HOH:O	2.58	0.51
1:A:451:ARG:HD2	3:A:801:HOH:O	2.11	0.50
1:A:561:GLU:H	1:A:561:GLU:CD	2.14	0.49
1:A:483:CYS:HB3	2:A:601:BME:C1	2.43	0.49

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:457:ARG:HH11	1:B:457:ARG:HG2	1.79	0.48
1:B:457:ARG:HG2	1:B:457:ARG:NH1	2.29	0.48
1:A:561:GLU:CA	3:A:733:HOH:O	2.57	0.47
1:A:457:ARG:HH12	1:A:482:LEU:HD12	1.78	0.47
1:A:560:LEU:HD11	1:B:558:ALA:HB1	1.97	0.47
1:B:549:MSE:O	1:B:553:GLN:HG3	2.16	0.46
1:A:483:CYS:O	1:B:539:ILE:HD12	2.18	0.43
1:B:550:LYS:HE3	1:B:550:LYS:HB3	1.39	0.43
1:A:561:GLU:HB3	3:A:733:HOH:O	2.19	0.42
1:B:447:LYS:CG	1:B:475:LYS:HE3	2.44	0.42
1:A:454:LEU:HD21	1:A:492:LEU:HG	2.02	0.42
1:B:469:SER:HB3	1:B:480:ASP:OD1	2.19	0.42
1:A:457:ARG:NH1	1:A:482:LEU:HD12	2.34	0.41
1:A:543:LYS:HG3	3:B:743:HOH:O	2.19	0.41
1:A:561:GLU:CB	3:A:733:HOH:O	2.69	0.41
1:B:474:HIS:HE1	3:B:742:HOH:O	2.03	0.40
1:A:543:LYS:HD2	2:A:602:BME:H22	2.04	0.40

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	122/155 (79%)	118 (97%)	4 (3%)	0	100	100
1	B	120/155 (77%)	120 (100%)	0	0	100	100
All	All	242/310 (78%)	238 (98%)	4 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	111/139 (80%)	109 (98%)	2 (2%)	59	45
1	B	109/139 (78%)	105 (96%)	4 (4%)	34	17
All	All	220/278 (79%)	214 (97%)	6 (3%)	44	28

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	543	LYS
1	A	561	GLU
1	B	443	GLU
1	B	454	LEU
1	B	469	SER
1	B	550	LYS

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	484	GLN
1	B	474	HIS

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 monosaccharides in this entry.

5.6 Ligand geometry

4 ligands are 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
2	BME	B	602	1	3,3,3	0.91	0	2,2,2	1.12	0
2	BME	B	601	1	3,3,3	0.29	0	2,2,2	0.33	0
2	BME	A	601	1	3,3,3	0.62	0	2,2,2	0.32	0
2	BME	A	602	1	3,3,3	1.11	0	2,2,2	1.68	1 (50%)

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
2	BME	B	602	1	-	0/1/1/1	-
2	BME	B	601	1	-	0/1/1/1	-
2	BME	A	601	1	-	1/1/1/1	-
2	BME	A	602	1	-	1/1/1/1	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	602	BME	O1-C1-C2	-2.34	101.68	110.82

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	601	BME	O1-C1-C2-S2
2	A	602	BME	O1-C1-C2-S2

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	601	BME	2	0
2	A	602	BME	1	0

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 [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	122/155 (78%)	-0.18	0 100 100	13, 22, 35, 51	0
1	B	120/155 (77%)	-0.22	2 (1%) 70 70	12, 21, 32, 62	0
All	All	242/310 (78%)	-0.20	2 (0%) 86 86	12, 22, 34, 62	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	439	ALA	3.1
1	B	440	PRO	2.2

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 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, 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
2	BME	A	602	4/4	0.86	0.39	38,42,43,44	0
2	BME	A	601	4/4	0.90	0.17	36,42,44,49	0
2	BME	B	601	4/4	0.94	0.13	32,38,39,52	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
2	BME	B	602	4/4	0.96	0.12	35,39,39,43	0

6.5 Other polymers [i](#)

There are no such residues in this entry.