



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 5, 2026 – 07:03 AM UTC

PDB ID : 9U75 / pdb_00009u75
Title : 5hmC specific restriction endonuclease Escherichia coli E. coli O157:H7 PD-T4-3
Authors : Yu, Y.; Liu, R.
Deposited on : 2025-03-24
Resolution : 2.80 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
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

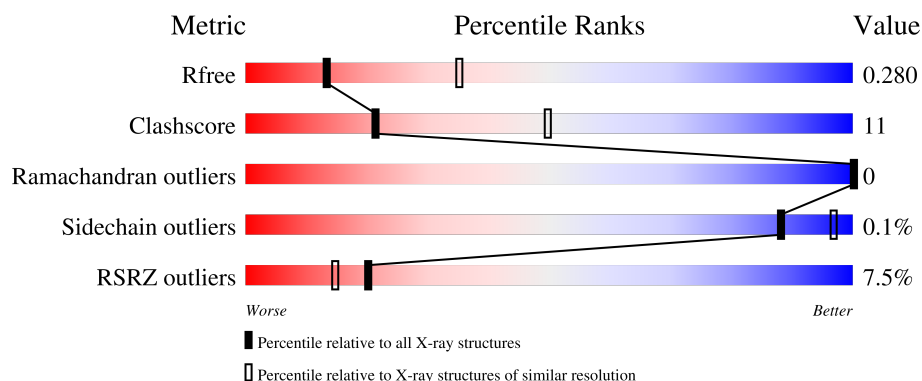
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.80 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	262	3% 78% 17% 5%
1	B	262	4% 77% 18% 5%
1	C	262	6% 78% 16% 5%
1	D	262	15% 63% 31% 6%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 7960 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 GIY-YIG domain-containing protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	248	Total 1992	C 1269	N 350	O 368	S 5	0	0	0
1	B	248	Total 1992	C 1269	N 350	O 368	S 5	0	0	0
1	C	248	Total 1992	C 1269	N 350	O 368	S 5	0	0	0
1	D	247	Total 1980	C 1260	N 349	O 366	S 5	0	0	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-4	SER	-	expression tag	UNP A0A9Q7EI50
A	-3	ALA	-	expression tag	UNP A0A9Q7EI50
A	-2	SER	-	expression tag	UNP A0A9Q7EI50
A	-1	GLY	-	expression tag	UNP A0A9Q7EI50
A	0	SER	-	expression tag	UNP A0A9Q7EI50
B	-4	SER	-	expression tag	UNP A0A9Q7EI50
B	-3	ALA	-	expression tag	UNP A0A9Q7EI50
B	-2	SER	-	expression tag	UNP A0A9Q7EI50
B	-1	GLY	-	expression tag	UNP A0A9Q7EI50
B	0	SER	-	expression tag	UNP A0A9Q7EI50
C	-4	SER	-	expression tag	UNP A0A9Q7EI50
C	-3	ALA	-	expression tag	UNP A0A9Q7EI50
C	-2	SER	-	expression tag	UNP A0A9Q7EI50
C	-1	GLY	-	expression tag	UNP A0A9Q7EI50
C	0	SER	-	expression tag	UNP A0A9Q7EI50
D	-4	SER	-	expression tag	UNP A0A9Q7EI50
D	-3	ALA	-	expression tag	UNP A0A9Q7EI50
D	-2	SER	-	expression tag	UNP A0A9Q7EI50
D	-1	GLY	-	expression tag	UNP A0A9Q7EI50
D	0	SER	-	expression tag	UNP A0A9Q7EI50

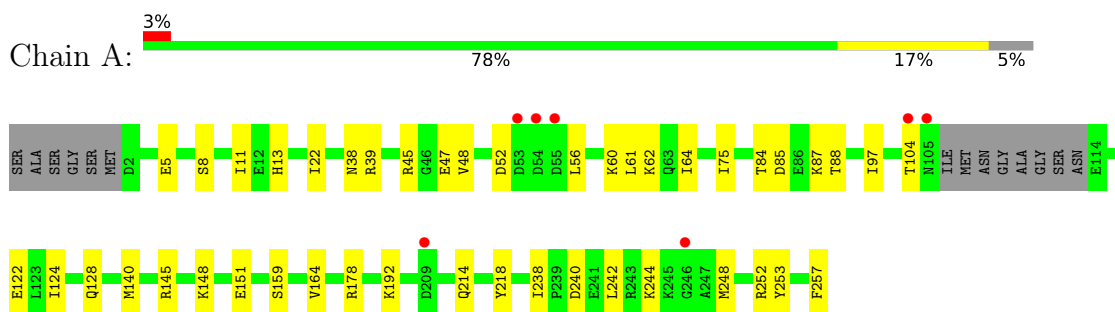
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	2	Total 2	O 2	0	0
2	C	1	Total 1	O 1	0	0
2	D	1	Total 1	O 1	0	0

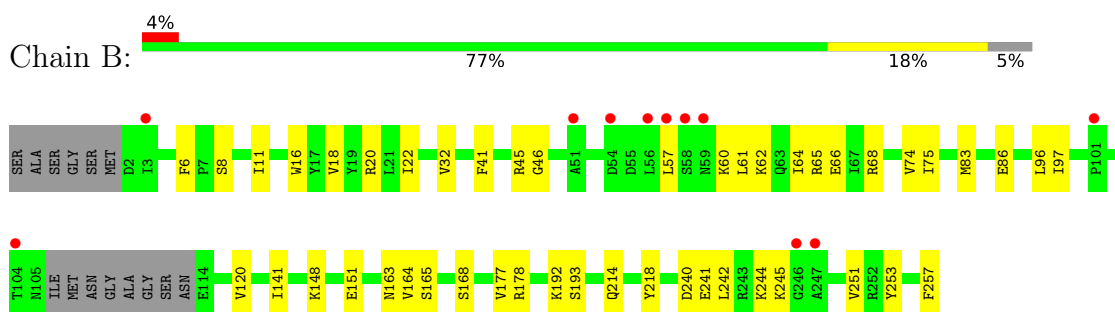
3 Residue-property plots [i](#)

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.

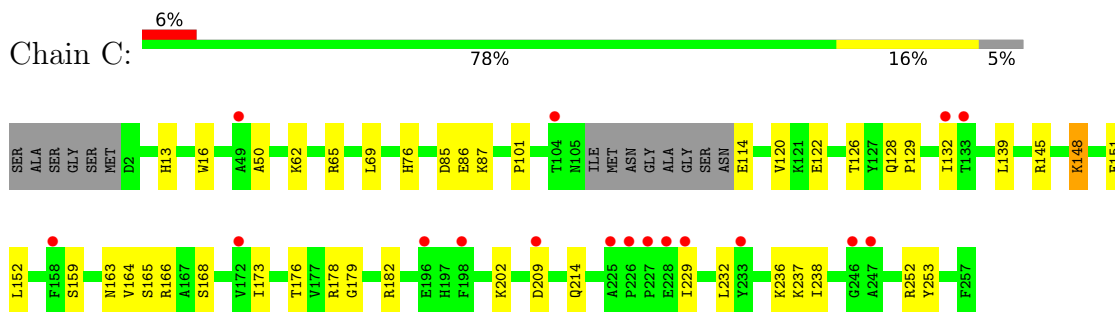
- Molecule 1: GIY-YIG domain-containing protein



- Molecule 1: GIY-YIG domain-containing protein

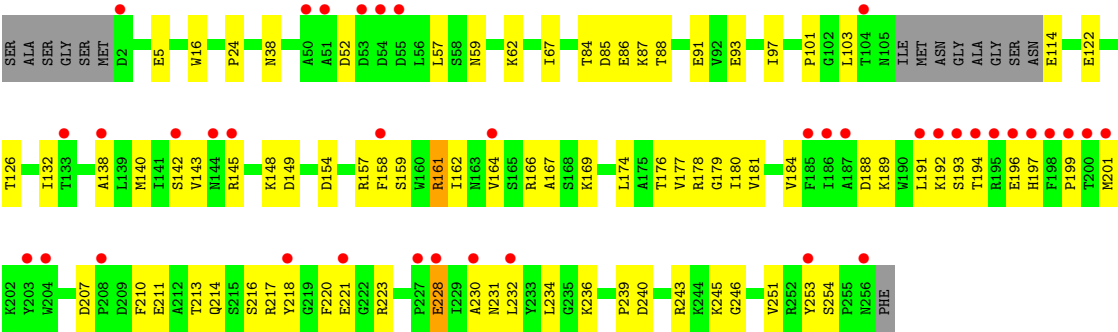


- Molecule 1: GIY-YIG domain-containing protein



- Molecule 1: GIY-YIG domain-containing protein





4 Data and refinement statistics

Property	Value	Source
Space group	P 41	Depositor
Cell constants a, b, c, α , β , γ	103.63Å 103.63Å 125.83Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.91 – 2.80 47.91 – 2.80	Depositor EDS
% Data completeness (in resolution range)	93.8 (47.91-2.80) 93.8 (47.91-2.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.17 (at 2.81Å)	Xtriage
Refinement program	PHENIX 1.8.4_1496	Depositor
R, R_{free}	0.226 , 0.286 0.226 , 0.280	Depositor DCC
R_{free} test set	994 reflections (3.05%)	wwPDB-VP
Wilson B-factor (Å ²)	46.3	Xtriage
Anisotropy	0.053	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 40.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.029 for h,-k,-l	Xtriage
F_o, F_c correlation	0.90	EDS
Total number of atoms	7960	wwPDB-VP
Average B, all atoms (Å ²)	52.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.05% 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](#)

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	0.21	0/2040	0.42	0/2756
1	B	0.17	0/2040	0.40	0/2756
1	C	0.23	0/2040	0.41	0/2756
1	D	0.19	0/2027	0.46	0/2740
All	All	0.20	0/8147	0.42	0/11008

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	C	0	1
1	D	0	2
All	All	0	3

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	148	LYS	Peptide
1	D	158	PHE	Peptide
1	D	228	GLU	Peptide

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	1992	0	1960	31	0
1	B	1992	0	1960	35	0
1	C	1992	0	1960	42	0
1	D	1980	0	1951	74	0
2	A	2	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
All	All	7960	0	7831	172	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 11.

The worst 5 of 172 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:C:229:ILE:HA	1:C:232:LEU:CD1	1.70	1.20
1:C:229:ILE:HA	1:C:232:LEU:HD12	1.27	1.11
1:D:245:LYS:HD3	1:D:246:GLY:N	1.71	1.05
1:C:132:ILE:HG12	1:C:236:LYS:HG3	1.52	0.91
1:D:159:SER:HB3	1:D:217:ARG:HH21	1.33	0.91

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	244/262 (93%)	240 (98%)	4 (2%)	0	100	100
1	B	244/262 (93%)	239 (98%)	5 (2%)	0	100	100
1	C	244/262 (93%)	240 (98%)	4 (2%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	D	243/262 (93%)	232 (96%)	11 (4%)	0	100	100
All	All	975/1048 (93%)	951 (98%)	24 (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	208/217 (96%)	208 (100%)	0	100	100
1	B	208/217 (96%)	208 (100%)	0	100	100
1	C	208/217 (96%)	208 (100%)	0	100	100
1	D	207/217 (95%)	206 (100%)	1 (0%)	81	93
All	All	831/868 (96%)	830 (100%)	1 (0%)	88	97

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

Mol	Chain	Res	Type
1	D	161	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	231	ASN
1	D	197	HIS
1	C	135	HIS
1	C	38	ASN
1	C	231	ASN

5.3.3 RNA ⓘ

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](#)

There are no ligands in this entry.

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	248/262 (94%)	0.02	7 (2%) 55 45	9, 36, 71, 112	0
1	B	248/262 (94%)	0.21	11 (4%) 39 31	17, 42, 78, 96	0
1	C	248/262 (94%)	0.53	17 (6%) 23 16	20, 54, 83, 105	0
1	D	247/262 (94%)	0.97	39 (15%) 5 4	21, 74, 123, 194	0
All	All	991/1048 (94%)	0.43	74 (7%) 20 15	9, 49, 96, 194	0

The worst 5 of 74 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	198	PHE	5.3
1	D	199	PRO	4.5
1	D	208	PRO	4.2
1	D	104	THR	4.1
1	D	185	PHE	3.8

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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.