



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 5, 2026 – 07:29 PM UTC

PDB ID : 2AR0 / pdb_00002ar0
Title : Crystal structure of Type I restriction enzyme EcoKI M protein (EC 2.1.1.72) (M.EcoKI)
Authors : Rajashankar, K.R.; Kniewel, R.; Lima, C.D.; Burley, S.K.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on : 2005-08-18
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
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

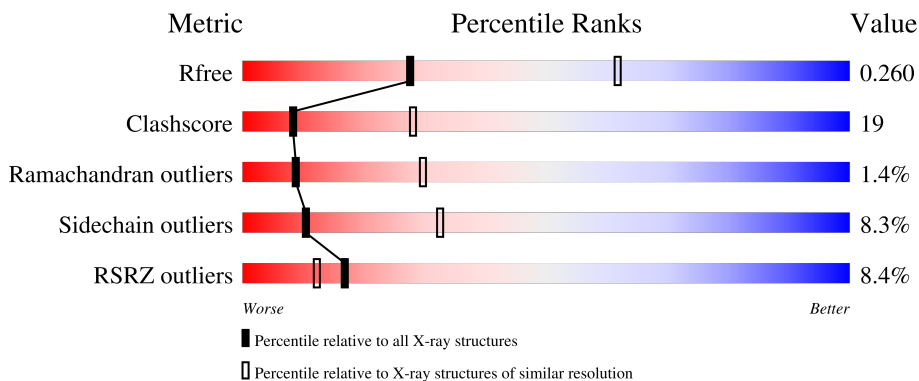
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.80 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	541	
1	B	541	

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 7828 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 Type I restriction enzyme EcoKI M protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	485	3843	2415	671	742	6	9	0	0	0
1	B	483	3832	2408	669	740	6	9	0	0	0

There are 50 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	-	cloning artifact	UNP P08957
A	2	SER	-	cloning artifact	UNP P08957
A	3	LEU	-	cloning artifact	UNP P08957
A	41	MSE	MET	modified residue	UNP P08957
A	76	MSE	MET	modified residue	UNP P08957
A	112	MSE	MET	modified residue	UNP P08957
A	132	MSE	MET	modified residue	UNP P08957
A	229	MSE	MET	modified residue	UNP P08957
A	295	MSE	MET	modified residue	UNP P08957
A	331	MSE	MET	modified residue	UNP P08957
A	386	MSE	MET	modified residue	UNP P08957
A	492	MSE	MET	modified residue	UNP P08957
A	506	MSE	MET	modified residue	UNP P08957
A	530	LYS	-	expression tag	UNP P08957
A	531	GLU	-	expression tag	UNP P08957
A	532	GLU	-	expression tag	UNP P08957
A	533	GLY	-	expression tag	UNP P08957
A	534	GLY	-	expression tag	UNP P08957
A	535	SER	-	expression tag	UNP P08957
A	536	HIS	-	expression tag	UNP P08957
A	537	HIS	-	expression tag	UNP P08957
A	538	HIS	-	expression tag	UNP P08957
A	539	HIS	-	expression tag	UNP P08957
A	540	HIS	-	expression tag	UNP P08957
A	541	HIS	-	expression tag	UNP P08957

Continued on next page...

Continued from previous page...

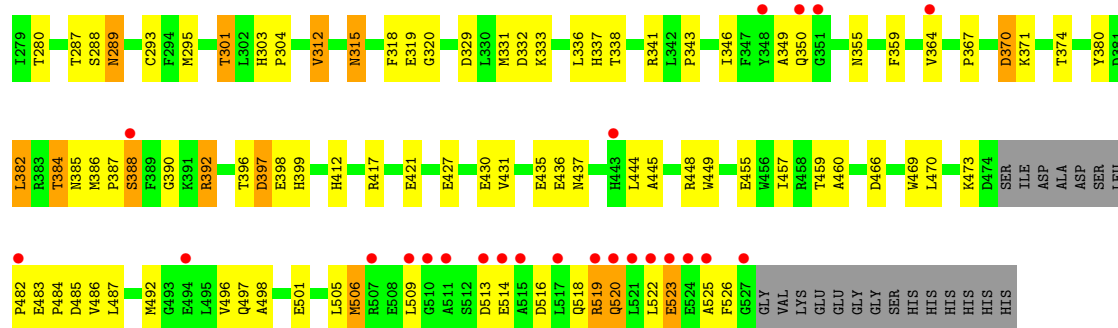
Chain	Residue	Modelled	Actual	Comment	Reference
B	1	MSE	-	cloning artifact	UNP P08957
B	2	SER	-	cloning artifact	UNP P08957
B	3	LEU	-	cloning artifact	UNP P08957
B	41	MSE	MET	modified residue	UNP P08957
B	76	MSE	MET	modified residue	UNP P08957
B	112	MSE	MET	modified residue	UNP P08957
B	132	MSE	MET	modified residue	UNP P08957
B	229	MSE	MET	modified residue	UNP P08957
B	295	MSE	MET	modified residue	UNP P08957
B	331	MSE	MET	modified residue	UNP P08957
B	386	MSE	MET	modified residue	UNP P08957
B	492	MSE	MET	modified residue	UNP P08957
B	506	MSE	MET	modified residue	UNP P08957
B	530	LYS	-	expression tag	UNP P08957
B	531	GLU	-	expression tag	UNP P08957
B	532	GLU	-	expression tag	UNP P08957
B	533	GLY	-	expression tag	UNP P08957
B	534	GLY	-	expression tag	UNP P08957
B	535	SER	-	expression tag	UNP P08957
B	536	HIS	-	expression tag	UNP P08957
B	537	HIS	-	expression tag	UNP P08957
B	538	HIS	-	expression tag	UNP P08957
B	539	HIS	-	expression tag	UNP P08957
B	540	HIS	-	expression tag	UNP P08957
B	541	HIS	-	expression tag	UNP P08957

- Molecule 2 is UNKNOWN ATOM OR ION (CCD ID: UNX) (formula: X).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	10	Total X 10 10	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	86	Total O 86 86	0	0
3	B	57	Total O 57 57	0	0



4 Data and refinement statistics

Property	Value	Source
Space group	P 41	Depositor
Cell constants a, b, c, α , β , γ	105.43Å 105.43Å 138.28Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.81 – 2.80 19.81 – 2.80	Depositor EDS
% Data completeness (in resolution range)	95.3 (19.81-2.80) 96.5 (19.81-2.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.11	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.40 (at 2.80Å)	Xtrriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.224 , 0.261 0.224 , 0.260	Depositor DCC
R_{free} test set	3534 reflections (4.93%)	wwPDB-VP
Wilson B-factor (Å ²)	38.9	Xtrriage
Anisotropy	0.058	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 51.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	0.044 for h,-k,-l	Xtrriage
F_o, F_c correlation	0.90	EDS
Total number of atoms	7828	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 3.73% 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: UNX

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	0.53	0/3915	0.90	6/5292 (0.1%)
1	B	0.53	0/3904	0.89	4/5277 (0.1%)
All	All	0.53	0/7819	0.90	10/10569 (0.1%)

There are no bond length outliers.

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	523	GLU	N-CA-C	-5.99	105.23	112.54
1	A	175	ASP	CA-C-N	5.58	125.42	119.28
1	A	175	ASP	C-N-CA	5.58	125.42	119.28
1	B	175	ASP	CA-C-N	5.49	125.32	119.28
1	B	175	ASP	C-N-CA	5.49	125.32	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	3843	0	3743	148	0
1	B	3832	0	3732	150	0
2	A	10	0	0	0	0
3	A	86	0	0	4	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	57	0	0	2	0
All	All	7828	0	7475	291	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 19.

The worst 5 of 291 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:287:THR:HG22	1:B:289:ASN:H	1.23	1.03
1:A:287:THR:HG22	1:A:289:ASN:H	1.23	1.01
1:B:117:TRP:HB2	3:B:563:HOH:O	1.65	0.95
1:B:219:LEU:HD22	1:B:277:THR:HG23	1.53	0.90
1:A:219:LEU:HD22	1:A:277:THR:HG23	1.57	0.86

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	479/541 (88%)	445 (93%)	27 (6%)	7 (2%)	8 28
1	B	477/541 (88%)	439 (92%)	32 (7%)	6 (1%)	9 31
All	All	956/1082 (88%)	884 (92%)	59 (6%)	13 (1%)	9 30

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	86	LYS
1	A	349	ALA
1	B	86	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	349	ALA
1	A	350	GLN

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	418/451 (93%)	386 (92%)	32 (8%)	12	36
1	B	417/451 (92%)	380 (91%)	37 (9%)	9	29
All	All	835/902 (93%)	766 (92%)	69 (8%)	10	32

5 of 69 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	427	GLU
1	B	435	GLU
1	B	519	ARG
1	A	430	GLU
1	A	427	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 47 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	71	GLN
1	B	242	HIS
1	B	94	ASN
1	B	196	ASN
1	B	263	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 oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 10 ligands modelled in this entry, 10 are unknown - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

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 [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	476/541 (87%)	0.29	40 (8%)	17 12	12, 40, 78, 87	0
1	B	474/541 (87%)	0.29	40 (8%)	17 12	13, 40, 79, 88	0
All	All	950/1082 (87%)	0.29	80 (8%)	17 12	12, 40, 79, 88	0

The worst 5 of 80 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	522	LEU	4.7
1	B	511	ALA	4.5
1	A	511	ALA	4.2
1	B	274	ALA	3.9
1	A	529	VAL	3.9

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
2	UNX	A	1006	1/1	0.27	0.25	47,47,47,47	0
2	UNX	A	1009	1/1	0.64	0.29	49,49,49,49	0
2	UNX	A	1001	1/1	0.69	0.26	42,42,42,42	0
2	UNX	A	1010	1/1	0.71	0.28	43,43,43,43	0
2	UNX	A	1002	1/1	0.75	0.17	25,25,25,25	0
2	UNX	A	1007	1/1	0.76	0.24	31,31,31,31	0
2	UNX	A	1008	1/1	0.81	0.26	24,24,24,24	0
2	UNX	A	1005	1/1	0.83	0.20	27,27,27,27	0
2	UNX	A	1004	1/1	0.87	0.15	34,34,34,34	0
2	UNX	A	1003	1/1	0.91	0.24	52,52,52,52	0

6.5 Other polymers [\(i\)](#)

There are no such residues in this entry.